MANAGING TECHNOLOGY PROJECTS



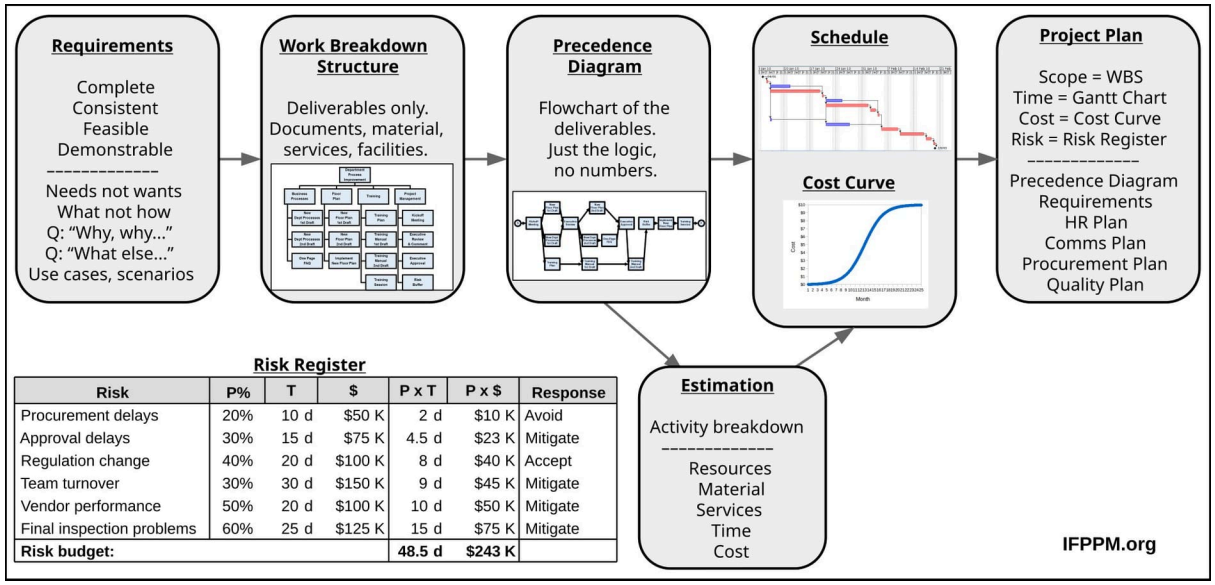
Robert Slate and Daniel Slate
De Anza College and the University of Notre Dame



Managing Technology Projects

Case Study Book

Project Management and Legal Considerations



Risk Register

Risk	P%	T	$	P x T	P x $	Response
Procurement delays	20%	10 d	$50 K	2 d	$10 K	Avoid
Approval delays	30%	15 d	$75 K	4.5 d	$23 K	Mitigate
Regulation change	40%	20 d	$100 K	8 d	$40 K	Accept
Team turnover	30%	30 d	$150 K	9 d	$45 K	Mitigate
Vendor performance	50%	20 d	$100 K	10 d	$50 K	Mitigate
Final inspection problems	60%	25 d	$125 K	15 d	$75 K	Mitigate
Risk budget:				**48.5 d**	**$243 K**	

Robert D. Slate

Adjunct Faculty, Computer Science and Information Systems, De Anza College

Daniel D. Slate

Visiting Assistant Professor, The Law School, University of Notre Dame

This text is disseminated via the Open Education Resource (OER) LibreTexts Project (https://LibreTexts.org) and like the thousands of other texts available within this powerful platform, it is freely available for reading, printing, and "consuming."

The LibreTexts mission is to bring together students, faculty, and scholars in a collaborative effort to provide an accessible, and comprehensive platform that empowers our community to develop, curate, adapt, and adopt openly licensed resources and technologies; through these efforts we can reduce the financial burden borne from traditional educational resource costs, ensuring education is more accessible for students and communities worldwide.

Most, but not all, pages in the library have licenses that may allow individuals to make changes, save, and print this book. Carefully consult the applicable license(s) before pursuing such effects. Instructors can adopt existing LibreTexts texts or Remix them to quickly build course-specific resources to meet the needs of their students. Unlike traditional textbooks, LibreTexts' web based origins allow powerful integration of advanced features and new technologies to support learning.



LibreTexts is the adaptable, user-friendly non-profit open education resource platform that educators trust for creating, customizing, and sharing accessible, interactive textbooks, adaptive homework, and ancillary materials. We collaborate with individuals and organizations to champion open education initiatives, support institutional publishing programs, drive curriculum development projects, and more.

The LibreTexts libraries are Powered by NICE CXone Expert and were supported by the Department of Education Open Textbook Pilot Project, the California Education Learning Lab, the UC Davis Office of the Provost, the UC Davis Library, the California State University Affordable Learning Solutions Program, and Merlot. This material is based upon work supported by the National Science Foundation under Grant No. 1246120, 1525057, and 1413739.

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation nor the US Department of Education.

Have questions or comments? For information about adoptions or adaptions contact info@LibreTexts.org or visit our main website at https://LibreTexts.org.

The LibreTexts name, logo, and book covers are trademarked by LibreTexts, Inc. (a 501(c)(3) not-for-profit organization) and protected against unauthorized use.

This text was compiled on 09/19/2026

TABLE OF CONTENTS

Licensing

Program Page

AI Use

Accessibility Statement

About the Authors

1: Introduction

2: Operation Warp Speed and COVID-19 Vaccines

2.1: Introduction

2.2: Operation Warp Speed

2.3: COVID-19 Vaccine Development

2.4: Drug Development for Traditional Drugs and Biologics

2.5: Immune System and Vaccines

2.5.1: Section Goals

2.5.2: Review of the Anatomy of the Immune System

2.5.3: Vaccines

2.5.4: References

2.5.5: Glossary

2.6: Introduction to Pharmacology

2.7: FDA Vaccine Development 101

2.8: mRNA Vaccines

2.9: Investigational New Drug (IND) Application

2.10: Biologics License Application (BLA)

2.11: Emergency Use Authorization (EUA)

2.12: Current Good Manufacturing Practice (CGMP)

2.13: CDC Advisory Committee on Immunization Practices (ACIP)

2.14: Project Management Considerations

2.14.1: Project Schedule Preparation

2.14.2: Project Acceleration Techniques

2.14.3: Risk Management

2.15: Legal Considerations

2.15.1: National Emergency Powers

2.15.2: Contracts

2.15.2.1: Selecting the Type of Contract

2.15.2.2: BARDA Contracts

2.15.2.3: BARDA contract with Moderna

2.15.2.4: BARDA Contract with SiO2 Medical Products, Inc. for Vials

2.15.3: Risk Mitigation and Liability

2.16: Glossary

3: EU GDPR Compliance

3.1: Introduction

3.2: The Regulation

3.2.1: Article 1 Subject-matter and objectives

3.2.2: Article 3 Territorial Scope

3.2.3: Article 4 Definitions

3.2.4: Article 5 Principles related to processing of data

3.2.5: Article 6 Lawfulness of processing

3.2.6: Article 7 Conditions for consent

3.2.7: Article 9 Processing of special categories of personal data

3.2.8: Article 13 Information to be provided where personal data are collected from the data subject

3.2.9: Article 15 Right of access by the data subject

3.2.10: Article 16 Right to rectification

3.2.11: Article 17 Right to erasure ('right to be forgotten')

3.2.12: Article 18 Right to restriction of processing

3.2.13: Article 20 Right to data portability

3.2.14: Article 21 Right to object

3.2.15: Article 22 Automated individual decision-making, including profiling

3.2.16: Article 24 Responsibility of the controller

3.2.17: Article 25 Data protection by design and by default

3.2.18: Article 28 Processor

3.2.19: Article 32 Security of processing

3.2.20: Article 33 Notification of a personal data breach to the supervisory authority

3.2.21: Article 34 Communication of a personal data breach to the data subject

3.2.22: Articles 35-36 Data Protection Impact Assessments

3.2.23: Article 37 Designation of the data protection officer

3.2.24: Article 38 Position of the data protection officer

3.2.25: Article 39 Tasks of the data protection officer

3.2.26: Article 42 Certification

3.2.27: Article 83 General conditions for imposing administrative fines

3.3: Implementing GDPR Compliance

3.3.1: Implementation Considerations

3.4: Project Management Considerations

3.5: Legal Considerations

3.6: Glossary

4: Network Switch Development

4.1: Introduction

4.2: Types of Network Switches
4.3: System Scope
4.4: Hardware System Design
4.4.1: Network ASICs
4.4.2: Chassis Design
4.4.3: Board Design
4.4.4: Signal integrity
4.4.5: Electro-magnetic Regulatory Compliance
4.4.6: Testing (HALT)
4.5: Software System Design
4.6: Project Management Considerations
4.6.1: Procurement
4.6.1.1: Bid Documents (RFI, RFP, etc.)
4.6.1.2: PR to PO procedure
4.6.2: Team Management
4.6.2.1: Content Theories of Motivation
4.6.2.2: Leadership
4.6.2.2.1: Introduction
4.6.2.2.2: The Nature of Leadership
4.6.2.2.3: The Leadership Process
4.6.2.2.4: Types of Leaders and Leader Emergence
4.6.2.2.5: The Trait Approach to Leadership
4.6.2.2.6: Behavioral Approaches to Leadership
4.6.2.2.7: Situational (Contingency) Approaches to Leadership
4.6.2.2.8: Substitutes for and Neutralizers of Leadership
4.6.2.2.9: Transformational, Visionary, and Charismatic Leadership
4.6.2.2.10: Leadership Needs in the 21st Century
4.6.2.2.11: Summary
4.7: Legal Considerations
4.7.1: Patents
4.7.1.1: The Patenting Process
4.7.1.2: Patent Disclosures
4.7.1.3: Patent cross-licensing
4.7.2: Trade Secrets
4.7.2.1: Trade Secret Protections
4.7.2.2: The Foundations of Trade Secrets Law
4.7.2.3: Elements of a Trade Secret
4.7.2.4: The Secrecy Requirement
4.7.2.5: Misappropriation of Trade Secrets

4.7.2.6: Remedies Available for the Misappropriation of Trade Secrets
4.7.2.7: Assessment Questions

4.7.3: Trademark Basics
4.7.3.1: Core Concepts
4.7.3.2: Early Trademark Systems
4.7.3.3: U.S. Trademark Law
4.7.3.4: Four Types of Trademarks
4.7.3.5: Subject Matter of Trademarks
4.7.3.6: The Spectrum of Distinctiveness
4.7.3.7: Bars to Trademark
4.7.3.8: Establishing Trademark Protection
4.7.3.9: Trademark Infringement
4.7.3.10: Trademark Remedies
4.7.3.11: Fair Use of Trademarks
4.7.3.12: Assessment Questions

4.8: Glossary

5: EU RoHS Compliance
5.1: Introduction
5.2: The Directive
5.3: Implementing the Directive
5.4: Project Management Considerations
5.5: Legal Considerations
5.6: Glossary

6: Boeing 737 Max Quality Issue Recovery 2018
6.1: Introduction
6.2: Project Scope- Steps to Recertification
6.3: Project Management Considerations
6.4: Managing Quality
6.4.1: Introduction to Managing Quality

6.4.2: Gurus of Quality
6.4.2.1: Walter Shewhart (1891-1967)
6.4.2.2: Edwards Deming (1900-1993)
6.4.2.3: Joseph Juran (1904-2008)
6.4.2.4: Philip Crosby (1926-2001) and Armand Fiegenbaum (1920-2014)

6.4.3: Cost of Quality

6.4.4: Quality Systems
6.4.4.1: Total Quality Management (TQM)
6.4.4.2: ISO
6.4.4.3: Six Sigma

6.4.5: Tools for Quality Improvement
6.4.5.1: Check Sheets
6.4.5.2: Histograms
6.4.5.3: Control Charts
6.4.5.4: Pareto Charts
6.4.5.5: Scatter Diagrams
6.4.5.6: Cause and Effect Diagrams
6.4.6: References
6.5: Legal Considerations
6.6: Glossary
7: Artificial Intelligence Usage by Utilities
7.1: Introduction
7.2: Specific Areas of Application of Artificial Intelligence
7.3: Artificial Intelligence Use by a California Utility (PGandE)
7.4: AI Initiatives by other California Utilities
7.4.1: Southern California Edison
7.4.2: San Diego Gas and Electric
7.5: Applying Artificial Intelligence to Wildfire Detection and Management
7.5.1: Vendor Solutions
7.5.1.1: N5 Sensors
7.5.1.2: Ladris
7.5.1.3: Pano.ai
7.6: Project Management Considerations
7.7: Legal Considerations
7.7.1: CPUC Governance
7.7.2: Wildfire Camera Privacy Issues
7.7.3: Impact of Potential AI Regulations
7.8: Glossary of Terms
8: COVID-19 Contact Tracing App
8.1: Description of the problem and potential solutions
8.2: Project Management Considerations
8.3: Legal Considerations
8.4: Glossary
Index
Glossary
Detailed Licensing
Detailed Licensing

This page titled Managing Technology Projects is shared under a license and was authored, remixed, and/or curated by Robert Slate and Daniel Slate.

Licensing

A detailed breakdown of this resource's licensing can be found in ***Back Matter/Detailed Licensing***.

This page titled Licensing is shared under a Public Domain license and was authored, remixed, and/or curated by LibreTexts.

Program Page



De Anza College sits in the heart of Silicon Valley, where innovation and opportunity are part of daily life, and also has one of the highest costs of living in the country. For our students, that makes every dollar count, and the price of a textbook can be the difference between staying enrolled or dropping a class. Our Zero Textbook Cost (ZTC) program closes that gap by giving students full access to high-quality, faculty-created course materials at no cost from the first day of class.

De Anza is proud of its diverse student community, and we hold our ZTC materials to the same standard of excellence that defines the college. Faculty across departments write, adapt, and review these resources themselves, keeping them accurate, current, and grounded in the perspectives of the students who use them, with accessibility and inclusive design built in from the start.

This program reflects the time and care our faculty invest in their students. If you have questions, feedback, or come across any issues with a ZTC resource, contact De Anza College's ZTC Grants Coordinator, Shagun Kaur.

Learn more about De Anza ZTC Program [www.deanza.edu/academic-services/oer/]

Note: The De Anza logo is a copyrighted trademark of De Anza College and may not be reproduced without permission.

This page titled Program Page is shared under a Public Domain license and was authored, remixed, and/or curated by LibreTexts.

AI Use

The two authors did not personally make use of AI in the creation of this book for text or illustrations as we were instructed that that was not permitted. The project manager, Shagun Kaur, demonstrated the use of AI to create accessibility alternative text for a single complex illustration and that was retained.

Chapter 7 Artificial Intelligence Usage by Utilities describes how the three major public utilities in California use or plan to use artificial intelligence in their operations. Chapter 7 also describes how some camera and sensor vendors make use of AI in wildfire detection applications.

This page titled AI Use is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Robert Slate and Daniel Slate.

Accessibility Statement

Accessibility

De Anza College is dedicated to creating learning materials that are accessible and benefit all students. Our materials follow the Web Content Accessibility Guidelines (WCAG) 2.1 Level AA. Our authors and contracted accessibility specialists checked all resources using this ASCCC OERI checklist, the accessibility tool in LibreTexts, WAVE, JAWS, and other more detailed checks for complex materials such as interactive components.

We selected LibreTexts as our preferred platform because of its commitment to accessibility. In addition to the platform's inherent accessibility, we collaborated with the LibreTexts team to fix any accessibility issues we found.

Our Accessibility Review Process

Accessibility was built into the resource as it was written. This resource went through the following stages of review:

1. **Template check.** The project coordinator and instructional designers checked the first chapter for accessibility as a model to ensure later chapters meet accessibility standards.
2. **Progress checks.** The manuscript was checked at approximately 40%, 60%, and 80% completion to catch and correct accessibility issues in production.
3. **Specialist review.** Once the manuscript was complete, an accessibility expert conducted an accessibility check.
4. **Remediation.** The author and the project manager who conducted the earlier progress checks worked together to remediate the issues identified by the accessibility expert.
5. **Final check.** The resource underwent another full check by the accessibility expert before the book was deemed ready for publication.

This process combines ongoing checks during writing with a final specialist review, so issues are caught early and confirmed fixed before publication.

Report an Accessibility Problem

We are always looking for ways to make our resources more accessible. If you run into a barrier while using this resource, please let us know so we can look into it. When you contact us, it helps to include:

- The name of the resource
- The location of the problem (a web address or page/chapter description)
- A description of the problem you encountered
- The device, browser, and any assistive technology you were using (e.g., Windows 11, Google Chrome, JAWS screen reader)

Contact us: De Anza ZTC team

This statement was last updated August 2026

This resource was project-managed by **Shagun Kaur**, overseeing scope, budget, and schedule; stakeholder coordination via monthly meetings; training and platform support (LibreTexts, troubleshooting, technical support); QA/accessibility check scheduling; and licensing and copyright governance.

A Note on Known Limitations

No accessibility check is perfect. We've made a genuine, sustained effort to catch and fix issues throughout the writing and review process described above, but some barriers may remain, especially in complex or interactive content. We'd rather be upfront about that than claim a standard of perfection we can't guarantee.

This page titled Accessibility Statement is shared under a Public Domain license and was authored, remixed, and/or curated by Robert Slate and Daniel Slate.

About the Authors

Team Information

Author	Bio
 Robert Slate, Adjunct Faculty, Computer Science and Information Systems, De Anza College	Robert Slate teaches in the Computer Science and Information Systems department of De Anza College. Robert holds a BSE (Electrical Engineering/Computer Science) from Princeton University and an MBA (Product Management) from Fairleigh Dickinson University. He also holds a Graduate Certificate in Management Science and Engineering from Stanford University and a Certificate in Accounting from UC Berkeley. **Professional Certifications** • Project Management Institute: PMP (Project Management Professional), PgMP (Program Management Professional), PfMP (Portfolio Management Professional), PMI-ACP (Agile Certified Practitioner), PMI-PBA (PMI-Professional in Business Analysis) • International Institute of Business Analysis: CBAP (Certified Business Analysis Professional) • Quality: Six Sigma Lean Black Belt Professional, Lean Six Sigma Master Black Belt **Courses Taught** • CIS 79 Managing Technology Projects • CIS 95H Business and Requirement Analysis • CIS 43 Systems Design **Work Experience** • Juniper Networks: Sr. Director Strategic Execution, Head of PMO for BU, Business Operations, Engineering Operations, Chief of Staff to GM • Ericsson: Sr. Director Product Management, PLCM Governance • Extreme Networks: Head of Adv. Eng. Program Management, GM Chief of Staff • Amdahl: Sr. Director, Product Planning/Product Management, roles in computer architecture, firmware development, and computer performance analysis.

Author	Bio
 Daniel Slate, Visiting Assistant Professor, The Law School, University of Notre Dame	Daniel Slate teaches at Notre Dame Law School, where his research engages questions in law and technology, emergency powers, and constitutional history. He holds a JD, PhD (Political Science), and BA (Philosophy, with honors and distinction) from Stanford University. During his time in the tech sector, he worked on product management, business development, and privacy and civil liberties teams. He is also a Faculty Affiliate of the AI Trust and Reliability Lab and the Human-Centered Responsible AI Lab at the Data, AI, and Computing Initiative of the University of Notre Dame. ## Courses Taught - LAW 73132 - Data Privacy Law ## Selected Publications - *Law, Furnace, Plow, Mind*, *in* Technology Ethics (Gregory J. Robson & Jonathan Y. Tsou, eds., 2d ed.) (forthcoming 2026/2027). - *Converting Raw Data Into Actionable Information: A topical review of artificial intelligence, machine learning, and digital twins in disaster management*, 14 IEEE Geosci. & Remote Sensing Mag. 76 (2026) (peer-reviewed) (IEEE) (DOI link) (with A. R. Albayrak et al.). - *Iterative Contextual Consent: AI-enabled Data Privacy Contracts*, HAIPS 25 Proc. Hum.-Centered AI Priv. & Sec. (ACM Comput. & Commc'ns Sec.) (2025) (ACM DL) (DOI link) (with Chaoran Chen, Tobi Ji-Jia Li & Yaxing Yao). - *Infringed*, 3 J. Am. Con. Hist. 381 (2025) (peer-reviewed). (SSRN) (www) (DOI link) (cited at oral argument in the United States Supreme Court in *Wolford v. Lopez*). - *Artificial Intelligence Adoption in Electric Utilities*, 45 Energy L. J. 1 (2024). (SSRN) (with Alexandre Parisot, Liang Min, Patrick Panciatici & Pascal Van Hentenryck). - *Crisis Government's Legitimacy Paradox: Foreseeability and Unobservable Success*, *in* Intersections, Reinforcements, and Cascades 248 (Daniel Zimmer, Trond Undheim, & Paul N. Edwards eds., 2023). (SSRN) (DOI link) - *Franklin's Talmud: Hebraic Republicanism in the Constitutional Convention and the Debates over Ratification, 1787–1788*, 1 J. Am. Con. Hist. 232 (2023). (SSRN) (DOI link) - *The Architecture of Privacy: On Engineering Technologies that Can Deliver Trustworthy Safeguards* (O'Reilly, 2015) (with Courtney Bowman, Ari Gesher, John K. Grant & Elissa Lerner).

This page titled About the Authors is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Robert Slate and Daniel Slate.

1: Introduction

Overview

CIS 79 Managing Technology Projects is a course in the Computer Science and Information Systems department at De Anza College in Cupertino, California in the heart of Silicon Valley. It is an elective course in the Project Management curriculum for the Certificate of Achievement, Certificate of Achievement-Advanced, and AA degree.

Case Study Approach

The course has utilized a case study approach since the Fall of 2020. The case studies are used for weekly assignments to reinforce and apply principles covered in course lectures.

One of the key pedagogical principles of CIS 79 is to prepare students to be industry-ready. This includes preparing some of the key phase exit presentations used for the project life cycle employed by many Silicon Valley technology companies. This includes Phase 0 exit (Concept), Phase 1 exit (Specifications; Project plan: schedule, budget, resources), and Phase 4 exit (readiness to ship to customers). In order to do this, some representative projects that technology companies have dealt with or will deal with need to be made available to students. Our preference is to use real projects rather than fictitious projects as they have more of a tendency to hold the interests of students.

Many of the case studies were developed based on the authors' direct experience. However, some of the case studies cover topics outside of the authors' direct experience, but which were current and relevant at the time they were developed and used in the course. For example, Operation Warp Speed that managed the creation and mass production of the initial COVID-19 vaccines was occurring in real time in the Fall of 2020 as the course was in progress. This was also the case with projects to develop COVID-19 vaccines within pharmaceutical companies. This provided students with exposure to important technology projects while they were actually happening, with outcomes far from certain. This helped simulate the uncertainties faced by real project managers when they are managing their projects. Even as new vaccines are being developed to address evolving mutated strains of the virus, the original projects remain of high interest due to how a project addressing a worldwide emergency was handled from a project management perspective to aggressively accelerate schedules from normal 5-10 year pharmaceutical development timeframes to something on the order of one year.

Project Types

The case studies cover a range of project types:

- Product development
- Computer networking technology (Ethernet switch development)
- COVID-19 vaccines and contact tracing apps
- Compliance
- Quality improvement (e.g., Boeing 737 Max 8 recertification effort because air transportation safety is paramount)
- Public utilities use of artificial intelligence technology

Compliance projects include implementation work to be in conformance with environmental initiatives as well as data privacy initiatives such as:

- EU Restriction of Hazardous substances (RoHS) and
- EU General Data Protection Regulation (GDPR).

The European Union (EU) has been in the forefront of these initiatives. Many US companies may have disregarded these initiatives initially, only to discover later that they are applicable when they realize that 25-30% of their revenues are derived from the EU or that many of their customers reside in the EU.

Both authors have worked on the EU Restriction of Hazardous Substances (RoHS) directive. One author did so in the role of project manager in two companies and as a consultant for several others. The other author ran a compliance checker tool against

the Bill of Materials (BOMs) for each SKU (Stock Keeping Unit) that needed to achieve compliance. RoHS compliance required a cross-functional team effort with a massive amount of data collection, detailed technical analysis, engineering development and testing effort, home-grown tool creation, and laborious repeated compliance attainment checking.

Our intent is to continuously add additional case studies as interesting technology projects address problems of interest to students. For example, the case study on artificial intelligence is new for the fall of 2026. Artificial intelligence tools are becoming pervasive and colleges such as De Anza College are adding certificates and degree programs in artificial intelligence to stay at the forefront of this technology. Specifically, the case study builds on a paper whose lead author is one of the authors of this textbook. It addresses public utilities and their current and planned usage of artificial intelligence for a variety of applications. Since public utilities are considered risk averse, the use of emerging technology may seem unusual. Usually there is oversight in the use of new technologies by bodies such as the California Public Utilities Commission. The most interesting usage of artificial intelligence to many who reside in California is its use in detecting and monitoring the spread of wildfires which have been quite destructive in recent years.

Case Study Structure

The structure of each case study is roughly as follows:

- Introduction/Overview
- Learning objectives
- Body of the case study
- Project Management considerations, reinforcing concepts covered in lectures. This includes basic project management concepts (e.g., task scheduling), special considerations in accelerating schedules, risk management, and team management.
- Legal Considerations: Legal considerations for each case study are discussed. This may include intellectual property considerations related to a development project (such as patents and trademarks) or special contractual requirements. A project manager may need to consult the organization's legal staff, so this material will help provide a working vocabulary for those discussions. Caveat: This is not a "case law" textbook (also known as a "casebook") that covers excerpts from legal cases, and nothing in this textbook should be taken as legal advice (for that, consult an attorney). The European Union's General Data Protection Regulation (GDPR) is covered in greater depth in LAW 73132 - Data Privacy Law at the Law School at the University of Notre Dame.
- Glossary/Lexicon (Technology-specific vocabulary) for each case study. This will help the project manager speak the language of each technical realm.

A question is often asked as to whether a project manager can be successful if he or she is unfamiliar with the technology, technical vocabulary, and technical processes associated with the project. One may not need to be an expert in a project's technology, but a familiarity with the technology, technical vocabulary, and technical processes will help the project manager gain the respect of the project teams, or not. This is the primary reason why the Glossary/Lexicon is included.

An important consideration is that not all companies are the same. There may be differences in the following areas, to name a few:

- Priorities. For some companies, time to market may be the most important element. This can be viewed in conjunction with the project management triangle (more on this later).
- Risk appetite. Some companies are risk averse (e.g., public utilities) whereas other companies are more willing to take risks.
- Budget management.
    - Some companies take a dim view of project managers padding their budgets. This is based on the concept of opportunity cost. The money tied up unused in a project could have been invested elsewhere by the company for a higher return.
    - Some projects, particularly those with the Unites States government, may require the use of Earned Value Management (EVM). Most private sector projects do not.

OER

De Anza College has many initiatives to produce Open Education Resources (OER), including an initiative to provide OER textbooks for all of the courses in the project management curriculum. This textbook is part of that initiative. Goals include improving learning outcomes and helping to ensure that the cost of a textbook is not a barrier to learning.

This is the first OER effort by both of the authors, requiring some new learning. It has involved both standing on the shoulders of prior authors and making use of their OER open licensed material or public domain material as well as creating our own material. We have made use of CC BY 4.0 academic papers (hopefully getting the attributions totally correct as required) and government material in the public domain, such as that from the United States Centers for Disease Control on vaccines. As appropriate, sections of certain regulations, laws, and directives have been extracted verbatim to give the reader the feel for what it takes to understand the requirements to be successful in a compliance project. Since some of the regulations are from the EU, the European spellings of words are retained (and hopefully unchanged by editors trying to be helpful). Additionally, the section numbers and letters in laws is carried forward as it appears in the laws, even though it may seem odd in some cases.

Some of the work of prior OER authors that was "remixed" into this textbook needed to be altered in order to meet evolving accessibility requirements. The accessibility work for new material and remixed material required some effort and we hope we did it all correctly.

Acknowledgements

- Shagun Kaur of De Anza College, for her project management of the Project Management OER textbook project.
- Felix Amoruwa, Adjunct Faculty at De Anza College, for reviewing the entire content of the textbook and providing feedback.

Chapter thumbnail image: File:Project Management - Project Planning Process Flowchart.svg - Wikimedia Commons. The SVG file was converted to JPG to be acceptable to the Libretexts system for inclusion as a thumbnail. Author: IFPPM.org (practicalpm in the Wikimedia citation). March 28, 2026. This file is licensed under the Creative Commons Attribution-Share Alike 4.0 International license.

This page titled 1: Introduction is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

2: OPERATION WARP SPEED AND COVID-19 VACCINES

2.1: Introduction

2.2: Operation Warp Speed

2.3: COVID-19 Vaccine Development

2.4: Drug Development for Traditional Drugs and Biologics

2.5: Immune System and Vaccines

- 2.5.1: Section Goals
- 2.5.2: Review of the Anatomy of the Immune System
- 2.5.3: Vaccines
- 2.5.4: References
- 2.5.5: Glossary

2.6: Introduction to Pharmacology

2.7: FDA Vaccine Development 101

2.8: mRNA Vaccines

2.9: Investigational New Drug (IND) Application

2.10: Biologics License Application (BLA)

2.11: Emergency Use Authorization (EUA)

2.12: Current Good Manufacturing Practice (CGMP)

2.13: CDC Advisory Committee on Immunization Practices (ACIP)

2.14: Project Management Considerations

- 2.14.1: Project Schedule Preparation
- 2.14.2: Project Acceleration Techniques
- 2.14.3: Risk Management

2.15: Legal Considerations

- 2.15.1: National Emergency Powers
- **2.15.2: Contracts**
  - 2.15.2.1: Selecting the Type of Contract
  - 2.15.2.2: BARDA Contracts
  - 2.15.2.3: BARDA contract with Moderna
  - 2.15.2.4: BARDA Contract with SiO2 Medical Products, Inc. for Vials
- 2.15.3: Risk Mitigation and Liability

2.16: Glossary

This page titled 2: Operation Warp Speed and COVID-19 Vaccines is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

2.1: Introduction

Operation Warp Speed, the United Stated federal government project that managed the development and mass production of the initial COVID-19 vaccines, was occurring in real time in the fall of 2020 as the CIS 79 Managing Technology Projects course was in progress. While creating the CIS 79 course from scratch that summer, Operation Warp Speed and the projects at pharmaceutical companies to develop COVID-19 vaccines seemed like it would be a good case study.

> **Note**
>
> Operation Warp Speed was a public–private partnership, initiated by the federal government of the United States, to facilitate and accelerate the development, manufacturing, and distribution of COVID-19 vaccines, therapeutics, and diagnostics.[1]

This case study provided students with exposure to an important project while it was happening in real time, with outcomes far from certain. This helped simulate the uncertainties faced by real project managers when they are managing their projects.

Even as new vaccines are being developed to address evolving mutated strains of the virus, the original project remains of high interest due to how a project addressing a worldwide emergency was handled from a project management perspective to aggressively accelerate schedules from normal 5-10 year pharmaceutical development timeframes to something on the order of one year.

For those of us who have worked in project management for years and in conversations with those in the pharmaceutical field, the successful development of the vaccine, manufacturing scale-up for billions of doses, successful execution of the complex logistics of distribution of vaccines that required low temperature handling, was quite remarkable.

In this chapter we will take a look at:

- the organization of Operation Warp Speed and the scope of its activities
- traditional pharmaceutical development cycles
- what mRNA vaccines are
- project management considerations, including schedule acceleration techniques and
- legal considerations, including looking at emergency powers, contracts between the US government and pharmaceutical companies, and handling legal liabilities.

Aside from the impact on human health, government measures to contain the spread of the virus had a significant impact on the economy, on livelihoods, and personal freedoms. In Silicon Valley where De Anza College is, Santa Clara County health officials were the first to impose lock downs. Workers could not go to work unless their work was classified as "essential." Grocery stores erected plexiglass barriers to protect cashiers. Footprints were place on floors of business establishments that were allowed to remain open to help enforce the six foot spacing guidelines. Watching the news and reading the news were sobering if not terrifying.

This resulted in a lot of pressure around the world for the scientific community to develop, test, and deploy a vaccine that would be effective and safe, and to do it as quickly as possible.

References

1. Wikipedia - Operation Warp Speed. Accessed August 24, 2020.

This page titled 2.1: Introduction is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

2.2: Operation Warp Speed

> **Note**
>
> Information in this section is derived or extracted from government publications in the public domain.

As part of the Unites States vaccine effort, on May 15, 2020, the federal government announced Operation Warp Speed (OWS).[8]

Goal/Mission

Operation Warp Speed's goal was to produce and deliver 300 million doses of safe and effective vaccines with the initial doses available by January 2021, as part of a broader strategy to accelerate the development, manufacturing, and distribution of COVID-19 vaccines, therapeutics, and diagnostics (collectively known as countermeasures).[2]

The US government expressed its commitment to providing free or low-cost COVID-19 countermeasures to the American people as fast as possible. Any vaccine or therapeutic doses purchased with US taxpayer dollars would be given to the American people at no cost.[6, page 4]



Figure 2.2.1: Operation Warp Speed Seal. (Public Domain; Department of Defense[10])

Leadership

HHS Secretary Alex Azar and Defense Secretary Mark Esper provided oversight of OWS, with Dr. Moncef Slaoui designated as chief advisor and General Gustave F. Perna confirmed as the chief operating officer.[6]

Funding

On March 27, 2020, the United States Congress passed the CARES Act (Coronavirus Aid, Relief, and Economic Security), which initially provided about $10 billion in funding for Operation Warp Speed.[5]

Congress directed almost $10 billion to this effort through supplemental funding, including the CARES Act. Congress has also appropriated other flexible funding. The almost $10 billion specifically directed includes more than $6.5 billion designated for countermeasure development through BARDA and $3 billion for NIH research.[6, page 5]

Funding was increased to about $18 billion by October 2020.[9]

Constituency

Operation Warp Speed was a public–private partnership, initiated by the federal government of the United States, to facilitate and accelerate the development, manufacturing, and distribution of COVID-19 vaccines, therapeutics, and diagnostics.[1]

Operation Warp Speed was a partnership among components of the Department of Health and Human Services (HHS), including the Centers for Disease Control and Prevention (CDC), the Food and Drug Administration (FDA), the National Institutes of Health (NIH), and the Biomedical Advanced Research and Development Authority (BARDA), and the Department of Defense (DoD). OWS engaged with private firms and other federal agencies, including the Department of Agriculture, the Department of Energy, and the Department of Veterans Affairs. OWS coordinated existing HHS-wide efforts, including the NIH's Accelerating COVID-19 Therapeutic Interventions and Vaccines (ACTIV) partnership, NIH's Rapid Acceleration of Diagnostics (RADx) initiative, and work by BARDA.[6, page 1]

Note

Operation Warp Speed agencies:[6]

- Department of Health and Human Services (HHS), including the
  - Centers for Disease Control and Prevention (CDC)
  - Food and Drug Administration (FDA)
  - National Institutes of Health (NIH), and
  - Biomedical Advanced Research and Development Authority (BARDA)
- Department of Defense
- Other federal agencies, including the Department of Agriculture, the Department of Energy, and the Department of Veterans Affairs.

Private Companies Involved

- "As of July, eight companies had been chosen for funding to expedite development and preparation for manufacturing their respective vaccine candidates: Johnson & Johnson (Janssen Pharmaceutical), AstraZeneca-University of Oxford, Pfizer-BioNTech, Moderna, Merck, Vaxart, Inovio, and Novavax. In June agreements of $1.2 billion for successful vaccine were signed with AstraZeneca and $483 million with Moderna."[4]
- "In July, Operation Warp Speed agreed to pay $1.6 billion to Novavax if clinical trials are successful to produce 100 million doses of its candidate vaccine. Novavax is a company with no history of marketed products. Also in July, Pfizer and BioNTech received an order of $2 billion to manufacture 100 million doses of a COVID-19 vaccine when it is shown to be safe and effective, and is licensed. On July 31st, Sanofi and GSK were funded for $2.1 billion."[4]

Managing Vaccine Development

To accelerate development while maintaining standards for safety and efficacy, Operation Warp Speed selected the most promising countermeasure candidates and provided coordinated government support.

Protocols for the demonstration of safety and efficacy were aligned, allowing these harmonized clinical trials to proceed more quickly, and the protocols for the trials will be overseen by the federal government (NIH), as opposed to traditional public-private partnerships, in which pharmaceutical companies decide on their own protocols.

Rather than eliminating steps from traditional development timelines, steps will proceed simultaneously, such as starting manufacturing of vaccines and therapeutics at industrial scale well before the demonstration of efficacy and safety as happens normally. This increases the financial risk, but not the product risk.[6, page 1]

Note

The bolding of some text in this section is to emphasize project management schedule acceleration techniques as an aid to the student.

The COVID-19 vaccine development plan was roughly:[6]

- Fourteen promising candidates were chosen from the 100+ vaccine candidates currently in development, some of them already in clinical trials with U.S. government support.
- The 14 vaccine candidates were narrowed down to about seven candidates, representing the most promising candidates from a range of technology options (nucleic acid, viral vector, protein subunit), which will go through further testing in early-stage clinical trials.
- Large-scale randomized trials for the demonstration of safety and efficacy proceeded for the most promising candidates.

Disbursements for Development[6]

Select actions to support Operation Warp Speed vaccine and therapeutic development included:[6]

- March 30: HHS announced $456 million in funds for Johnson & Johnson's (Janssen) candidate vaccine. Phase 1 clinical trials began in Belgium on July 24th and in the U.S on July 27th.
- April 16: HHS made up to $483 million in support available for Moderna's candidate vaccine, which began Phase 1 trials on March 16 and received a fast-track designation from FDA. This agreement was expanded on July 26 to include an additional $472 million to support late-stage clinical development, including the expanded Phase 3 study of the company's mRNA vaccine, which began on July 27th.
- May 21: HHS announced up to $1.2 billion in support for AstraZeneca's candidate vaccine, developed in conjunction with the University of Oxford. The agreement is to make available at least 300 million doses of the vaccine for the United States, with the first doses delivered as early as October 2020 and Phase 3 clinical studies beginning this summer with approximately 30,000 volunteers in the United States.
- July 7: HHS announced $450 million in funds to support the large-scale manufacturing of Regeneron's COVID-19 investigational anti-viral antibody treatment, REGN-COV2. This agreement is the first of a number of OWS awards to support potential therapeutics all the way through to manufacturing. As part of the manufacturing demonstration project, doses of the medicine will be packaged and ready to ship immediately if clinical trials are successful and FDA grants EUA or licensure.
- July 7: HHS announced $1.6 billion in funds to support the large-scale manufacturing of Novavax's vaccine candidate. By funding Novavax's manufacturing effort, the federal government will own the 100 million doses expected to result from the demonstration project.
- July 22: HHS announced up to $1.95 billion in funds to Pfizer for the large-scale manufacturing and nationwide distribution of 100 million doses of their vaccine candidate. The federal government will own the 100 million doses of vaccine initially produced as a result of this agreement, and Pfizer will deliver the doses in the United States if the product successfully receives FDA EUA or licensure, as outlined in FDA guidance, after completing demonstration of safety and efficacy in a large Phase 3 clinical trial, which began July 27th.
- July 31: HHS announced approximately $2 billion in funds to support the advanced development, including clinical trials and large scale manufacturing, of Sanofi and GlaxoSmithKline's (GSK) investigational adjuvanted vaccine. By funding the manufacturing effort, the federal government will own the approximately 100 million doses expected to result from the demonstration project. The adjuvanted vaccine doses could be used in clinical trials or, if the FDA authorizes use, as outlined in agency guidance, the doses would be distributed as part of a COVID-19 vaccination campaign.
- August 5: HHS announced approximately $1 billion in funds to support the large-scale manufacturing and delivery of Johnson & Johnson's (Janssen) investigational vaccine candidate. Under the terms of the agreement, the U.S. Government will own the resulting 100 million doses of vaccine, and will have the option to acquire more. The company's investigational vaccine relies on Janssen's recombinant adenovirus technology, AdVac, a technology used to develop and manufacture Janssen's Ebola vaccine with BARDA support; that vaccine received European Commission approval and was used in the Democratic Republic of the Congo (DRC) and Rwanda during the 2018-2020 Ebola outbreak that began in the DRC.
- August 11: HHS announced up to $1.5 billion in funds to support the large-scale manufacturing and delivery of Moderna's investigational vaccine candidate. Under the terms of the agreement, the U.S. Government will own the resulting 100 million doses of vaccine, and will have the option to acquire more. The vaccine, called mRNA-1273, has been co-developed by Moderna and scientists from the National Institute of Allergy and Infectious Diseases (NIAID), part of the National Institutes of Health. NIAID has continued to support the vaccine's development including nonclinical studies and clinical

trials. Additionally, BARDA has supported phase 2/3 clinical trials, vaccine manufacturing scale up and other development activities for this vaccine. The Phase 3 clinical trial, which began July 27, is the first government-funded Phase 3 clinical trial for a COVID-19 vaccine in the United States.

Manufacturing[6]

The United States federal government made investments in the necessary manufacturing capacity at its own risk, giving private pharmaceutical firms the confidence to invest aggressively in development, facilitating faster distribution of an eventual vaccine.

The plan was to proceed with ramping up Manufacturing capacity for selected candidates while they were still in development, rather than scaling up after approval or authorization.

Actions to support Operation Warp Speed manufacturing efforts included:

- The May 21, April 16, and March 30 HHS agreements with AstraZeneca, Moderna, and Johnson & Johnson respectively include investments in manufacturing capabilities.
- June 1: HHS announced a task order with Emergent BioSolutions to advance domestic manufacturing capabilities and capacity for a potential COVID-19 vaccine as well as therapeutics, worth approximately $628 million, using Emergent's BARDA-supported Center for Innovation in Advanced Department and Manufacturing.
- July 27: HHS announced a task order with Texas A&M University and FUJIFILM to advance domestic manufacturing capabilities and capacity for a potential COVID-19 vaccine, worth approximately $265 million, using another BARDA-supported CIADM.
- August 4: Grand River Aseptic Manufacturing Inc., (GRAM) Grand Rapids, Michigan, was awarded a $160 million firm-fixed-price contract for domestic aseptic fill and finish manufacturing capacity for critical vaccines and therapeutics in response to the COVID-19 pandemic.

Distribution[6]

Before the countermeasures were approved or authorized, the program built the necessary plans and infrastructure for distribution.

HHS planned for a tiered approach to vaccine and therapeutic distribution, which built on allocation methodology developed as part of pandemic flu planning and was adjusted based on experience from the COVID-19 response, data on the virus and its impact on populations, and the performance of a given countermeasure, and the needs of the essential workforce. Operation Warp Speed expanded domestic manufacturing and supplies of specialized materials and resources, such as glass vials, that can be necessary for distribution. DoD's involvement enabled faster distribution and administration than would have otherwise been possible.

Select actions to support Operation Warp Speed distribution efforts include:

- May 12: DoD and HHS announced a $138 million contract with ApiJect for more than 100 million prefilled syringes for distribution across the United States by year-end 2020, as well as the development of manufacturing capacity for the ultimate production goal of over 500 million prefilled syringes in 2021.
- June 9: HHS and DoD announced a joint effort to increase domestic manufacturing capacity for vials that may be needed for vaccines and treatments.
- June 11: HHS announced $204 million in funds to Corning to expand the domestic manufacturing capacity to produce approximately 164 million Valor Glass vials per year if needed. Valor Glass provides chemical durability to minimize particulate contamination. The specialized glass allows for rapid filling and capping methods that can increase manufacturing throughput by as much as 50 percent compared with conventional filling lines, which in turn can reduce the overall manufacturing time for vaccines and therapies.
- June 11: HHS announced $143 million to SiO2 Materials Science to ramp up capacity to produce the company's glass-coated plastic container, which can be used for drugs and vaccines. The new lines provide the capacity to produce an additional 120 million vials per year if needed.

Government Accountability Office report[7]

Operation Warp Speed was a federal effort that supported multiple COVID-19 vaccine candidates to speed up development.

The General Accountability Office (GAO) analyzed the program's vaccine candidates and found that their development followed traditional practices, with some adaptations. For example, **some clinical trial phases overlapped with each other and with animal studies to accelerate development**. Two vaccines were authorized for emergency use.

The GAO found that agencies were working to help mitigate vaccine manufacturing challenges. For example, agencies have taken steps to require contractors to prioritize vaccine production supplies.

Operation Warp Speed (OWS) aimed to help accelerate the development of a COVID-19 vaccine. GAO found that OWS and vaccine companies adopted several strategies to accelerate vaccine development and mitigate risk. For example, OWS selected vaccine candidates that use different mechanisms to stimulate an immune response (i.e., platform technologies). **Vaccine companies also took steps, such as starting large-scale manufacturing during clinical trials and combining clinical trial phases or running them concurrently**. Clinical trials gather data on safety and efficacy, with more participants in each successive phase (e.g., phase 3 has more participants than phase 2).

As of January 30, 2021, five of the six Operation Warp Speed vaccine candidates had entered phase 3 clinical trials, two of which (Moderna's and Pfizer/BioNTech's vaccines) received an emergency use authorization (EUA) from the Food and Drug Administration (FDA). For vaccines that received EUA, additional data on vaccine effectiveness will be generated from further follow-up of participants in clinical trials already underway before the EUA was issued.

Technology readiness

GAO's analysis of the Operation Warp Speed vaccine candidates' technology readiness levels (TRL), an indicator of technology maturity, showed that COVID-19 vaccine development under OWS generally followed traditional practices, with some adaptations. FDA issued specific guidance that identified ways that vaccine development may be accelerated during the pandemic. Vaccine companies told GAO that the primary difference from a non-pandemic environment was the compressed timelines. To meet OWS timelines, some vaccine companies relied on data from other vaccines using the same platforms, where available, or **conducted certain animal studies at the same time as clinical trials**. However, as is done in a non-pandemic environment, all vaccine companies gathered initial safety and antibody response data with a small number of participants before proceeding into large-scale human studies (e.g., phase 3 clinical trials). The two EUAs issued in December 2020 were based on analyses of clinical trial participants and showed about 95 percent efficacy for each vaccine. These analyses included assessments of efficacy after individuals were given two doses of vaccine and after they were monitored for about 2 months for adverse events.

Manufacturing

As of January 2021, five of the six Operation Warp Speed vaccine companies had started commercial scale manufacturing. OWS officials reported that as of January 31, 2021, companies had released 63.7 million doses—about 32 percent of the 200 million doses that, according to OWS, companies with EUAs have been contracted to provide by March 31, 2021. Vaccine companies face a number of challenges in scaling up manufacturing to produce hundreds of millions of doses under OWS's accelerated timelines. DOD and HHS worked with vaccine companies to help mitigate manufacturing challenges, including:

- **Limited manufacturing capacity:** A shortage of facilities with capacity to handle the vaccine manufacturing needs can lead to production bottlenecks. Vaccine companies worked in partnership with OWS to expand production capacity. For example, one vaccine company told GAO that HHS's Biomedical Advanced Research and Development Authority helped them identify an additional manufacturing partner to increase production. Additionally, the U.S. Army Corps of Engineers oversaw construction projects to expand capacity at vaccine manufacturing facilities.
- **Disruptions to manufacturing supply chains:** Vaccine manufacturing supply chains have been strained by the global demand for certain goods and workforce disruptions caused by the global pandemic. For example, representatives from one facility manufacturing COVID-19 vaccines stated that they experienced challenges obtaining materials, including reagents and certain chemicals. They also said that due to global demand, they waited 4 to 12 weeks for items that before the pandemic were typically available for shipment within one week. Vaccine companies and DOD and HHS officials told GAO they have undertaken several efforts to address possible manufacturing disruptions and mitigate supply chain challenges. These efforts included federal assistance to (1) expedite procurement and delivery of critical manufacturing equipment, (2) develop a list of critical supplies that are common across the six OWS vaccine candidates, and (3) expedite the delivery of

necessary equipment and goods coming into the United States. Additionally, DOD and HHS officials said that as of December 2020 they had placed prioritized ratings on 18 supply contracts for vaccine companies under the Defense Production Act, which allows federal agencies with delegated authority to require contractors to prioritize those contracts for supplies needed for vaccine production.

- **Gaps in the available workforce:** Hiring and training personnel with the specialized skills needed to run vaccine manufacturing processes can be challenging. OWS officials stated that they have worked with the Department of State to expedite visa approval for key technical personnel, including technicians and engineers to assist with installing, testing, and certifying critical equipment manufactured overseas. OWS officials also stated that they requested that 16 DOD personnel be detailed to serve as quality control staff at two vaccine manufacturing sites until the organizations can hire the required personnel.

Why the Government Accountability Office Did This Study

As of February 5, 2021, the U.S. had over 26 million cumulative reported cases of COVID-19 and about 449,020 reported deaths, according to the Centers for Disease Control and Prevention. The country also continued to experience serious economic repercussions, with the unemployment rate and number of unemployed in January 2021 at nearly twice their pre-pandemic levels in February 2020.

In May 2020, Operation Warp Speed was launched and included a goal of producing 300 million doses of safe and effective COVID-19 vaccines with initial doses available by January 2021. Although FDA has authorized two vaccines for emergency use, Operation Warp Speed has not yet met its production goal by January 2021. Such vaccines are crucial to mitigate the public health and economic impacts of the pandemic.

GAO was asked to review Operation Warp Speed vaccine development efforts. This GAO report examined:

1. the characteristics and status of the Operation Warp Speed vaccines,
2. how developmental processes have been adapted to meet Operation Warp Speed timelines, and
3. the challenges that companies have faced with scaling up manufacturing and the steps they are taking to address those challenges.

GAO administered a questionnaire based on HHS's medical countermeasures TRL criteria to the six Operation Warp Speed vaccine companies to evaluate the COVID-19 vaccine development processes. GAO also collected and reviewed supporting documentation on vaccine development and conducted interviews with representatives from each of the companies on vaccine development and manufacturing.

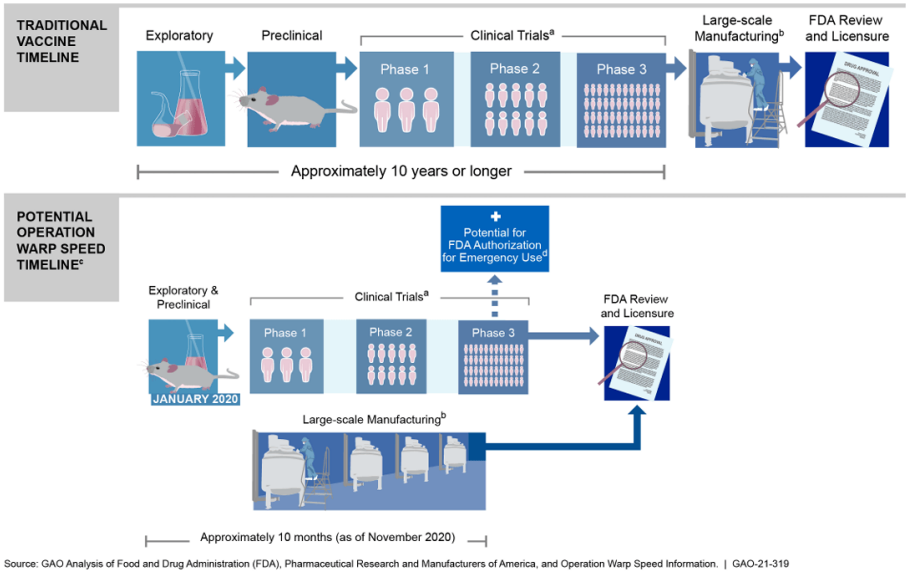


Figure 2.2.2: Traditional Vaccine Development Timeline Compared To Potential Operation Warp Speed Timeline. (Public Domain; General Accountability Office.[8, page 3])

Note: The timelines for vaccine development depicted in Figure 2.2.2 are not drawn to scale. These timelines depict examples, and the specific development steps and timelines for a given vaccine may vary from this example.

Accessibility Note

Figure 2.2.2 depicts a generic traditional vaccine timeline at the top and a potential Operation Warp Speed timeline on the bottom:

- **Traditional Vaccine timeline:** A serial timeline with approximately 10 years spanning Exploratory (depicted by laboratory flasks), Pre-clinical (depicted by a laboratory rat), and Clinical Trials Phases 1-3 (Phase 1 showing a small number of figures denoting people, Phase 2 showing a medium number of people, and Phase 3 with a large number of people), followed by large-scale manufacturing (stylized figure of vaccine manufacturing), and FDA Review and Licensure (depicted by a magnifying glass over a document).
- **Potential Operation Warp Speed Timeline:** A compressed timeline of approximately 10 months labeled by a demarcation at the bottom of the figure spanning Exploratory and Pre-clinical (depicted by a flask and laboratory rat), followed by Clinical Trials Phases 1, 2, and part of 3 (Phase 1 showing a small number of figures denoting people, Phase 2 showing a medium number of people, and Phase 3 with a large number of people), but with an arrow pointing to a box showing the potential for FDA Authorization for Emergency use before the conclusion of Phase 3, and with large-scale manufacturing (stylized figure of multi-station vaccine manufacturing with an arrow going to FDA Review and Licensure) overlapping the clinical trials. This is followed by FDA Review and Licensure (depicted by a magnifying glass over a document).

Footnote details

The footnotes from Figure 2.2.2 are:

- Footnote a:
  - Phase 1 clinical trials generally test the safety of a product with a small group of healthy volunteers (usually fewer than 100). These trials are designed to determine the product's initial safety profile and the side effects associated with increasing doses, among other things.
  - Phase 2 clinical trials are designed to evaluate the effectiveness of a product for a particular use and determine the common short-term side effects and risks associated with the product. These trials are conducted with a medium-size population of volunteers (usually a few dozen to hundreds).
  - Phase 3 clinical trials are performed after preliminary evidence suggesting effectiveness of a product has been obtained, and are intended to gather additional information about safety and effectiveness. These trials usually involve several hundred to thousands of volunteers, including participants who are at increased risk for infection. According to Food and Drug Administration (FDA) officials, these clinical trial phases may overlap.
- Footnote b:
  - According to FDA, manufacturing processes are reviewed as part of the vaccine licensure process. Thus, even under a traditional vaccine timeline, some initial manufacturing occurs during development, so the manufacturing processes can be adequately validated. According to an OWS fact sheet, in some cases, the federal government is taking on the financial risk to enable large-scale manufacturing to start while clinical trials are ongoing, with the goal of having millions of doses available for distribution upon authorization or licensure of a COVID-19 vaccine.
- Footnote c:
  - The Operation Warp Speed timeline depicts an example of a potential accelerated timeline for COVID-19 vaccine development. However, the development process of any given OWS vaccine candidate may vary from this example. As of January 2021, approximately 12 months have elapsed since exploratory and preclinical research began in January 2020, after the first U.S. cases of COVID-19 were reported. The timing for any remaining steps have yet to be determined as of this report. According to OWS documentation, certain steps may overlap or be shortened to accelerate the development of a COVID-19 vaccine.
- Footnote d:
  - During an emergency, as declared by the Secretary of Health and Human Services under 21 U.S.C. § 360bbb-3(b), FDA may temporarily authorize unapproved medical products or unapproved uses of approved medical products through an emergency use authorization (EUA), provided certain statutory criteria are met. FDA has indicated that issuance of an EUA for a COVID-19 vaccine for which there is adequate manufacturing information would require the submission of certain clinical trial information from phase 3 clinical trials that demonstrate the safety and effectiveness of the vaccine

in a clear and compelling manner, among other things. Any COVID-19 vaccine that initially receives an EUA from FDA is expected to ultimately be reviewed and receive licensure through a biologics license application, according to FDA guidance.

Technology Readiness Levels and Vaccine Development

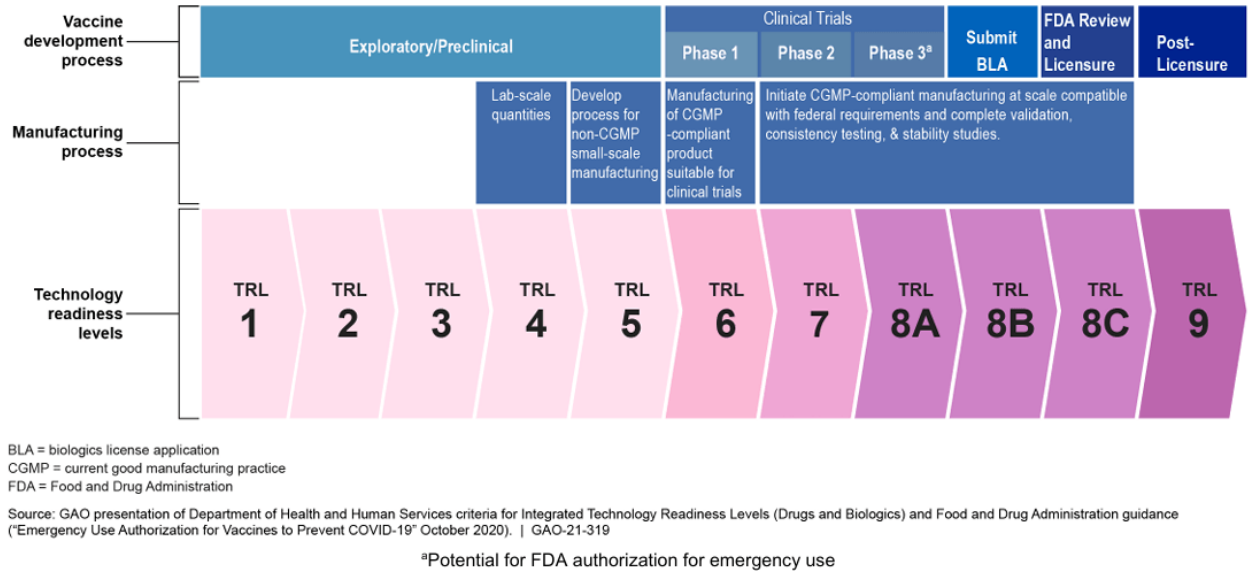


Figure 2.2.3: Department of Health and Human Services' Integrated Technology Readiness Level (TRL) Countermeasure Scale and the Traditional Vaccine Development and Manufacturing Process. ([8,page 10])

Accessibility Note

Figure 2.2.3 depicts the mapping of the vaccine development process and manufacturing process to Technology Readiness Levels (TRLs).

Department of Health and Human Services' Integrated Technology Readiness Level (TRL) Countermeasure Scale and the Traditional Vaccine Development and Manufacturing Process

- **Vaccine development process:** A serial timeline with with a horizontal bar representing Exploratory/Pre-clinical (TRLs 1 through 5), followed by a horizontal bar representing Clinical Trials Phases 1 (TRL 6), Phase 2 (TRL 7), and Phase 3 (TRL 8A), followed by a horizontal bar representing submission of the BLA (biologics license application) (TRL 8B), followed by a horizontal bar representing FDA Review and Licensure (TRL 8C), and followed by a horizontal bar representing Post-Licensure (TRL 9).
- **Manufacturing process:** shows a horizontal bar representing lab scale quantities (TRL 4), followed by a horizontal bar representing "Develop process for non-CGMP small-scale manufacturing" (TRL 5), in parallel with the end of Exploratory/ Pre-clinical in the Vaccine development process. This is followed by a horizontal bar representing "Manufacturing of CGMP-compliant product suitable for clinical trials":, overlapped with Clinical Trials Phase 1 (TRL 6). This is followed by by a horizontal bar representing "Initiate CGMP-compliant manufacturing at scale compatible with federal requirements and complete validation, consistency testing, and stability studies", covering Clinical Trials Phases 2 and 3, Submit BLA, and FDA Review and Licensure, covered by TRL 7, TRL 8A, TRL 8B, and TRL 8C.
- **Technology Readiness Levels:** Horizontal bars for each of the following: TRL 1, TRL 2, TRL 3, TRL 4, TRL 5, TRL 6, TRL 7, TRL 8A, TRL 8B, TRL 8C, and TRL 9. The mapping to the Vaccine development process and Manufacturing process is described above. Technology Readiness Levels (TRLs) in drug and vaccine development range from basic research (TRL 1) to regulatory approval and market launch (TRL 9). The TRLs provide a framework for assessing the maturity of a drug or vaccine candidate.

References

1. Wikipedia - Operation Warp Speed. August 24, 2020.
2. HHS Operation Warp Speed. U.S. Department of Health and Human Services. Explaining Operation Warp Speed. Accessed August 24, 2020.
3. Wikipedia - Operation Warp Speed. August 23, 2020.
4. Wikipedia - Operation Warp Speed. August 24, 2020.
5. History of Operation Warp Speed. Accessed February 1, 2026.
6. HHS Operation Warp Speed. U.S. Department of Health and Human Services. Explaining Operation Warp Speed. Accessed February 1, 2026.
7. GAO-21-319. United States Government Accountability Office. OPERATION WARP SPEED Accelerated COVID 19 Vaccine Development Status and Efforts to Address Manufacturing Challenges. Publication number GAO-21-319. February 2021. Accessed February 1, 2026.
8. GAO-21-319. United States Government Accountability Office. OPERATION WARP SPEED Accelerated COVID 19 Vaccine Development Status and Efforts to Address Manufacturing Challenges. Publication number GAO-21-319. February 2021. Accessed February 1, 2026.
9. Wikipedia - Operation Warp Speed. Accessed February 2, 2026.
10. Operation Warp Speed seal. U.S. Department of Defense. *This file is a work of a U.S. military or Department of Defense employee, taken or made as part of that person's official duties. As a work of the U.S. federal government, the file is in the* ***public domain*** *in the United States.*

This page titled 2.2: Operation Warp Speed is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

2.3: COVID-19 Vaccine Development

The development of a COVID-19 vaccine was crucial to mitigating the public health and economic impacts of the virus. By the end of March 2020, with the initiation of the first clinical trials, the race was on in the United States to develop a vaccine. On December 14, 2020, the United States took an important step to protect the public against the virus as the first vaccine shots, developed in a shorter time than any previous vaccine, were administered.[2]

Vaccines provide protection for individuals and communities, to lower transmission and disease burden once a large enough portion of the population, typically 70 to 90 percent develops immunity. Reaching this "**herd immunity** threshold" limits the likelihood that a non-immune person will be infected. Herd immunity helps protect people who are not immune to a disease by reducing their chances of interacting with an infected individual, thereby slowing or stopping the spread of the disease. Achieving herd immunity can require a high rate of vaccination in the community, and can bring about a safe return to use of restaurants, theaters, and gyms, and the resumption of community-based activities. In this way, vaccines can save lives, reduce the sometimes debilitating effects of COVID-19, and contribute to the restoration of the economy.[2,p4-5]

Let's now revisit the traditional vaccine development timeline compared to the aggressive timeline of Operation Warp Speed, as discussed briefy in the prior section.

Traditional Vaccine Timeline: Approximately 10 years or longer[2]

The traditional processes include:

- Exploratory
- Preclinical
- Clinical Trials
  - Phase 1
  - Phase 2
  - Phase 3
- Large-scale Manufacturing
- FDA Review and Licensure

In the exploratory phase, the target and candidate vaccine are identified.

In the preclinical phase, researchers use cells and animals to assess safety and produce evidence of clinical promise, evaluated by the candidate's ability to elicit a protective immune response.

During clinical trials, more human subjects are added at each successive phase. Safety, efficacy, proposed doses, schedule of immunizations, and method of delivery are evaluated.

The next phase is FDA review of the biologics license application (BLA) and licensure, which includes oversight of manufacturing and planning for post-market surveillance. At any phase, the process can be terminated for various reasons including detection of adverse events, such as serious side effects.

Potential Operation Warp Speed Timeline Approximately 10 months

The processes include:

- Exploratory & Preclinical (January 2020)
- Clinical Trials
  - Phase 1
  - Phase 2
  - Phase 3: Potential for FDA Authorization for Emergency Use
- Large-scale Manufacturing, overlapped with clinical trials
- FDA Review and Licensure

Operation Warp Speed Vaccine Platform Technologies[2]

Data table for Vaccine Platform Technologies Supported by Operation Warp Speed, as of January 2021

	mRNA platform	Replication-defective live-vector platform	Recombinant-subunit-adjuvanted protein platform
Description	Encapsulated genetic instructions that allow vaccinated individuals to produce the spike protein of SARS-CoV-2 to stimulate immune system but cannot cause COVID-19.	Non-replicating virus that delivers genetic instructions to allow vaccinated individuals to produce the spike protein of SARS-CoV-2 to stimulate immune system but cannot cause COVID-19.	Fully-formed spike protein of SARS-CoV-2 delivered with adjuvant, which helps to stimulate immune system of vaccinated individuals but cannot cause COVID-19.
Operation Ward Speed candidates (most advanced clinical trial phase)	Moderna (phase 3) Pfizer/BioNTech (phase 3)	Janssen (phase 3) AstraZeneca (phase 3)	Sanofi/GSK (phase 2) Novavax (phase 3)

Figure 2.3.1: Operation Warp Speed Vaccine Platform Technologies. (GAO [2])

Accessibility Note

Figure 2.3.1 describes the Operation Warp Speed Vaccine Platform Technologies:

1. mRNA,
2. replication defective live-vector, and
3. recombinant-subunit-adjuvanted protein.

These are described in detail in subsequent sections.

Vaccine Platform Technologies: Mechanisms

Vaccine platform technologies use different mechanisms for stimulating immune responses, Let's take a quick look at those mechanisms.

Human cells use lock-and-key-style security to allow for the necessary exchange of proteins while preventing the intrusion of disease-causing microbes, such as viruses. Before entering a cell, a protein needs to present a unique ‘key’, a molecular pattern that opens a specific ‘lock.’ Coronaviruses, such as SARS-CoV-2, use counterfeit keys, called spike “S” proteins, to enter human cells. All COVID-19 vaccines share a common strategy: teach the immune system to recognize the SARS-CoV 2 spike protein and neutralize the virus, providing immunity. The immune system response that neutralizes the virus is largely mediated by antibody production and associated immune cells (e.g., T cells). The OWS vaccine candidates differ in what method, or platform, they use to initiate these immune responses.

There are three main platforms:

1. mRNA,
2. replication defective live-vector, and
3. recombinant-subunit-adjuvanted protein.

These three vaccine platforms, unlike other vaccine platforms, do not require researchers to grow the SARS-CoV-2 virus, which has sped the time of development and avoided safety concerns associated with using a disease-causing virus.

Table 2: Characteristics for Each Operation Warp Speed Vaccine Candidate, as of January 2021

Vaccine identifier	Candidate company	Vaccine platform	Doses	Dose spacing (days)	Mixing required[a]	Storage temperature[b]
BNT162b2	Pfizer/BioNTech	mRNA[c]	2	21	Yes	2 to 8°C up to 5 days Longer periods -80 to -60°C
mRNA-1273	Moderna	mRNA[c]	2	28	No	2 to 8°C up to 30 days Longer periods -25 to -15°C
Ad26.COV2.S	Janssen	Replication-defective live-vector	1	N/A	No	At least 3 months at 2 to 8°C Up to 2 years at -20°C
AZD1222	AstraZeneca	Replication-defective live-vector	2	28	No	2 to 8°C up to 6 months
VAT01	Sanofi/GSK	Recombinant-subunit-adjuvanted protein	2	21	Yes	2 to 8°C
NVX-CoV2373	Novavax	Recombinant-subunit-adjuvanted protein	2	21	No	2 to 8°C

Source: GAO analysis of information from vaccine companies, Food and Drug Administration (FDA), and other government sources. | GAO-21-319.

[a]Mixing required means the addition of an adjuvant (to increase immune response) or diluent (to dilute the vaccine) is required at the time of vaccine administration.

[b]Storage temperature is the recommended temperature range for vaccine storage.

[c]Messenger RNA (mRNA)

Figure 2.3.2: Characteristics for Each Operation Warp Speed Vaccine Candidate, as of January 2021. (GAO [2,page 19)

Accessibility Note

Figure 2.3.2 describes the characteristics for each Operation Warp Speed vaccine candidate, including vaccine ID, company, platform, doses, storage temperature. These are described in sebsequent sections.

mRNA platform

The Moderna and Pfizer/BioNTech mRNA vaccines deliver the genetic sequence of the SARS-CoV-2 spike protein directly to the cell.

The mRNA molecule includes a code that causes the cell to make the spike protein. Immune system cells recognize the spike protein and a protective immune response results. The spike protein genetic code does not enter the cell's nucleus, only the cytoplasm.

The mRNA needs to be encased in a lipid (fat) nanoparticle to enter the cell.

- **Pfizer/BioNTech:** Pfizer/BioNTech's mRNA vaccine, BNT162b2, consists of mRNA encoding the viral spike protein of SARS-CoV-2, transported inside lipid nanoparticles that allow the mRNA to enter cells. The vaccine remains shelf-stable in an ultra-low temperature freezer between -80°C to -60°C. Vials must be kept frozen between 80°C to -60°C and protected from light until ready to use. The vaccine remains shelf-stable for up to five days at standard refrigerator temperatures (between 2° and 8°C).
- **Moderna:** Moderna's mRNA-1273 vaccine also consists of mRNA encoding the viral spike protein of the SARS-CoV-2 virus transported in lipid nanoparticles. The Moderna vaccine can be stored at refrigerator temperatures (between 2° and 8°C) for 30 days, and it is stable for 6 months during shipping and long-term storage at freezer temperatures of -20°C.

The process for mRNA vaccines generating antibodies to protect individuals from COVID-19:

1. Key genetic information for SARS-COV-2 virus
2. Lipid nanoparticle containing mRNA coding the SARS-CoV-2 spike protein
3. SARS-CoV-2 spike proteins produced by vaccinated individual
4. Generation of antibodies against SARS-CoV-2 within vaccinated individual.

Replication-defective live-vector platform

The Janssen and AstraZeneca vaccine candidates use a weakened adenovirus, a virus that can cause the common cold but that

is altered so that it cannot reproduce or cause disease.

Known as a viral vector, it carries a DNA code to make the SARS-CoV-2 spike protein that will stimulate the immune system to produce antibodies. The vector interacts with the target cell and delivers its genetic material into the nucleus, where cellular enzymes generate the spike protein, but not the adenovirus itself. The vaccinated person will produce the spike protein, priming their immune system to target SARS-CoV-2.

- **Janssen:** Janssen's vaccine candidate uses a non-replicating human Adenovirus 26 vector platform, the same platform Janssen used to develop a vaccine for Ebola. This virus, which normally causes the common cold, contains the genetic material of the SARS-CoV-2 spike protein. This vaccine candidate can be stored between 2 and 8°C for at least three months.
- **AstraZeneca:** AstraZeneca's AZD1222 vaccine candidate consists of a non-replicating chimpanzee adenovirus, ChAdOx1, which is a weakened version of the virus that causes infections in non-human primates. This vaccine candidate can be stored between 2 and 8°C.

The process for replication-defective live-vector vaccines generating antibodies to protect individuals from COVID-19:

1. Key genetic information from SARS-CoV-2 virus
2. Viral vector containing gene coding the SARS-CoV-2 spike protein
3. SARS-CoV-2 spike proteins produced by vaccinated individual
4. Generation of antibodies against SARS-CoV-2 within vaccinated individual.

Recombinant-subunit-adjuvanted protein platform

The Sanofi/GSK and Novavax vaccine candidates use purified SARS-CoV-2 spike proteins to stimulate an immune response. Often, recombinant-subunit adjuvanted protein platforms require an adjuvant, a component of the vaccine that helps the immune system response.

Examples of the vaccines produced using this platform include Hepatitis B, human papilloma virus, and tetanus vaccines.

- **Sanofi/GSK:** Sanofi/GSK's vaccine candidate, developed in partnership by Sanofi and GSK, uses the same recombinant protein based technology as one of Sanofi's seasonal influenza vaccines with GSK's established pandemic adjuvant technology. This vaccine candidate can be stored between 2° and 8° C.
- **Novavax:** Novavax's NVX-CoV2373 vaccine candidate is a recombinant nanoparticle spike protein vaccine candidate that includes a proprietary adjuvant to increase the immune response. It can be stored between 2° and 8°C.

The Process for Recombinant-subunit-adjuvanted Protein Vaccines Generating Antibodies to Protect Individuals from COVID-19:

1. Key genetic information from SARS-CoV-2 virus
2. Purified SARS-CoV-2 spike proteins with adjuvant
3. Generation of antibodies against SARS-CoV-2 within vaccinated individual.

Vaccine Clinical Trials

As of January 2021, five of the six Operation Warp Speed vaccine candidates had begun phase 3 clinical trials in the United States, and two had received an EUA. In November 2020, the GAO reviewed four of the vaccine candidates' clinical trial protocols.

The GAO found that they generally appeared to follow a typical clinical trial design by enrolling mostly healthy adults and excluding such groups as children, pregnant women, and those with certain comorbid or unstable conditions. Excluding these groups in the initial phase 3 clinical trials is not unusual, but a potential consequence is that the data on vaccine safety and effectiveness is based on mostly healthy adults and may not apply to these excluded populations.[2, p20]

As of January 2021, the vaccines that had received EUAs had not been licensed by FDA and continue to be studied in clinical trials. For vaccines that received EUA, additional data on vaccine effectiveness will be generated from further follow-up of participants in clinical trials already underway before the EUA was issued.

The two EUAs issued in December 2020 were based on analyses of clinical trial participants and showed about 95 percent efficacy for each vaccine. These analyses included assessments of efficacy after individuals were given two doses of vaccine and from follow-up for a median duration of 2 months to monitor for adverse events. For the two vaccines that received EUAs, Pfizer/

BioNTech and Moderna, phase 3 clinical trials did not study the efficacy of a single dose regimen in a manner that allowed for definitive conclusions, according to FDA.

According to FDA, a BLA typically includes safety data from the entire study population through at least 6 months of follow-up following the last vaccination, though most adverse events are observed within 1.5 months of vaccine administration. In clinical trials for the Pfizer/BioNTech and Moderna vaccines, participants reported side effects such as pain at the injection site, fatigue, and headache. These side effects are not unusual for vaccines. The initial phase 3 clinical trials for these vaccines excluded people with a history of severe adverse reactions to any vaccine or allergic reactions to any component of this vaccine.

Traditional Practices with Adaptations

COVID-19 vaccine development generally followed traditional practices, with some adaptations.

The HHS integrated TRLs generally reflect the traditional process used for vaccine development. Therefore, they provide the measurement standard by which we can assess the development process to understand where and how the process may have been modified for COVID-19 vaccines. In an effort to understand the readiness of each OWS vaccine candidate, the GAO conducted a TRL analysis, which showed that vaccine companies generally followed the traditional development process.

After reviewing vaccine companies' questionnaire responses and supporting information, the GAO assigned TRLs to each vaccine candidate based on the HHS integrated TRLs using associated FDA guidance and supporting documentation for each vaccine candidate.

Table 3: Technology Readiness Levels (TRL) for Each Operation Warp Speed (OWS) Vaccine Candidate, as of January 2021

Vaccine Company	TRL	Description
Pfizer/BioNTech	8A	Vaccine has achieved completion of current good manufacturing practice (CGMP) validation and consistency lot manufacturing, and pivotal clinical trials demonstrating sufficient efficacy and safety to receive an emergency use authorization (EUA).
Moderna	8A[a]	Vaccine has achieved completion of CGMP validation and consistency lot manufacturing, and pivotal clinical trials demonstrating sufficient efficacy and safety to receive an EUA.
AstraZeneca	7B	Vaccine has achieved scale-up, initiation of CGMP process validation, and expanded clinical trials as appropriate for the product.
Janssen	7B[a]	Vaccine has achieved scale-up, initiation of CGMP process validation, and expanded clinical trials as appropriate for the product.
Novavax	6C[ab]	Vaccine has achieved CGMP pilot lot production, investigational new drug (IND) application submission, and phase 2 clinical trials that establish an initial safety, pharmacokinetics, and immunogenicity assessment as appropriate.
Sanofi/GSK	6C[a]	Vaccine has achieved CGMP pilot lot production, IND application submission, and phase 1 clinical trials that establish an initial safety, pharmacokinetics, and immunogenicity assessment as appropriate.

Source: GAO analysis of vaccine companies' questionnaire responses and vaccine development documentation. I GAO-21-319

Figure 2.3.3: TRL for Vaccine Candidates. (GAO [2,[p24])

Note: GAO assigned these TRLs based on questionnaire responses and documentation, where available.

CGMP regulations for drugs and biologics, including vaccines, contain minimum requirements for the methods, facilities, and controls used in manufacturing, processing, and packing of a drug product. The regulations help to ensure that a product is safe for use, and that it has the ingredients and strength it claims to have. See 21 C.F.R. pts. 210 and 211 (2020).

An IND is a formal notice to the Food and Drug Administration (FDA) of a company's intent to begin human clinical trials. An IND must include evidence that the product is reasonably safe for proposed clinical trials, based on preclinical data, among other information. FDA has 30 days to object to an IND before it becomes effective. 21 C.F.R. pt. 312 (2020).

TRL 8A indicates that additional expanded clinical safety trials may be required for the product. In the case of a COVID-19 vaccine, the GAO assigned a TRL 8A when an emergency use authorization (EUA) was issued for that product. During an emergency, as declared by the Secretary of Health and Human Services under 21 U.S.C. § 360bbb-3(b), FDA may temporarily authorize unapproved medical products or unapproved uses of approved medical products through an EUA, provided certain statutory criteria are met. FDA has indicated that issuance of an EUA for a COVID-19 vaccine for which there is adequate

manufacturing information would require the submission of certain clinical trial information from phase 3 clinical trials that demonstrate the safety and effectiveness of the vaccine in a clear and compelling manner, among other things. Any COVID-19 vaccine that initially receives an EUA from FDA is expected to ultimately be reviewed and receive licensure through a biologics license application, according to FDA guidance.

Footnote a: TRL based at least in part on testimonial evidence; GAO could not verify all information supporting our TRL determination through documentary evidence.

Footnote b: Although Novavax is currently conducting phase 3 clinical trials, they reported that as of February 2, 2021, they had not completed the step to scale up and validate the CGMP manufacturing process, and therefore did not yet meet the criteria for TRL 7.

Adaptations to accelerate development

The primary difference between COVID-19 vaccine development and vaccine development in a non-pandemic environment was the compressed timelines under which the vaccine companies were working.

In addition, to speed up the availability of the vaccines, companies initiated large-scale manufacturing while collecting data on clinical trial participants.

In a June 2020 guidance document, FDA identified some ways that COVID-19 vaccine development may be accelerated. For example, the guidance document states that companies may accelerate development by relying on knowledge gained from similar products manufactured with the same well-characterized platform technology, to the extent legally and scientifically permissible. All OWS vaccine companies indicated that prior experience on the vaccine platform helped support key steps that would normally be conducted for each individual vaccine.[2,p25]

At least half of the OWS vaccine companies indicated that they had not completed certain animal safety and efficacy studies before beginning phase 1 clinical trials. Instead, in order to begin collecting data in clinical trials more quickly, the companies relied on data from other vaccines using the same platforms, where available, or conducted animal studies concurrently with clinical trials. As of January 2021, some animal studies are still ongoing for COVID-19 vaccines that are in late-stage (e.g., phase 3) clinical trials.[2,p26]

Concurrent or combined clinical trials

OWS vaccine companies conducted concurrent or combined clinical trials.[2,p26-27]

By conducting different phases of clinical trials concurrently (e.g., **phase 3 clinical trials beginning as phase 1 trials are ongoing**), OWS vaccine companies increased the speed of the vaccine development process. According to HHS, working on clinical trial phases in parallel instead of taking the traditional sequential approach to vaccine development potentially shaves months off the timeline for vaccine development.

One company noted that using efficient clinical trial strategies, such as concurrent or overlapping trials, is particularly important to quickly determine disease protection (i.e., vaccine efficacy) in a pandemic.

For instance, this approach was successfully used during the Ebola epidemic in Africa where vaccine efficacy was assessed while the epidemic was still ongoing. Though some overlap of phases is not unusual even in traditional vaccine development, officials from two companies stated that in non-pandemic environments it can take months to review clinical trial data before starting a new phase. For example, officials from one company said they might normally take 6 months to review data from phase 2 trials before initiating a phase 3 trial, and they would have a meeting with FDA about the plan before proceeding.

For COVID-19 vaccine development, officials from this particular company said they took 3 weeks to review data and initiate efforts to move to phase 3 trials. All six OWS vaccine companies gathered initial human safety and immunogenicity data in phase 1 or combined phase 1/2 clinical trials with a small number of participants before proceeding into trials with more participants, namely large-scale phase 3 clinical trials, consistent with traditional processes. All companies that have started phase 3 clinical trials as of January 2021 did so before completing phase 1 clinical trials.

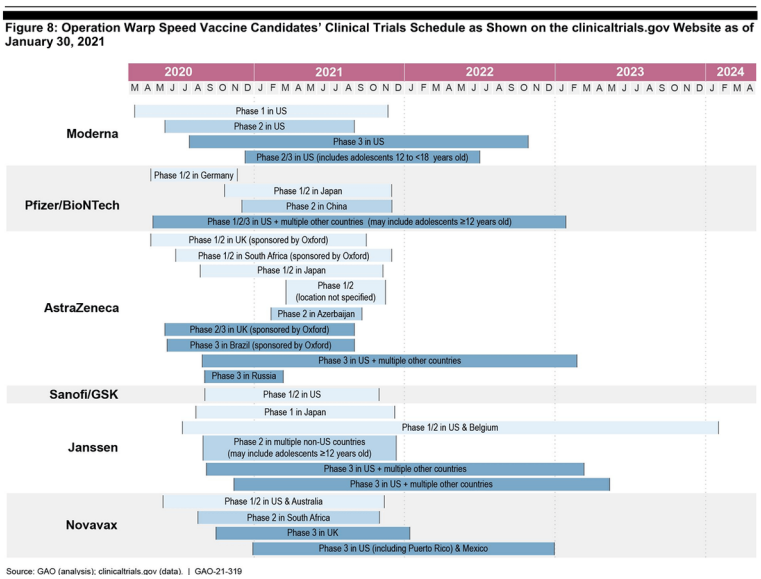


Figure 2.3.4: Vaccine clinical trials schedule summaries. (GAO [2,p28])

Discussion of Figure 2.3.4

The following is a synopsis of the Clinical trial phase overlapping as employed by various pharmaceutical companies from a project management perspective:

- Moderna: Phase 2 started and completed during Phase 1 in the Unites States. Phase 3 in the United States started while Phase 1 and Phase 2 were in progress.
- Pfizer/BioNTech: Conducted Phase 1 and Phase 2 in Germany, and started Phase 1/Phase 2 in Japan toward the end of Phase 1/2 in Germany. It initiated Phase 2 in China shortly after initiating Phase 1/Phase 2 in Japan. It also initiated Phase 1/2/3 in the US and other countries at the same time as it started the Phase 1/2 in Germany.
- AstraZeneca: After initiating Phase 1/2 in the UK with Oxford, it initiated an overlapping Phase 1/2 in South Africa, also with Oxford. While the Phase 1/2 in South Africa was in progress, it initiated a Phase 1/2 in Japan and a Phase 2 in Azerbaijan. A Phase 2/3 in the UK overlapped with the Phase 1/2 in the UK, as did a Phase 3 in Brazil, followed by the initiation of a Phase 3 in the United States and other countries, including Russia.
- Sanofi/GSK: Conducted a Phase 1/2 in the United States.
- Janssen: Conducted a Phase 1/2 in the United States and Belgium, with an overlapping Phase 1 in Japan, and a Phase 2 in multiple non-US countries,
- Novavax: It initiated a Phase 1/2 in the United States and Australia, followed by an overlapping Phase 2 in South Africa, which in turn overlapped with a Phase 3 in the UK and later by a overlapping Phase 3 in the United States.

At least half of the OWS vaccine companies selected a dose for phase 1 clinical trials that was based in part on disease protection data generated in animal studies for that vaccine candidate or from studies for other vaccines using the same vaccine platform. One company noted that the public health emergency precluded them from taking the time to determine the minimum effective dose. Instead, they focused on a dose that resulted in an acceptable tolerability and immunogenicity profile and with the greatest chance of efficacy. This company recognized that they might end up delivering a higher dose than is necessary in the short-term, but indicated they could explore a minimal dose that may be as effective, but more efficient, at a later time.[2,p29]

References

1. Publication GAO-21-319. United States Government Accountability Office. OPERATION WARP SPEED Accelerated COVID 19 Vaccine Development Status and Efforts to Address Manufacturing Challenges. Publication GAO-21-319 Accessible Version. Accessed February 1, 2026.
2. Publication GAO-21-319. United States Government Accountability Office. OPERATION WARP SPEED Accelerated COVID 19 Vaccine Development Status and Efforts to Address Manufacturing Challenges. Publication GAO-21-319 Accessible Version. Accessed February 2, 2026.

This page titled 2.3: COVID-19 Vaccine Development is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

2.4: Drug Development for Traditional Drugs and Biologics

Learning Objectives:

- Explain the phases of the drug development process.
- Compare and contrast the drug development process between traditional drugs and biologics.
- Identify the role of the Food and Drug Administration (FDA) in ensuring that drugs are safe and effective.
- Explain why the FDA may expedite drug approval or encourage the development of certain drugs.
- Describe the role of the health care team in post-marketing surveillance.

Introduction to the Drug Development Process

In the United States, the drug development process occurs within a structured framework to determine that a candidate drug has proven benefits that outweigh known and potential risks for the intended patient population. This process includes several stakeholders, including the pharmaceutical company (sponsor) developing the candidate drug and the Food and Drug Administration (FDA). The Center for Drug Evaluation and Research (CDER) is the FDA division responsible for regulating the approval and use of drug products as defined by the Federal Food, Drug, and Cosmetic (FD&C) Act, including prescription and over-the-counter medications, certain biological therapeutics, and generic drugs.[1,2] In addition, CDER considers products such as fluoride toothpaste, antiperspirants, dandruff shampoos, and sunscreens to be drugs and, therefore, is also responsible for regulating their approval and use.[2] The Center for Biologics Evaluation and Research (CBER) is the FDA division responsible for the regulation of biological products under the authority of the Public Health Services (PHS) Act and the FD&C Act.[3-5] Examples of products regulated by CDER and CBER are provided in Table 2.4.1.

Table 2.4.1 CDER and CBER Product Regulatory Responsibility[4,5]

CDER Responsible	CBER Responsible
Naturally occurring substances from mineral or plant source materials (excluding vaccines or allergenics)	Vaccines and vaccine-associated products
Products produced from nonhuman animal or human solid tissue sources	In vivo diagnostic allergenic products
Chemically synthesized molecules, including those that bind to the receptors for biological products and mononucleotide or polynucleotide products (e.g., cytokines, thrombolytics, monoclonal antibodies, immunomodulators)	Human blood or human blood-derived products, animal-derived procoagulant products, and animal- or cell culture-derived hemoglobin-based products (e.g., red blood cell substitutes)
Antibiotics	Immunoglobulin products
Hormone products (e.g., insulin, human growth hormone, pituitary hormones)	Products composed of or intended to contain intact cells or intact microorganisms, including bacteria, fungi, viruses, or virus pseudotypes, or viral vectors (e.g., gene therapy products)

CDER Responsible	CBER Responsible
Certain substances produced by fungi or bacteria (e.g., disaccharidase inhibitors, HMG-Co-A reductase inhibitors)	Protein, peptide (40 or fewer amino acids), or carbohydrate products produced by cell culture, except antibiotics, hormones, and products previously derived from human or animal tissue and regulated as approved drugs
	Animal venoms or constituents of venoms
	Synthetically-produced allergenic products are intended to specifically alter the immune response to a specific antigen or allergen
	Certain drugs used in conjunction with blood banking and/or transfusion

Accessibility Note

Table 2.4.1 shows that Product Regulatory Responsibility lies with the CBER for vaccines and with the CDER for substances from mineral or plant sources (excluding vaccines).

The drug development process includes 5 phases:

1. discovery and development,
2. preclinical research,
3. clinical research,
4. FDA review, and
5. post-marketing safety monitoring.

For most drugs, the typical duration from discovery to FDA approval is between 10 and 15 years (Figure 2.4.1) and may cost more than $1 billion.[6] Each of these phases is described in more detail later in this section.

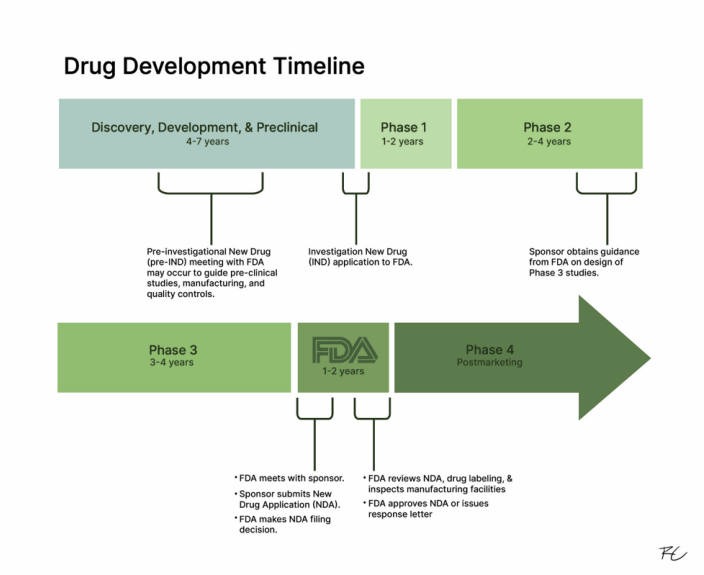


Figure 2.4.2: Drug Development Timeline. (CC-BY 4.0; Riley Cutler)

Accessibility Note

Figure 2.4.2: Drug Development Timeline displays a timeline for drug development showing the number of years for each of the phases:

- Discovery, Development, and Pre-clinical: 4-7 years. After this, an Investigational New Drug (IND) application goes to the FDA.
- Phase 1: 1-2 years
- Phase 2: 2-4 years
- Phase 3: 3-4 years
- FDA processing: 1-2 years
- Phase 4: Post-marketing.

Discovery, Development, and Preclinical Research

The discovery, development, and preclinical research phases (Figure 2.4.2) encompass the drug development work before investigational drugs are tested in humans. This process may take 4 – 7 years to complete, with several steps occurring simultaneously (Figure 2.4.2).[7]

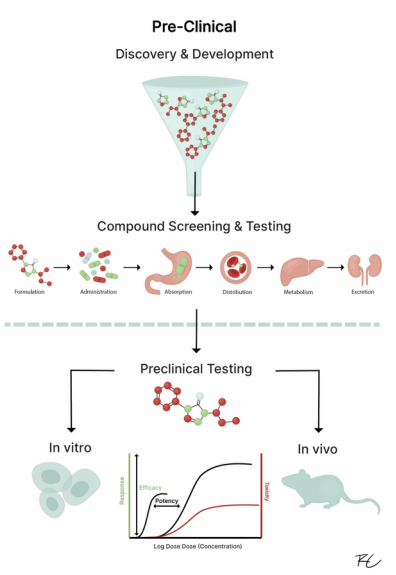


Figure 2.4.3: Preclinical Research. (CC-BY 4.0; Riley Cutler)

Accessibility Note

Figure 2.4.3 depicts the Pre-clinical phase of drug development:

- Discovery and Development (depicted by a funnel showing various compounds)
- Compound Screening and Development
  - Formulation (depicted by a compound)
  - Administration (depicted by pills)
  - Absorption (depicted by a stomach)
  - Distribution (depicted by cellular transport)
  - Metabolism (depicted by a liver)
  - Excretion (depicted by kidneys)
- Pre-clinical testing
  - In vitro (depicted by cells)
  - In vivo (depicted by a lab rat)
  - Chart showing dose efficacy, potency, toxicity.

Discovery and Development

During the discovery phase, scientists identify disease-related mechanisms and processes (e.g., disease pathways, genes, proteins) that may serve as targets for developing drugs to stop, reverse, or prevent the disease of interest. These discoveries are made in a laboratory setting.

Once a disease-related target has been identified, hundreds to thousands of chemical and/or biological compounds are screened using *in silico*, or computational, testing to evaluate their ability to interact with the disease-related target and potentially treat or prevent the disease of interest.[7,8] The goal of this process is to identify a small number of compounds—and ultimately a lead compound—that offer the most promise for further development. Promising compounds are further evaluated using *in silico* and *in vitro* models to predict physicochemical parameters, pharmacokinetics, and drug metabolism.[8,9]

Once a lead compound is identified, the preclinical development includes analytical and bioanalytical methods development and validation, dosage form design, and establishment of good manufacturing processes for the active ingredient (e.g., lead compound) and final investigational drug product to be used in clinical research phases.[8,9] Additionally, *in vivo* studies are performed on the investigational drug using one or more animal models to evaluate basic pharmacology (e.g., metabolism, pharmacokinetics), perform dose range-finding, and assess experimental toxicology.[9]

Preclinical Research

Before an investigational drug can be tested in humans, definitive, *in vivo* preclinical studies in one or more animal models are required to assess the safety and toxicology (LD_{50}) of candidate drugs using good laboratory practices (GLP), good manufacturing practice (GMP) in the dosage form and maximally tolerated dose from dose range-finding studies.[9,10] Additional *in vitro* studies are also performed to assess organ system toxicity (e.g., central nervous, cardiovascular, respiratory, and immune systems) and genetic toxicology (e.g., DNA, cellular genetic processes).[9] Investigational drugs that exhibit potentially beneficial pharmacological activity to treat the disease of interest and have demonstrated reasonable safety in preclinical research may proceed to the clinical research phase once the sponsor receives approval for an **investigational new drug (IND)** application (described later in this section). Only about 1 in 1,000 potential drugs evaluated in the discovery, development, and preclinical phase proceed to the clinical research phase.[6]

Clinical Research

Once the FDA approves an IND application, a three-step clinical research phase begins, during which the investigational drug is tested in humans. The goal of the clinical research phase is to establish the safety, efficacy, and effectiveness of new drugs in humans.

Phase 1

The first step of the clinical research phase is to conduct Phase 1 studies (Figure 2.4.4:). The primary focus of Phase 1 studies is to determine a safe dosage range and administration method and evaluate the investigational drug's toxicity profile. Part of the determination of a safe dosage range includes evaluations of the drug's pharmacokinetic properties.

Phase 1 studies are typically conducted in a relatively small number (20 – 80 subjects) of healthy volunteers. However, Phase 1 studies evaluating drugs to treat cancer are typically conducted in patients with the targeted type of cancer.[1,7,11] Phase 1 studies may take 1 – 2 years or more to complete (Figure 2.4.4:) and inform the design of Phase 2 studies.[7,11,12]

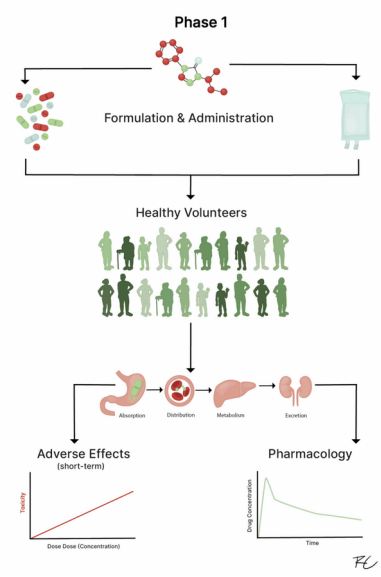


Figure 2.4.4: Phase 1 Studies. (CC-BY 4.0; Riley Cutler)

Accessibility Note

Figure 2.4.4 depicts Phase 1 Studies:

- Formulation & Administration
- Healthy volunteers (depicted by a small number of figures of people)
- Testing
  - Absorption (depicted by a stomach)
  - Distribution (depicted by cellular transport)
  - Metabolism (depicted by a liver)
  - Excretion (depicted by kidneys)
- Adverse effects (depicted by a chart plotting Toxicity versus Dose)
- Pharmacology (depicted by a chart plotting Drug Concentration versus Time).

Phase 2

Approximately 40 – 70% of investigational drugs tested in Phase 1 demonstrate a reasonable safety profile and move on to Phase 2 studies (Figure 2.4.5).[11,12] The purpose of Phase 2 studies is to evaluate the efficacy of the investigational drug to treat the condition of interest while gaining additional safety data. Phase 2 studies often incorporate dose-ranging evaluations to explore the investigational drug's therapeutic range.[6,7,11]

Because Phase 2 studies are designed to evaluate efficacy, they are conducted in patients with the condition of interest.

Note

This is not the case with vaccines. The statement applies to therapeutics.

Furthermore, the number of patients studied in Phase 2 studies is larger (100 – 300) compared to Phase 1 studies. Phase 2 studies are generally randomized, controlled studies utilizing a placebo or alternative treatment (e.g., standard of care) and study patients for up to 2 years.[11] Efficacy endpoints often relate to symptom relief, pharmacodynamic/physiologic changes expected based on the pharmacology (e.g., blood pressure), and/or biomarker changes indicative of alterations in the disease course. Phase 2 studies serve as a "proof-of-concept" to inform the design of Phase 3 studies.[7] Overall, it may take 2 – 4 years for Phase 2 studies to be completed (Figure 2.4.5).[6,11,12]

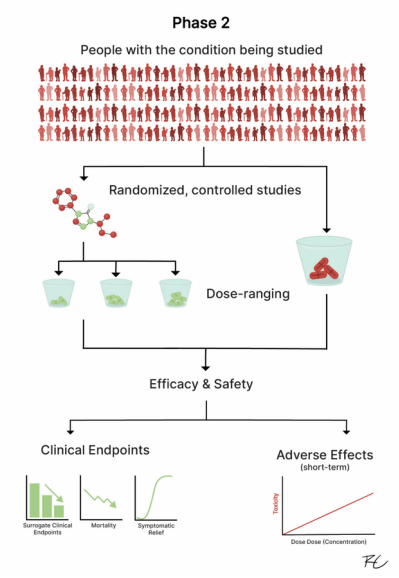


Figure 2.4.5: Phase 2 Studies. (CC-BY 4.0; Riley Cutler)

Accessibility Note

Figure 2.4.5 depicts Phase 2 Studies:

- Testing (depicted by a moderate number of people)
- Randomized controlled studies (depicted by containers with different dose amounts)
- Efficacy and Safety
  - Clinical Endpoints (depicted by three charts: Surrogate Clinical Endpoints, Mortality, and Symptomatic Relief)
  - Adverse Effects (depicted by a chart plotting Toxicity versus Dose)

Phase 3

Approximately 30% of investigational drugs tested in Phase 2 demonstrate efficacy and a reasonable safety profile, enabling them to proceed to Phase 3 studies.[11,12] The purpose of Phase 3 studies is to confirm that the investigational drug is both effective and safe in patients with the disease of interest.

To accomplish these goals, Phase 3 studies include much larger (300–3,000 or more) cohorts of patients with the disease of interest and longer durations of follow-up, providing a better assessment of serious, long-term, and/or rare adverse effects.[6,7,11] Like Phase 2 studies, Phase 3 studies compare the investigational drug to a control group, typically a placebo. Efficacy endpoints may also include more robust clinical endpoints such as disease-specific adverse events, including mortality. Phase 3 often takes 1–4 years to complete (Figure 2.4.6).[6,7,11,12]

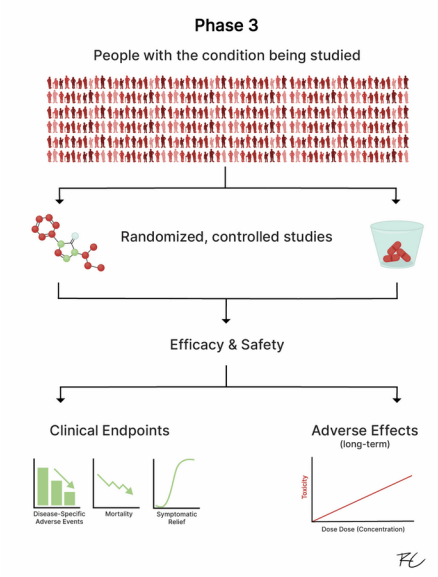


Figure 2.4.6: Phase 3 Studies. (CC-BY 4.0; Riley Cutler)

Accessibility Note

Figure 2.4.6 depicts Phase 3 Studies:

- Testing (depicted by a large number of people)
- Randomized controlled studies (depicted by a chemical and container containing a dose)
- Efficacy and Safety
  - Clinical Endpoints (depicted by three charts: Surrogate Clinical Endpoints, Mortality, and Symptomatic Relief)
  - Adverse Effects (depicted by a chart plotting Toxicity versus Dose)

Regulatory Process and Timeline

The FDA and CDER have a clearly defined structure to guide sponsors and regulate the drug development process, including timelines and deadlines related to formal reviews and approvals. Two critical components of this structure are the IND application and New Drug Application (NDA) or Biologics License Application (BLA), which are described below and incorporated into the timeline depicted in Figure 2.4.2. While sponsors must follow the IND and NDA/BLA processes, they are encouraged to seek guidance from the FDA throughout the drug development process. In particular, sponsors are encouraged to have a pre-IND meeting with the FDA to review guidance documents, share their proposed drug development plan with the FDA, and solicit FDA feedback on strategies to address specific questions related to the investigational drug's development.[6,9,11] Sponsors are also encouraged to seek guidance from the FDA following completion of Phase 2 studies to guide the design of the pivotal Phase 3 studies.[1,11]

Investigational New Drug Application

For investigational drugs that show promise in preclinical studies, a critical step in the drug development process is the IND application, the approval of which allows research of the investigational drug in humans to proceed. The IND application is submitted upon completion of the Preclinical Phase. The IND application must include data from preclinical studies (e.g., pharmacology, safety/toxicology); data from previous human experience (if any); chemistry, manufacturing, and control information; protocols for the conduct of each study proposed during the clinical research phase, and information about the investigator(s).[9,11]

Once submitted, the FDA will notify the sponsor of the date it was received. The FDA may place the IND application on a Clinical Hold or request additional information before permitting the Clinical Research Phase to proceed.[9,13] If the sponsor does not receive notification of a Clinical Hold or request for more information within 30 days of receiving the application, the IND may go into effect. The IND may go into effect earlier if notified by the FDA that the IND may begin. The manufacturer may ship the investigational new drug to investigators who may administer the drug to human subjects participating in the clinical studies once the IND goes into effect. For more information regarding the components, format, and other details related to the IND application, readers are referred to the Code of Federal Regulations Title 21 Part 312.

New Drug Application & Biologics License Application

For investigational drugs shown to be safe and effective in Phase 3 clinical studies, the sponsor must submit an NDA (or BLA if the drug is a biologic) to the FDA formally requesting approval to market the drug. Before submitting the NDA, the sponsor may meet with the FDA to obtain guidance on the application. The NDA (or BLA) provides the FDA comprehensive information and data regarding the manufacturing process and facilities, quality control, and assurance; complete product description, including the chemical formula/structure, specifications, pharmacodynamics, and pharmacokinetics; data from all preclinical and clinical studies; indications; labeling including directions for use and patient information; safety updates and proposed risk evaluation and mitigation strategies (REMS) that may be applicable; drug abuse information; and institutional review board compliance.[6,14]

Once received from the sponsor, the FDA has 60 days to decide to accept it for review.[1,6] Once accepted for review, the FDA aims to complete its comprehensive review and render a decision within 6 – 10 months.[1,6,14] In addition to reviewing the data contained within the NDA (or BLA), the FDA review also includes inspection of production facilities and clinical study sites. Additionally, the FDA review may include review by one or more of the FDA Advisory Committees to get additional, independent advice from experts, including patient representatives.[14] The NDA/BLA review process often takes 1–2 years to complete before the FDA renders a decision (Figure 5.1).[6,12] For more information regarding the details and requirements for submission, readers are referred to the Code of Federal Regulations Title 21 Part 314 for an NDA or Code of Federal Regulations, Title 21 Part 601 for a BLA.

The FDA estimates that approximately 25–30% of drugs entering Phase 3 move on to the next phase: NDA.[11] However, a recent analysis of clinical drug development between 2011 – 2020 reported a much higher success rate (57.8%) for investigational drugs moving from Phase 3 to NDA.[12] In this same analysis, the approval rate of NDAs was approximately 90%.

Accelerated Approval & Drug Development Designations

For some candidate drugs, the FDA has several mechanisms in place to expedite drug approval and encourage the development of certain drugs. These mechanisms can abbreviate the typical drug development timeline to as little as ~9 years.[15]

- **Accelerated Approval**: Promising candidate drugs that treat serious or life-threatening conditions and provide greater benefit and/or lower risk over currently available therapies can be granted accelerated approval.[1,16] Drugs granted accelerated approval must demonstrate improvements in surrogate endpoints likely to predict clinical benefit or intermediate clinical endpoints less robust than standard endpoints.[16] For drugs granted accelerated approval, the sponsor is required to conduct Phase 4 post-marketing studies to verify and further elucidate the drug's benefits.[1,16] Approval of a drug granted accelerated approval can be withdrawn if Phase 4 studies fail to verify clinical benefits outweigh the risks associated with the drug.[16]
- **Fast Track**: Candidate drugs that treat serious conditions and fill an unmet medical need, based on preclinical and/or early clinical research, may be designated as fast track to expedite the approval process and make the drug available to patients sooner.[17] The sponsor must request the fast track process.
- **Breakthrough Therapy**: Similarly, candidate drugs intended to treat serious conditions for which preliminary evidence has demonstrated substantial benefit over currently available therapies may be designated as breakthrough therapy.[18] Like fast-track drugs, the sponsor must request breakthrough therapy designation. Drugs designated as breakthrough therapy may also be eligible for fast-track designation.
- **Priority Review**: Drugs that significantly improve the treatment, diagnosis, or prevention of serious illnesses may be designated for priority review, which seeks to reduce the time for the FDA to take action on the NDA from 10 months to 6 months or less.[1,19]

Phase 4: Post-Marketing Surveillance

Phase 4 of drug development research refers to the post-marketing evaluation of newly approved drugs. These studies are designed to provide additional information related to the safety, increasing pharmacovigilance, and efficacy of the new drug.[1,7,11] Phase 4 studies are conducted in much larger populations (e.g., thousands of patients) with the disease of interest, often in "real-world" settings.[11] These studies may be required by the FDA as conditions of approval or conducted voluntarily as part of the long-term scientific evaluation by the sponsor or investigators. Importantly, during Phase 4, the evaluation of adverse events voluntarily reported by health professionals, patients, and consumers using the FDA's *MedWatch* program is also done. Reporting adverse events is an expectation for all healthcare workers.

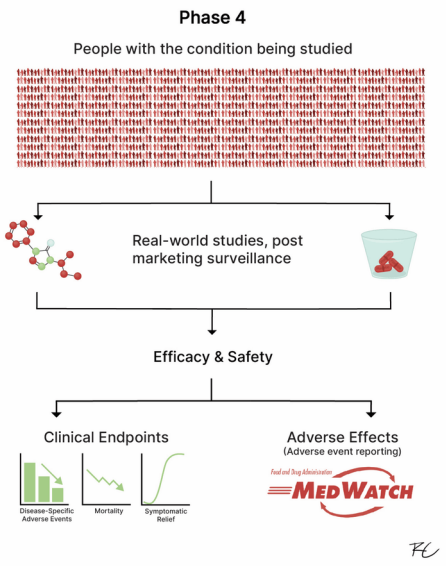


Figure 2.4.7: Phase 4 Studies. (CC-BY 4.0; Riley Cutler)

Accessibility Note

Figure 2.4.7 depicts Phase 4 Studies:

- Testing (depicted by a very large number of people)
- Real-world, post-marketing surveillance studies (depicted by a chemical and container containing a dose)
- Efficacy and Safety
  - Clinical Endpoints (depicted by three charts: Surrogate Clinical Endpoints, Mortality, and Symptomatic Relief)
  - Adverse Effects (depicted by a logo of MedWatch)

References

1. Gibofsky A. The FDA and the Drug Development Process. In: Lockshin MD, Crow MK, Barbhaiya M, eds. *Diagnoses Without Names*. Springer International Publishing; 2022:69-74. doi:10.1007/978-3-031-04935-4_7
2. U.S. Food & Drug Administration. Center for Drug Evaluation and Research | CDER. Published August 14, 2023. Accessed September 18, 2023. FDA CDER.
3. Van Norman GA. Drugs, Devices, and the FDA: Part 1. *JACC: Basic to Translational Science*. 2016;1(3):170-179. doi:10.1016/j.jacbts.2016.03.002
4. BioStock. Drug development: the four phases. Published January 2, 2023. Accessed September 19, 2023. Drug development: the four phases(opens in new window)
5. U.S. Food & Drug Administration. Step 1: Discovery and development. Published January 4, 2018. Accessed September 18, 2023. FDA Step 1 discovery and development(opens in new window)
6. Steinmetz KL, Spack EG. The basics of preclinical drug development for neurodegenerative disease indications. *BMC Neurol*. 2009;9(Suppl 1):S2. doi:10.1186/1471-2377-9-S1-S2
7. U.S. Food & Drug Administration. Step 2: Preclinical research. Published January 4, 2018. Accessed September 18, 2023. FDA Step 2: Preclinical Research(opens in new window)
8. U.S. Food & Drug Administration. Step 3: Clinical research. Published January 4, 2018. Accessed September 18, 2023. FDA Step 3 Clinical Research (opens in new window)
9. Thomas D, Chancellor D, Micklus A, et al. *Clinical Development Success Rates and Contributing Factors 2011-2020*. Biotechnology Innovation Organization | Informa Pharma Intelligence | QLS Advisors LLC; 2021. Clinical Development Success Rates (opens in new window)
10. U.S. Food & Drug Administration. Investigational New Drug (IND) application. Published July 20, 2022. Accessed September 19, 2023. FDA IND application(opens in new window)
11. U.S. Food & Drug Administration. Step 4: FDA drug review. Published January 4, 2018. Accessed September 18, 2023. Step 4: FDA drug review(opens in new window)
12. Brown DG, Wobst HJ, Kapoor A, Kenna LA, Southall N. Drug development times for innovative drugs. *Nat Rev Drug Discov*. 21:793-794. doi:10.1038/d41573-021-00190-9
13. U.S. Food & Drug Administration. Accelerated Approval. Published February 24, 2023. Accessed September 18, 2023. FDA Accelerated Approval (opens in new window)
14. U.S. Food & Drug Administration. Fast track. Published January 4, 2018. Accessed September 18, 2023. FDA Fast Track (opens in new window)
15. U.S. Food & Drug Administration. Breakthrough therapy. Published January 4, 2018. Accessed September 18, 2023. FDA Breakthrough therapy (opens in new window)
16. U.S. Food & Drug Administration. Priority review. Published January 4, 2018. Accessed September 18, 2023. FDA Priority review(opens in new window)

This chapter titled Drug Development for Traditional Drugs and Biologics is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Robert J. DiDomenico from Introduction to Pharmacology by Carl Rosow, David Standaert, & Gary Strichartz (MIT OpenCourseWare) via source content that was edited to the style and standards of the LibreTexts platform; a detailed edit history is available upon request. Figures by Riley Cutler.

Note

Edits that were made include:

- Added Section summary (that will display in the table of contents)
- Edits for accessibility in figures, tables, etc. Converted items to numbered lists.
- Handled figure and table numbers.
- Fixed typos.
- In the above list of references, determine how to handle "opens in a new window." Fixed during accessibility review.
- Converted the list of references to a numbered list as part of the accessibility review.
- In order to change alt text descriptions in Fugures, it was necessary to export and re-import illustrations in them Figures to avoid Libretexts bringing in random illustrations.
- Changed hard coding of Table numbers and Figures to use Libretexts conventions.
- Added extensive Accessibility Notes for each Figure.

This page titled 2.4: Drug Development for Traditional Drugs and Biologics is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Karen Vuckovic.

- User:00c238fa-6460-468d-9639-864e15b4407c has no license indicated.

2.5: Immune System and Vaccines

Note

- Changes to the original OER material: Removed Allergies from the title.
- Deleted the sections on:
  - Allergies
  - Anaphylactic shock
  - Autoimmune disorders.
- Since the sections of "Review of the Anatomy of the Immune System" in the original source are very short, they were combined into one subsection, maintaining full attribution. Deleted some text that was repetitious.
- Since the sections of "Vaccines" in the original source are relatively short, they were combined into one subsection, maintaining full attribution. Deleted figures.
- In the section "Immunization Information Systems" information was added about the California Immunization Registry (CAIR).

This page titled 2.5: Immune System and Vaccines is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Anna Jannak (Consortium of Academic and Research Libraries in Illinois (CARLI)).

2.5.1: Section Goals

- Discuss the immune system, adaptive immunity, and innate immunity.
- Explain the benefits of vaccinations and the medical assistant's responsibilities in vaccine administration and documentation.
- Discuss the CDC's role in vaccines.

This page titled 2.5.1: Section Goals is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Anna Jannak (Consortium of Academic and Research Libraries in Illinois (CARLI)).

2.5.2: Review of the Anatomy of the Immune System

The immune system is a remarkable defense system that the human body relies on to protect against pathogens like bacteria, viruses, and other pathogens. It is a highly intricate and sophisticated network of cells, tissues, organs, and molecules that protect the body from infections and foreign invaders while maintaining overall health.

The critical components of the immune system include white blood cells (leukocytes), which can be further categorized into neutrophils, lymphocytes (including B cells and T cells), monocytes, eosinophils, and basophils.

These white blood cells serve as the immune system's frontline soldiers, involved in phagocytosis, antibody production, and cytokine release.

The lymphatic system, consisting of lymph nodes, the spleen, and the thymus, acts as command centers for immune cell interaction and maturation.

Bone marrow is responsible for producing immune cells, while antibodies and complement proteins, as well as the concept of immunological memory, are crucial components of adaptive immunity.

This page titled 2.5.2: Review of the Anatomy of the Immune System is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Anna Jannak (Consortium of Academic and Research Libraries in Illinois (CARLI)).

2.5.3: Vaccines

As mentioned previously, vaccines expose the immune system to a harmless or weakened form of a pathogen or a part of it. Vaccines are one of the most critical and influential tools in public health for preventing infectious diseases. They have saved countless lives and significantly reduced the burden of many deadly diseases worldwide.

Vaccines are biological preparations that stimulate the immune system to recognize and fight specific pathogens, such as bacteria or viruses. They typically contain either a weakened or inactivated form of the pathogen, a small piece of the pathogen (like a protein), or a similar, harmless version.

When a vaccine is administered, it introduces the pathogen or its components to the immune system without causing the disease. The immune system then produces antibodies and activates immune cells to recognize and remember the pathogen. If the vaccinated person later encounters the actual pathogen, their immune system can rapidly mount a strong and targeted response, preventing or reducing the severity of the disease.

There are various types of vaccines, including:

- live attenuated,
- inactivated,
- recombinant or conjugate, and
- mRNA.

Live attenuated vaccines contain a weakened, live form of the pathogen (e.g., measles, mumps, rubella).

Inactivated vaccines contain killed versions of the pathogen (e.g., polio, hepatitis A).

Recombinant or conjugate vaccines contain specific components of the pathogen (e.g., hepatitis B, human papillomavirus).

mRNA vaccines (e.g., Pfizer-BioNTech and Moderna COVID-19 vaccines) are a newer type that uses a small piece of the pathogen's genetic material to stimulate an immune response.

Vaccines undergo rigorous testing in clinical trials before receiving approval for use. Ongoing monitoring and surveillance systems ensure that vaccines are safe and effective. Adverse reactions to vaccines are typically rare and mild, such as soreness at the injection site or a mild fever. Serious side effects are exceedingly rare.

Centers for Disease Control and Prevention (CDC)

The Centers for Disease Control and Prevention (CDC) in the United States plays a central role in managing and coordinating vaccination programs to protect the public from vaccine-preventable diseases.

The CDC's vaccination program encompasses various aspects, including vaccine research, development, regulation, distribution, education, and monitoring. The CDC researches to develop new vaccines and improve existing ones. They collaborate with other organizations, such as the National Institutes of Health (NIH) and pharmaceutical companies, to advance vaccine science. The CDC also conducts studies to determine the safety and effectiveness of vaccines.

The CDC works closely with manufacturers, healthcare providers, and public health agencies to ensure an adequate supply of vaccines. They maintain the Vaccines for Children (VFC) program, which provides vaccines at no cost to eligible children who might not otherwise have access to immunizations.

The CDC plays a crucial role in educating healthcare professionals, the public, and community leaders about the importance of vaccines. They develop educational materials, host webinars, and engage in public awareness campaigns to promote vaccination.

The CDC monitors vaccine safety through the **Vaccine Adverse Event Reporting System (VAERS)** and the **Vaccine Safety Datalink (VSD)**. These systems track and investigate potential adverse events following vaccination. Timely surveillance helps identify rare or unexpected side effects, contributing to vaccine safety. In the event of vaccine-preventable disease outbreaks, the CDC provides guidance and support to contain the spread of the disease and increase vaccination coverage in affected

communities.

Vaccine Schedules

The CDC's Advisory Committee on Immunization Practices (ACIP) provides expert recommendations on vaccine use in the United States. ACIP develops guidelines for vaccination schedules, including those for infants, children, adolescents, and adults. These recommendations serve as a basis for vaccination policies and guidelines at the state and local levels. The CDC's vaccine schedules are comprehensive and evidence-based plans that outline the recommended vaccinations for individuals of all ages in the United States. The CDC vaccine schedules are divided into age-specific categories, including vaccines for infants, children, adolescents, and adults. The schedules outline the recommended vaccines for each age group and the recommended timing for vaccination.

The Infant and Childhood Vaccination Schedule covers vaccines recommended for children from birth to 18 years old, including vaccines against diseases like measles, mumps, rubella, diphtheria, tetanus, pertussis, and more. The Adolescent Vaccination Schedule includes vaccines recommended for adolescents, such as those protecting against meningitis, human papillomavirus (HPV), and annual influenza vaccination. The adult schedule guides vaccines for adults based on age, medical conditions, and risk factors. It includes vaccines like influenza, shingles, and pneumococcal vaccines. The Catch-up Vaccination Schedule helps individuals who may have missed or fallen behind on their vaccinations catch up to the recommended vaccine series. The schedules also address vaccinations for specific populations, such as pregnant women and travelers, and provide recommendations for healthcare workers. The CDC updates the vaccine schedules annually to reflect the latest scientific research, vaccine approvals, and changes in vaccine recommendations.

Vaccine Information Statement (VIS)

CDC Vaccine Information Statements (VISs) are essential educational materials provided by the Centers for Disease Control and Prevention (CDC) to inform vaccine recipients, their parents, or legal representatives about the benefits and risks associated with vaccines. VISs facilitate informed decision-making regarding vaccinations and promote vaccine safety and public health. VISs serve as a crucial communication tool to ensure that individuals receiving vaccines have access to clear and comprehensive information about the vaccine, its potential side effects, and the importance of vaccination. The goal is to help recipients make informed decisions about vaccination based on their personal circumstances and medical history.

Healthcare providers are required by law in the United States to provide a copy of the appropriate VIS to vaccine recipients or their parents/legal representatives before administering specific vaccines. This requirement is mandated by the National Childhood Vaccine Injury Act (NCVIA) to enhance vaccine safety and the informed consent process.

Each VIS contains specific information related to a particular vaccine, including:

- The name of the vaccine.
- Why the vaccine is recommended.
- A description of the disease(s) the vaccine prevents.
- The recommended vaccination schedule.
- Possible side effects or adverse reactions.
- Who should or should not receive the vaccine.
- What to do in case of an adverse event.

VISs are available in multiple languages to accommodate diverse populations in the United States. Healthcare providers should offer the appropriate language version to ensure that individuals can fully understand the information provided. Healthcare providers typically provide a copy of the relevant VIS to the vaccine recipient, parent, or legal representative before vaccination. Depending on their practice's policies and patient preferences, they may use printed copies, electronic versions, or both. After reviewing the VIS, healthcare providers should address any questions or concerns the recipient or their representative may have.

The medical assistant is responsible for ensuring that the most current issue of the VIS is presented to the patient. Presentation of the VIS to the patient must be documented along with the publication date of the VIS. The date can be found at the bottom of the VIS PDF form available from the website CDC Vaccine Information Statements [www.cdc.gov].

CDC Pink Book

The CDC Pink Book is a reliable resource for the medical assistant to use when administering vaccines. It provides the most comprehensive information on routinely used vaccines and the diseases they prevent. It includes general immunization recommendations, vaccine safety, storage and handling instructions, and detailed information about vaccine-preventable diseases and the vaccines administered to avoid them (National Center for Immunization and Respiratory Diseases, 2021).

Vaccine Administration Documentation

In addition to documenting the presentation of the VIS to the patient or the patient's guardian, there are other items that the MA must legally enter into the patient's health record.

The National Childhood Vaccine Injury Act mandates that the following must be documented in the patient's health record:

- Date of administration
- Vaccine manufacturer
- Vaccine lot number
- Name and title of the person who administered the vaccine and address of the facility where the permanent record will reside
- Vaccine information statement (VIS)
- Date printed on the VIS
- Date the VIS was given to the patient or parent/guardian

Most facilities also require the MA to document the site of vaccine administration and the patient's response (National Center for Immunization and Respiratory Diseases, 2019).

Immunization Information Systems

Immunization information systems (IISs) are confidential, computerized databases that record information on all vaccine doses administered.

In Illinois, this allows patients to access Vax Verify from the Illinois Department of Public Health to view their personal vaccination history. This can help the patient and the provider keep patients up to date with the CDC vaccine schedules (Illinois Department of Public Health, 2024).

In California, the California Immunization Registry (CAIR) is used to maintain vaccination history.[1]

The Vaccine Adverse Event Reporting System (VAERS)

The Vaccine Adverse Event Reporting System (VAERS) is a national vaccine safety surveillance program for reporting adverse events after administering any vaccine in the United States.

VAERS is co-managed by the Centers for Disease Control and Prevention (CDC) and the U.S. Food and Drug Administration (FDA). Anyone can report an adverse event to VAERS, but healthcare professionals must report certain adverse events, and vaccine manufacturers must report all adverse events that come to their attention. Adverse events should be reported even if the cause of the adverse event is uncertain (U.S. Department of Health and Human Services, n.d.)

References

1. California Immunization Registry. Accessed January 17, 2026.

Attributions

Pharmacology-Principles and Application for Medical Assistants by Anna Jannak (Consortium of Academic and Research Libraries in Illinois (CARLI)(opens in new window)) is licensed CC BY 4.0(opens in new window)

This page titled 2.5.3: Vaccines is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Anna Jannak (Consortium of Academic and Research Libraries in Illinois (CARLI)).

2.5.4: References

Centers for Disease Control and Prevention. (2023a, May 12). *Hepatitis B vaccine: What you need to know*. CDC Hepatitis B.

Centers for Disease Control and Prevention. (2023b, October 26). *Recommended child and adolescent immunization schedule for ages 18 years or younger*. CDC schedule.

Cueni, L. N., & Detmar, M. (2008). The lymphatic system in health and disease. *Lymphatic research and biology*, *6*(3-4), 109–122. Lymphatic system.

Illinois Department of Public Health. (2024). *Vax verify*. Illinois Vax Verify.

Institute for Quality and Efficiency in Health Care. (2020, April 23). *How does the immune system work*? Retrieved January 10, 2024, from NIH Immune system.

Janeway, C.A., Travers, P., Walport, M. & Shlomchik, M. (2001). Chapter 10: Adaptive immunity to infection; immunological memory. In *Immunobiology: The Immune System in Health and Disease* (5th ed.). Garland Science. NIH Adaptive Immunity.

MedlinePlus. (2023, December 26). *Immune system and disorders*. Retrieved January 10, 2024. Available from: Immune system and disorders.

National Center for Immunization and Respiratory Diseases. (2019, May 7). *Healthcare providers/professional: Document the vaccination(s)*. U.S. Department of Health & Human Services, Centers for Disease Control and Prevention. CDC Vaccines.

National Center for Immunization and Respiratory Diseases. (2021, August 18). *Epidemiology and prevention of vaccine-preventable diseases*. U.S. Department of Health & Human Services, Centers for Disease Control and Prevention. CDC Pinkbook.

Tigner, A., Ibrahim, S.A., & Murray, I. V. (2022, Nov 14). Histology, white blood cell. In *StatPearls*. NIH Histology, white blood cell.

U.S. Department of Health and Human Services. (n.d.) *About VAERS*. Vaccine Adverse Event Reporting System. HHS VAERS.

Note

Changes made to the original OER material:

- Fixed accessibility issues with links in References called out by the accessibility audit.

This page titled 2.5.4: References is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Anna Jannak (Consortium of Academic and Research Libraries in Illinois (CARLI)).

Anaphylaxis: a severe and life-threatening allergic reaction that occurs rapidly after exposure to an allergen

Antibodies: proteins that bind specifically to antigens to mark them for destruction by immune cells

Antigens: molecules on the surface of pathogens that trigger an immune response

Immune system: defense system that the human body relies on to protect against pathogens

Leukocytes: white blood cells

Pathogen: a foreign invader in the human body

Vaccine: contains a weakened form of a pathogen or a part of it to trigger the formation of memory cells without causing illness

VAERS: Vaccine Adverse Event Reporting System

Note

Changes to the original OER material:

- Changed name of section and put all terms in bold to be consistent with the rest of the textbook.

This page titled 2.5.5: Glossary is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Anna Jannak (Consortium of Academic and Research Libraries in Illinois (CARLI)).

2.6: Introduction to Pharmacology

A drug is a chemical agent which can affect living processes. We will mainly be talking about small molecules which affect cellular processes. Most of these are Xenobiotics (from the Greek xenos - stranger) chemicals that are not synthesized by the body, but introduced into it from outside. There is inevitably a certain amount of ambiguity in this definition: Is oxygen or water a drug? How about Vitamin C in a glass of orange juice? How about an injection of Vitamin C to treat scurvy?

Pharmacology (from the Greek pharmakon - a drug or poison, logos - word or discourse) is the science dealing with actions of drugs on the body (pharmacodynamics) and the fate of drugs in the body (pharmacokinetics). It overlaps with pharmacy, the science of preparation of drugs; much of it deals with therapeutics, the treatment of disease (by whatever means). Toxicology is the branch of pharmacology dealing with the "undesirable" effects of drugs on biological processes (in the case of a nerve gas the bad effect may be a desired one).

In order for a drug to work, it must enter the body and somehow be distributed in such a way that it gets to its site of action. In most cases the site of action is a macromolecular "receptor" located in the target tissue. Most drug effects are temporary, because the body has systems for drug detoxification and elimination. We will consider these issues broadly in this section. As you read, please refer to Figure 2.6.1:

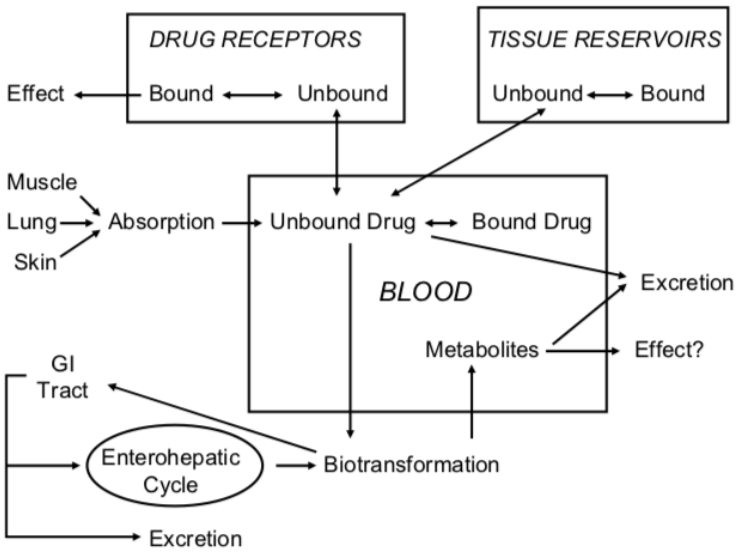


Figure 2.6.1: Drug entering the body, entering bloodstream and receptors. (CC BY-NC-SA 4.0(opens in new window); Carl Rosow, David Standaert, & Gary Strichartz via Principles of Pharmacology)

Accessibility Note

Figure 2.6.1 is a schematic diagram illustrating the movement and fate of a drug within the body. At the center is a large box labeled blood. Inside the blood compartment, drug exists in two forms, labeled unbound drug and bound drug, with arrows showing that the drug can move back and forth between these two states. From the unbound drug in the blood, arrows indicate movement upward to drug receptors, where the drug can shift between bound and unbound states, and the bound state is associated with an effect. From the blood, arrows also point to tissue reservoirs, showing that unbound drug can move into tissues and become bound, and that bound drug can move back into the unbound form.

On the left side of the diagram, arrows labeled muscle, lung, and skin point toward absorption, which then leads into the blood compartment, indicating routes of drug entry into the bloodstream. Below the blood compartment, an arrow leads downward to biotransformation, representing metabolic processing of the drug. From biotransformation, an arrow points to the gastrointestinal tract. From the GI tract an arrow points to an oval labeled enterohepatic cycle (showing recycling of drug or metabolites between the liver and the GI tract), and also points toward excretion.

On the right side of the blood compartment, arrows indicate excretion of the drug and metabolites. Within the blood box, metabolites are shown forming and moving upward from biotransformation, with arrows leading toward excretion and a label questioning whether metabolites may have an effect.

Overall, the diagram shows absorption of a drug into the blood, reversible binding to receptors and tissues, metabolism through biotransformation, possible recycling through the enterohepatic cycle, and eventual excretion, with effects associated primarily with drug binding at receptors.

Overview of Pharmacokinetics - "What the body does to the drug"

1. The drug may enter the body in a variety of ways: as an oral liquid, pill, or capsule; as an inhaled vapor or aerosol; absorbed through intact skin or a mucous membrane; injected into muscle, subcutaneous tissue, spinal fluid, or directly into the bloodstream. As we shall see, the physical properties of the drug and the specific way it is prepared greatly influence the speed of absorption.
2. If the drug is given orally and swallowed, it must be absorbed from the GI tract into the portal circulation. If it is absorbed from the skin, mouth, lungs or muscle it will go directly into the systemic circulation. If drug is injected directly into the bloodstream (e.g., intravenous injection), 100% of it is available for distribution to tissues. This is not usually the case for other modes of administration. For example, drug which is absorbed via the portal circulation must first pass through the liver which is the primary site of drug metabolism (biotransformation). Some of the drug may therefore be metabolized before it ever reaches the systemic blood. In this case, "first-pass" metabolism reduces the bioavailability to less than 100%.
3. Once the drug is in the bloodstream a portion of it may exist as free drug, dissolved in plasma water. Some drug will be reversibly taken up by red cells and some will be reversibly bound to plasma proteins. For many drugs, the bound forms can account for 95-98% of the total. This is important because it is the free drug which traverses cell membranes and produces the effect. It is also important because protein-bound drug can act as a reservoir which releases drug slowly and thus prolongs its action.
4. The unbound drug may then follow its concentration gradient and distribute into peripheral tissues. In some cases, the tissue contains the target site and in others the tissue is not affected by the drug. Sites of non-specific binding act as further reservoirs for the drug. This total volume of distribution determines the equilibrium concentration of drug after a specified dose.
5. Tissue-bound drug eventually reenters the bloodstream where it perfuses the liver and kidneys. The liver metabolizes most drugs into inactive or less active compounds which are more readily excreted. These metabolites and some of the parent compound may be excreted in the bile and eventually may pass out of the body in the feces. Alternatively, some of the drug may be reabsorbed again, farther down the GI tract (the so-called enterohepatic cycle). Any biotransformed drug which is not excreted in bile passes back into the systemic circulation.
6. Parent drug and metabolites in the bloodstream may then be excreted: most are filtered by the kidney, where a portion undergoes reabsorption, and the remainder is excreted in the urine. Some drugs are actively secreted into the renal tubule. Another route of excretion is the lung: Drugs like alcohol and the anesthetic gases are eliminated by this route. Smaller amounts of drug are eliminated in the sweat, tears and breast milk.
7. Biotransformation may sometimes produce metabolites with a great deal of activity. Occasionally, we administer a parent drug which is inactive (a pro-drug) and only the metabolite has activity.

Overview of Pharmacodynamics - "What the drug does to the body"

As stated above, the majority of drugs bind to specific receptors on the surface or interior of cells, but there are many other cellular components and non-specific sites which can serve as sites of drug action.

1. **Water** can be a target. Osmotic diuretics like mannitol are not reabsorbed by the kidney, and the osmotic load they create in the renal tubule obligates the loss of water. Laxatives like magnesium sulfate work in the intestine by the same principle.
2. **Hydrogen** ions can be targets. Ammonium chloride is sometimes used to acidify the urine. When it is taken orally, the liver metabolizes ammonium ion to urea, while the chloride is excreted in the urine. The loss of Cl- obligates the loss of H+ in the urine, thus the pH is lowered.
3. **Metal** ions can be targets. Chelating agents like EDTA may be used to bind divalent cations like Pb++. Metal ions are most frequently drug targets in cases of poisoning.
4. **Enzymes** are targets of many therapeutically useful drugs. Drugs may inhibit enzymes by competitive, non-competitive, or irreversible blockade at a substrate or cofactor binding site. Digitalis glycosides increase myocardial contractility by inhibiting the membrane enzyme, Na+-K+ - ATPase. Antimicrobial and antineoplastic drugs commonly work by inhibiting

enzymes which are critical to the functioning of the cell. In order to be effective, these drugs must have at least some selective toxicity toward bacterial or tumor cells. This usually means that there is a unique metabolic pathway in these cells or some difference in enzyme selectivity for a common metabolic pathway. An example of this is the inhibition of folate synthesis by sulfonamides. These drugs are effective antibacterial agents because the bacteria depend upon folate synthesis, while the host doesn't. This example will be covered in detail in one of our case discussions.

5. **Nucleic acids** are targets for antimetabolites and some antibiotics. In the case of 5- fluorouracil, the compound acts as a counterfeit substitute for uracil and becomes incorporated into a faulty mRNA. Antisense oligonucleotides are another very specific way to interfere with a restricted part of the genome.
6. Some drugs, like general anesthetics, appear to act by non-specific binding to a macromolecular receptor target. These drugs are thought to alter the function of membrane proteins, in part, by disordering the structure of the surrounding lipid membranes. Their lack of specificity is reflected in very low chemical structural requirements. The general anesthetics include compounds as chemically diverse as nitrogen, xenon, halogenated ethers, and steroids. They exhibit very little stereoselectivity, that is, there are not marked differences in anesthetic activity between enantiomers.
7. Finally, we have the drugs which act by binding to specific receptors. These drugs have both high structural specificity and stereoselectivity, i.e., relatively small changes in chemical structure can radically alter the activity of these drugs.

Definitions

Let us finish with some important definitions.

- **Agonist** is a drug which binds to its "receptor" and produces its characteristic effect. A drug may be a full agonist or partial agonist, depending on the maximal effect it produces. An antagonist binds to the receptor without causing an effect, thereby preventing an active substance from gaining access. Antagonists, like enzyme inhibitors, may be competitive, non-competitive or irreversible.
- **Dose-Response**. The sine qua non of drug effect. Simply put, as the dose of drug increases, the response should increase. The curve generated is usually sigmoidal when effect is plotted against log dose (Dr. Strichartz will discuss the theoretical basis for this). Effect may be measured as a graded variable (change in blood pressure, force of contraction) or as a quantal variable (number dead/alive). The slope of the curve is characteristic of the particular drug-receptor interaction. When two drugs act by the same receptor mechanism, we expect to see two parallel log-dose response curves.
- **ED_{50}**. The median effective dose, or the dose which produces a response in 50% of subjects. If the response is death (lethality) we call it the LD_{50}. The EC50 refers to concentration rather than dose. Similar abbreviations are used for other response levels: ED_{99}, LD_1, etc.
- **Potency**. A terribly misused word – the lay public uses it to mean “effectiveness.” The potency of a drug refers to the dose (actually the molar concentration) required to produce a specific intensity of effect. If the ED50 of drug A and B are 5 and 10 mg, respectively, the Relative Potency of A is twice that of B. Relative potency specifically applies to the comparison of drugs which act by the same mechanism, and therefore have parallel dose-response curves.
- **Efficacy**. Also called Maximal Efficacy or Intrinsic Activity. This is the maximum effect of which the drug is capable. A potent drug may have a low efficacy, and a highly efficacious drug may have a low potency. For the clinician, efficacy is much more important than potency (within limits). Who cares if the pill contains 5 or 10 mg of drug?
- **Affinity**. This refers to the strength of binding between drug and receptor. It is quantified by the dissociation constant kD (covered in the next lecture).
- **Selectivity**. This refers to the separation between desired and undesired effects of a drug. In the ideal case, a drug is completely specific, and an effective dose does not elicit any undesired effect. Penicillin is an example of a highly selective drug, since it works specifically by inhibiting cell wall synthesis, and (other than allergic responses) it has very little effect on human cells at normal doses. Unfortunately, many therapeutic agents, like digoxin and theophylline, produce dose-related side effects near their therapeutic dose range. For some drugs like cancer chemotherapeutic agents, their selectivity is their dose-limiting property, i.e., they are given to kill tumor cells until they produce toxicity in normal cells as well.
- **Therapeutic Window**. For every drug, there exists some concentration which is just barely effective (the Effective Concentration) and some dose which is just barely toxic (the Toxic Concentration). Between them is the therapeutic window where most safe and effective treatment will occur.
- **Therapeutic Index**. This is the ratio of toxic to effective doses at the level of 50% response: TD_{50}/ED_{50}. In animal toxicology studies, it is usually the LD_{50}/ED_{50}. Another measure sometimes utilized is the Certain Safety Factor, which is TD_1/ED_{99}.

Note

Changes made to original OER materials:

- Removed underlining of words.
- Removed bolding of text when it did not appear to add value.
- Moved figure to an Elements Template to comply with Libretexts authoring standards.
- Put a detailed Accessibility Note below Figure 2.6.1.

This page titled 2.6: Introduction to Pharmacology is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Carl Rosow, David Standaert, and Gary Strichartz (MIT OpenCourseWare) via source content that was edited to the style and standards of the LibreTexts platform.

- 2.8: Integer Data Type is licensed CC BY-SA 4.0.

2.7: FDA Vaccine Development 101

Note

The material in this section has been extracted from an FDA site almost verbatim, with some edits. The material is in the public domain and this textbook provides attribution to the FDA at the end of this section.

Vaccine Development – 101 | FDA

The U.S. Food and Drug Administration (FDA) is the regulatory authority that has oversight of the safety, effectiveness and quality of vaccines that are used in the United States.

Vaccines to prevent infectious diseases are given to millions of babies, children, adolescents and adults and it is critical that they are demonstrated to be safe and effective. Vaccines have prevented countless cases of disease and disability and have saved millions of lives. Ensuring the safety and effectiveness of vaccines is one of FDA's top priorities.

FDA's Center for Biologics Evaluation and Research (CBER) ensures that FDA's rigorous scientific and regulatory processes are followed by those who pursue the development of vaccines. Vaccine development is a complex science. FDA's scientific and regulatory advice to vaccine developers, as well as FDA's evaluation to determine the safety and effectiveness of vaccines, are among the most robust in the world.

How Vaccines Work

Vaccines work by mimicking the infectious bacteria or viruses that cause disease. Vaccination stimulates the body's immune system to build up defenses against the infectious bacteria or virus (organism) without causing the disease. The parts of the infectious organism that the immune system recognizes are foreign to the body and are called antigens. Vaccination exposes the body to these antigens.

Some vaccines contain weakened versions of a bacteria or virus, other vaccines contain only part of the bacteria or virus. Some vaccines contain only the genetic material for a specific protein and direct the body to produce a small amount of that protein. The body's immune system reacts defensively once it detects this protein.

After vaccination, the immune system is prepared to respond quickly and forcefully when the body encounters the real disease-causing organism.

Below is the typical process that FDA expects vaccine developers to follow to generate the information it needs to assess the safety and effectiveness of a vaccine to prevent an infectious disease.

Laboratory Research and Discovery Stage

Research and Discovery Stage Scientists develop a rationale for a vaccine based on how the infectious organism causes disease. The scientists then conduct laboratory research to test their idea for a vaccine candidate; sometimes this testing occurs in animals. This is considered the Research and Discovery Stage. Once a scientific finding is thought to have practical applications, in that it may be feasible to develop a vaccine candidate based on that finding, the research moves forward.

Research Moves Forward

Pre-clinical

Before a vaccine can be tested in people, a company or researcher performs additional laboratory research and testing in animals to obtain information about how the vaccine works and whether it's likely to be safe and work well in humans. These tests are known as the Pre-clinical phase.

Clinical Development

When the company/researcher is ready to begin studies in humans, they compile the results of their laboratory and other preclinical testing, as well as information pertaining to the manufacturing technology and the quality of the vaccine and submit these to FDA in the form of an Investigational New Drug application (IND). FDA evaluation includes an assessment of the preclinical data and a determination whether these tests were conducted according to Good Laboratory Practices. FDA also conducts an assessment of the product, its quality and safety, and the technology to manufacture it, to determine whether it is reasonably safe for testing of the vaccine to move forward in people. Studies conducted in people are known as the *Clinical Development* stage and typically cover three phases under the oversight of FDA.

Clinical trials are conducted according to plans that reflect FDA's considerable expertise in clinical trial design. These plans are called "protocols." Vaccines intended for children are generally tested first in adults, with a step-down clinical development program to children and infants.

> Note
>
> The protocols would include the design of the study including estimates of the sample size to be used, enlisting the clinical trial participant, documentation of follow-up visit plans after the administration of the vaccine, response variables, and plans for data analysis.

The phases of the studies may progress sequentially, but it is also not uncommon for the phases of development to overlap.

Phase 1: Emphasis during this phase is on safety and generally includes 20–100 volunteers who haven't been exposed to the disease being studied and who are generally otherwise healthy. These studies are used to determine whether there are adverse reactions with increasing doses and, if possible, to gain early information about how well the vaccine works to induce an immune response in people.

Phase 2: In the absence of safety concerns from phase 1 studies, phase 2 studies include more people, where various dosages are tested on hundreds of people with typically varying health statuses and from different demographic groups, in randomized-controlled studies. These studies provide additional safety information on common short-term side effects and risks, examine the relationship between the dose administered and the immune response, and provide initial information regarding the effectiveness of the vaccine in its ability to generate an immune response. Standardized and validated tests are used to evaluate the immune responses. These vaccine studies typically also include a control group consisting of people who may receive an FDA-approved vaccine, a placebo or another substance. People receiving the vaccine under study are compared to people in the control group.

Phase 3: The vaccine is generally administered to thousands of people and the study generates critical information on effectiveness and additional important safety data. This phase includes additional information about immune response and compares those who receive the vaccine to those who receive a control, such as a placebo. For example, the number of cases of disease in the vaccinated group is compared to the number in the control group to see whether the vaccine reduces the incidence of disease. These studies also provide information about the vaccine's safety including the identification of less common side effects.

Special Considerations

- In public health emergencies, such as a pandemic, the development process may be atypical or expedited. For example, as demonstrated by the response to the COVID-19 pandemic, the U.S. government may coalesce government agencies, international counterparts, academia, nonprofit organizations and pharmaceutical companies to develop a coordinated strategy for prioritizing and speeding development of the most promising treatments and vaccines. In addition, the federal government may make investments in the necessary manufacturing capacity at its own risk, giving companies confidence that they can invest aggressively in development and allowing faster distribution of an eventual vaccine.
  Recognizing the urgent need for safe and effective vaccines FDA utilizes its various authorities and expertise to facilitate the expeditious development and availability of safe and effective vaccines. Early in a public health crisis, FDA provides clear communication to the pharmaceutical industry pertaining to the scientific data and information needed for safe and effective vaccines and works quickly to provide advice on their proposed development plans and assessment of the data that are generated. During a public health emergency, if certain criteria are met, manufacturers may submit a request for Emergency Use Authorization (EUA) to FDA to facilitate the availability and use of their vaccine during this time.

- When evaluating the need for pre-clinical studies for a vaccine, FDA considers all data relevant to that vaccine and closely related vaccines, as well as the design of the specific clinical study for the vaccine in question. If FDA's evaluation of these data support initiating human studies of a candidate vaccine intended to prevent an infectious disease without first completing additional toxicology studies, then FDA would consider allowing such human studies to proceed.
- There is no predetermined timeline for vaccine development. Typically, the better the scientific understanding of a pathogen and the disease it causes, the more efficient vaccine development.
- Adaptive Trial Designs: These are clinical study designs which aim to expedite clinical trial decisions based on preliminary results derived from earlier trials and, in some cases, from the same trial. Using this approach can facilitate efficient clinical development. The goal of these designs is to reduce the size and duration of the trial and demonstrate an effect if one exists. These adaptations are performed with close attention to statistical rigor.

Assessment of Manufacturing

While the vaccine is being tested in people, FDA is also assessing information pertaining to the manufacturing of the vaccine and the facility where it will be made. Vaccine manufacturing is complex. The process of making the candidate vaccine for the phase 3 studies in batches called "lots" helps the manufacturer ramp up for commercial-scale manufacturing. FDA requires vaccine manufacturers to submit data to support manufacturing processes, facilities, product characterization, and demonstration of lot-to-lot consistency. FDA works with the manufacturer to develop a lot release protocol (a template of tests to be conducted on the vaccine) that will be used for each lot of vaccine post-approval. Experienced FDA-investigators carefully examine and evaluate the facility and operation for compliance with FDA regulations.

Once a manufacturing process is developed that ensures that the vaccine can be produced reliably and consistently, and the preclinical and clinical development programs have been successfully completed, companies submit a Biologics License Application (BLA) to the FDA. A BLA is a comprehensive submission that is submitted to the Agency. It includes preclinical and clinical data and information, as well as details of the manufacturing process and facility(ies).

Seeking Approval

By submitting a BLA to the FDA, a company is seeking permission to distribute and market a vaccine for use in the United States. FDA evaluates the data to determine whether the safety and effectiveness of the vaccine has been demonstrated and whether the manufacturing and facility information assure product quality and consistency. After its evaluation, FDA decides whether to approve (also known as to license) the vaccine for use in the United States. If FDA approves the vaccine, the company is permitted to market it in the United States for use in the population for which it is approved.

FDA makes its decisions based on its analysis of the benefits and risks for the intended population who will receive the vaccine, as well as the disease(s) to be prevented. FDA's scientific team works collaboratively to evaluate all of the scientific data and information included in the BLA and makes the determination whether to approve a vaccine. A typical FDA team is comprised of: physicians, chemists, statisticians, pharmacologists/toxicologists, microbiologists, experts in post-marketing safety, clinical study site inspectors, manufacturing and facility inspectors, and labeling and communications experts.

In some cases, FDA seeks the input of its Vaccines and Related Biological Products Advisory Committee (VRBPAC). This committee is comprised of a panel of outside, independent, technical experts from various scientific and public health disciplines that provide input on scientific data and its public health significance in a public forum. The FDA will consider, but is not bound by, the input received from the VRBPAC when determining whether to approve a vaccine.

Prescribing Information and Labeling

Prescribing information for a vaccine is based on scientific data that are submitted by the manufacturer in the BLA and determined by the FDA to be satisfactory to support the approved indication(s), usage, dosing, and administration. The prescribing information is updated as needed to include the most current information about the vaccine that is available to and reviewed by FDA. The prescribing information does not necessarily address all aspects of vaccine use, such as recommendations that are specific to disease outbreaks, vaccine shortages, and all subpopulations with underlying medical conditions.

FDA Oversight after Approval

Monitoring of Safety and Effectiveness

It is important to note that a vaccine is a drug. Like any drug, vaccines have benefits and risks, and even when highly effective, no vaccine is 100 percent effective in preventing disease or 100 percent safe in all individuals. Most side effects of vaccines are usually minor and short-lived. For example, a person may feel soreness at the injection site or experience a mild fever. Serious vaccine reactions are extremely rare, but they can happen.

Although the vaccine development process and FDA's evaluation are rigorous and comprehensive, there is still a need for ongoing surveillance of vaccines after FDA-approval to identify uncommon adverse events or long-term complications that may occur, and sometimes to monitor effectiveness. In certain cases, the FDA may require the manufacturer to conduct post-marketing studies to further assess known or potential serious risks. (These studies are sometimes called Phase 4 of development).

Vaccines are closely monitored using various surveillance systems, such as the Vaccine Adverse Event Reporting System (VAERS), the FDA BEST (Biologics Effectiveness and Safety) program and the FDA Sentinel Program, the FDA and Centers for Medicare & Medicaid Services (CMS) partnership, and the Centers for Disease Control and Prevention's (CDC) Vaccine Safety Datalink.

Lot Release

Lot release is a mechanism that provides FDA with a real-time system to continuously monitor product quality. As previously noted, vaccines are generally made in batches called lots. FDA requires vaccine manufacturers to submit data to support the demonstration of lot-to-lot consistency. After approval, the manufacturer must submit the following materials relating to that vaccine lot (or "batch"):

- Protocols: contain the agreed-upon tests.
- Results: the results of the testing performed by the manufacturer. Testing typically includes assessment of purity, potency, identity and sterility.
- Samples: generally, the manufacturer must submit samples of the vaccine from the lot in question to permit FDA to perform confirmatory testing.

Manufacturers are not permitted to distribute a specific lot of vaccine until the FDA releases it.

FDA Research

Research is fundamental to FDA's ability to provide effective scientific and regulatory evaluation of vaccines. FDA conducts its research activities in conjunction with its regulatory activities, which provides the Agency a unique perspective on both fronts. A wide variety of rapidly evolving technical and scientific issues concerning the safety, potency, and effectiveness of vaccines requires knowledge of new developments in basic research in the relevant biological disciplines. For this reason, FDA scientists conduct a variety of research that contributes to policy, risk assessments, new methods and standards, and changes to product labeling, including promoting new techniques for assessing vaccine safety and potency, as well as strategies for vaccine development.

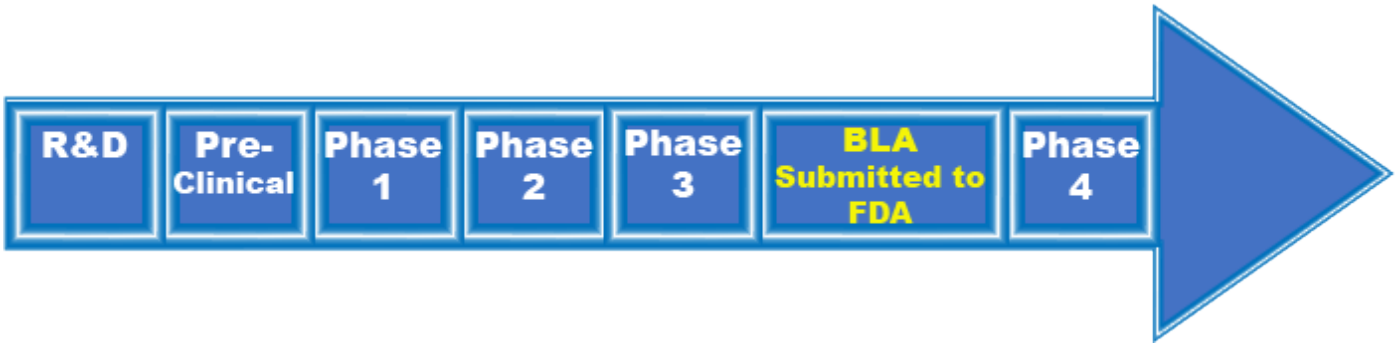


Figure 2.7.1: Illustration of the Vaccine Approval Process. Includes R&D, Pre-Clinical, Phase 1, Phase 2, Phase 3, BLA Submitted to FDA, and Phase 4. (FDA, public domain [2])

Note

The overall process for vaccine development and approval includes:

- Research and Development
- Pre-clinical trials with non-humans
- The filing of an Investigational New Drug (IND) application with the FDA at the conclusion of pre-clinical trials requesting proceeding to in-human testing
- Pre-licensure vaccine Clinical Trials (Phases 1, 2, and 3)
- Filing a Biologics License Application (BLA) at the conclusion of Clinical Trial Phase 3 with the FDA
- Inspection of the pharmaceutical company's manufacturing facility or facilities and achieving approval
- Presenting the pharmaceutical company's findings to the FDA's Vaccines and Related Biological Products Advisory Committee (VRBPAC)
- Usability testing of the vaccine product labeling.

Attributions/References

1. Food and Drug Administration, FDA Vaccine Development 101.
2. Food and Drug Administration, Illustration of the FDA Approval Process.

Note

FDA Website Policies. Downloaded January 7, 2026.

Linking to or Copying Information on the FDA Website

Unless otherwise noted, the contents of the FDA website (www.fda.gov) — both text and graphics — are not copyrighted. They are in the public domain and may be republished, reprinted and otherwise used freely by anyone without the need to obtain permission from FDA. Credit to the U.S. Food and Drug Administration as the source is appreciated but not required.

People are also free to link to any URL on FDA's site. FDA's preference is that people link to the material on the FDA site (rather than copying it to their personal websites) because the agency continuously updates the information on the website as better information becomes available. A person copying documents to another website, instead of linking to them, would then have to monitor the original documents to know when these documents were updated by FDA or else risk giving bad or incorrect advice to visitors to their website. Providing consumers or health professionals with advice that is not fully up to date can lead to serious public health consequences. Providing industry advice that is not fully up to date can lead to companies being out of compliance with regulatory requirements.

If a person, nonetheless, decides to copy content or images, FDA strongly recommends that the copied item lists the date that the material was copied and provides a link back to its source on the FDA website. Users can then see for themselves if the copied material has been updated or changed.

This page titled 2.7: FDA Vaccine Development 101 is shared under a Public Domain license and was authored, remixed, and/or curated by LibreTexts.

2.8: mRNA Vaccines

As mentioned earlier, mRNA vaccines (e.g., Pfizer-BioNTech and Moderna COVID-19 vaccines) are newer types of vaccines that uses a small piece of the pathogen's (virus's) genetic material to stimulate an immune response.

Definition: mRNA vaccine

An **mRNA vaccine** is a type of vaccine that uses a copy of a molecule called messenger RNA (mRNA) to produce an immune response. The vaccine delivers molecules of antigen-encoding mRNA into cells, which then use the designed mRNA as a blueprint to build foreign protein that would normally be produced by a pathogen (such as a virus) or by a cancer cell. These protein molecules stimulate an adaptive immune response that teaches the body to identify and destroy the corresponding pathogen or cancer cells. The mRNA is delivered by a co-formulation of the RNA encapsulated in lipid nanoparticles that protect the RNA strands and help their absorption into the cells.[1]

The advantages of mRNA vaccines over traditional vaccines are:

- ease of design,
- speed and lower cost of production,
- the induction of both cellular and humoral immunity,
- and lack of interaction with the genomic DNA.[1]

Definition: humoral

Humoral: relating to the body fluids, especially with regard to immune responses involving antibodies in body fluids as distinct from cells.

Mechanism of operation[1]

The goal of a vaccine is to stimulate the adaptive immune system to create antibodies that precisely target that particular pathogen. The markers on the pathogen that the antibodies target are called antigens.

Traditional vaccines stimulate an antibody response by injecting either antigens , an attenuated (weakened) virus, an inactivated (dead) virus, or a recombinant antigen-encoding viral vector (harmless carrier virus with an antigen transgene) into the body. These antigens and viruses are prepared and grown outside the body.

In contrast, mRNA vaccines introduce a short-lived synthetically created fragment of the RNA sequence of a virus into the individual being vaccinated. These mRNA fragments are taken up by dendritic cells through phagocytosis. The dendritic cells use their internal machinery (ribosomes) to read the mRNA and produce the viral antigens that the mRNA encodes. The body degrades the mRNA fragments within a few days of introduction. Although non-immune cells can potentially also absorb vaccine mRNA, produce antigens, and display the antigens on their surfaces, dendritic cells absorb the mRNA globules much more readily. The mRNA fragments are translated in the cytoplasm and do not affect the body's genomic DNA, located separately in the cell nucleus.

Once the viral antigens are produced by the host cell, the normal adaptive immune system processes are followed. Antigens are broken down by proteasomes. Class I and class II MHC molecules then attach to the antigen and transport it to the cellular membrane, "activating" the dendritic cell. Once activated, dendritic cells migrate to lymph nodes, where they present the antigen to T cells and B cells. This triggers the production of antibodies specifically targeted to the antigen, ultimately resulting in immunity.

Target: Spike protein

Definition: Term

In virology, a spike protein or peplomer protein is a protein that forms a large structure known as a spike or peplomer projecting from the surface of an enveloped virus. [4]

Coronaviruses exhibit coronavirus spike protein, also known as the S protein, on their surfaces; S is a class I fusion protein and is responsible for mediating viral entry as the first step in viral infection. It is highly antigenic and accounts for most antibodies produced by the immune system in response to infection. For this reason the spike protein has been the focus of development for COVID-19 vaccines in response to the COVID-19 pandemic caused by the virus SARS-CoV-2.[4]

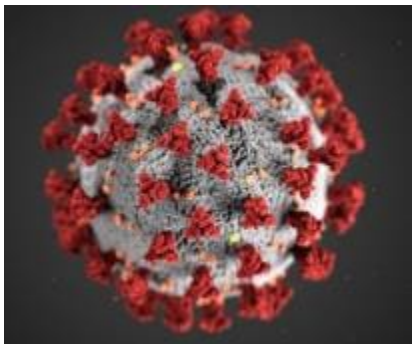


Figure 2.8.1: Image of SARS-CoV-2 virus. (Centers for Disease Control and Prevention (CDC), COVID-19 Vaccination Program Interim Playbook for Jurisdiction Operations, September 16, 2020. Image on page 1 and 2.)[5]

Storage at low temperatures

The mRNA vaccines need to be maintained at low temperatures:

- Pfizer–BioNTech's BNT162b2 mRNA vaccine has to be kept between −80 and −60 °C (−112 and −76 °F) and
- Moderna's mRNA-1273 vaccine requires storage between −25 and −15 °C (−13 and 5 °F).[1]

This low storage requirement brought into popular parlance the term "Cold chain logistics." Project Warp Speed needed to procure the appropriate refrigeration equipment for use in transporting the vaccines through the distribution chain and required that providers maintain the required temperatures to prevent degradation of the mRNA vaccine.

Definition: Cold chain

A **cold chain** is a supply chain that uses refrigeration to maintain perishable goods, such as pharmaceuticals, produce or other goods that are temperature-sensitive.[2]

The cold chain is used in the supply of vaccines to distant clinics in hot climates served by poorly developed transport networks. Disruption of a cold chain due to war or logistical challenges may produce severe consequences, as distributed vaccines can become inert due to a lack of temperature control during transport. The integrity of the vaccine cold chain is therefore a critical public health concern.[2]

For vaccines, there are different types of cold chains:[2]

- the **ultralow**, or deep freeze, cold chain for vaccines that require -70 degrees C, such as the Ebola and Pfizer-BioNTech COVID-19 vaccines, and some animal vaccines.
- the **frozen** chain requires -20 degrees C. Varicella and zoster vaccinations require this level.
- the **refrigerated** chain, which requires temperatures between two and eight degrees C. Most flu vaccinations only require refrigeration.

In 2020, during the COVID-19 pandemic, vaccines being developed needed ultracold storage and transportation temperatures as cold as −70 °C (−94 °F), requiring what has been referred to as a "colder chain" infrastructure. This creates some issues of distribution for the Pfizer vaccine. It is estimated that only 25 to 30 countries in the world have the infrastructure for the required ultracold cold chain.[2]

The World Health Organization (WHO) provided guidance for Ultra-low temperature (ULT) storage and transport for vaccines.[3]

Continuous data logging is fundamental for validation and compliance. This is achieved using standalone temperature data loggers, RFID tags that record temperature, or, most commonly, through the integrated telematics system. This creates an unbroken digital record of a shipment's temperature history from origin to destination.[3]

References

1. Wikipedia - mRNA vaccine. Accessed January 18, 2026.
2. Wikipedia - Cold chain. Accessed January 18, 2026.
3. WHO vaccine low temp. World Health Organization, Ultra-low temperature (ULT) storage and transport for vaccines, An overview of options and challenges. February 2021. Accessed January 18, 2026.
4. Wikipedia - Spike protein. Accessed January 18, 2026.
5. Centers for Disease Control and Prevention (CDC), COVID-19 Vaccination Program Interim Playbook for Jurisdiction Operations, September 16, 2020. Image on page 1 and 2.

This page titled 2.8: mRNA Vaccines is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

2.9: Investigational New Drug (IND) Application

What are INDs?

In this section we will cover the Investigational New Drug Application (IND) used for CBER-Regulated products such as vaccines.

An Investigational New Drug Application (IND) is a request from a clinical study sponsor to obtain authorization from the Food and Drug Administration (FDA) to administer an investigational drug or biological product to humans. Clinical studies are often conducted to collect safety and effectiveness information in support of marketing applications for biologic and drug products. Unless exempted, the sponsor for a clinical study must obtain authorization from FDA for conducting the study by submitting an IND application. Such authorization must be secured prior to interstate shipment and administration of any new drug or biological product that is not the subject of an approved New Drug Application or Biologics Product License Application.

Current Federal law requires that a drug be the subject of an approved marketing application before it is transported or distributed across state lines. Because a sponsor will probably want to ship the investigational drug to clinical investigators in many states, it must seek an exemption from that legal requirement. The IND is the means through which the sponsor technically obtains this exemption from the FDA.

During a new drug's early preclinical development, the sponsor's primary goal is to determine if the product is reasonably safe for initial use in humans, and if the compound exhibits pharmacological activity that justifies commercial development. When a product is identified as a viable candidate for further development, the sponsor then focuses on collecting the data and information necessary to establish that the product will not expose humans to unreasonable risks when used in limited, early-stage clinical studies.

FDA's role in the development of a new drug begins when the drug's sponsor (usually the manufacturer or potential marketer), having screened the new molecule for pharmacological activity and acute toxicity potential in animals, wants to test its diagnostic or therapeutic potential in humans. At that point, the molecule changes in legal status under the Federal Food, Drug, and Cosmetic Act and becomes a new drug subject to specific requirements of the drug regulatory system.

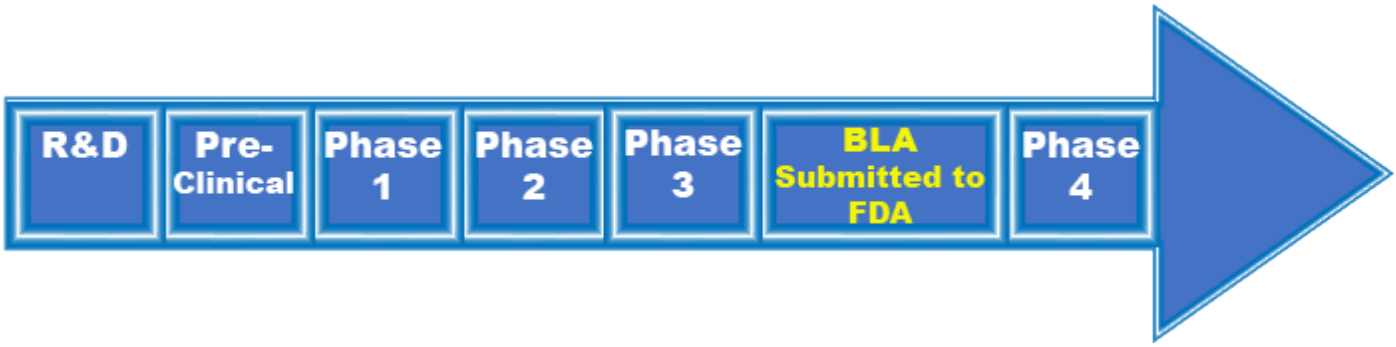


Figure 2.9.1: Vaccine Approval Process. Includes R&D, Pre-Clinical, Phase 1, Phase 2, Phase 3, BLA Submitted to FDA, and Phase 4. (FDA, public domain [5])

Note

The overall process for vaccine development and approval includes:

- Research and Development
- Pre-clinical trials with non-humans
- The filing of an Investigational New Drug (IND) application with the FDA at the conclusion of pre-clinical trials requesting proceeding to in-human testing
- Pre-licensure vaccine Clinical Trials (Phases 1, 2, and 3)
- Filing a Biologics License Application (BLA) at the conclusion of Clinical Trial Phase 3 with the FDA
- Inspection of the pharmaceutical company's manufacturing facility or facilities and achieving approval

- Presenting the pharmaceutical company's findings to the FDA's Vaccines and Related Biological Products Advisory Committee (VRBPAC)
- Usability testing of the vaccine product labeling.

There are three IND types:

1. An **Investigator IND** is submitted by a physician who both initiates and conducts an investigation, and under whose immediate direction the investigational drug is administered or dispensed. A physician might submit a research IND to propose studying an unapproved drug, or an approved product for a new indication or in a new patient population.
2. An **Emergency Use IND** allows the FDA to authorize use of an experimental drug in an emergency situation that does not allow time for submission of an IND in accordance with 21CFR, Sec.312.23, or Sec. 312.20. It is also used for patients who do not meet the criteria of an existing study protocol, or if an approved study protocol does not exist.
3. A **Treatment IND** is submitted for experimental drugs showing promise in clinical testing for serious or immediately life-threatening conditions while the final clinical work is conducted and the FDA review takes place.

There are two IND categories:

1. Commercial
2. Research (non-commercial)

The IND application must contain information in three broad areas:

1. **Animal Pharmacology and Toxicology Studies:** Preclinical data to permit an assessment as to whether the product is reasonably safe for initial testing in humans. Also included are any previous experience with the drug in humans (often foreign use).
2. **Manufacturing Information:** Information pertaining to the composition, manufacturer, stability, and controls used for manufacturing the drug substance and the drug product. This information is assessed to ensure that the company can adequately produce and supply consistent batches of the drug.
3. **Clinical Protocols and Investigator Information:**
    - Detailed protocols for proposed clinical studies to assess whether the initial-phase trials will expose subjects to unnecessary risks.
    - The qualifications of clinical investigators: professionals (generally physicians) who oversee the administration of the experimental compound in order to assess whether they are qualified to fulfill their clinical trial duties.
    - Commitments:
        - to obtain informed consent from the research subjects,
        - to obtain review of the study by an institutional review board (IRB), and
        - to adhere to the investigational new drug regulations.

Once the IND is submitted, the sponsor must wait 30 calendar days before initiating any clinical trials. During this time, FDA has an opportunity to review the IND for safety to assure that research subjects will not be subjected to unreasonable risk.

Emergency Use INDs:

Emergency Use IND allows the FDA to authorize use of an experimental drug in an emergency situation that does not allow time for submission of an IND in accordance with 21CFR , Sec. 312.23 or Sec. 312.20. It is also used for patients who do not meet the criteria of an existing study protocol, or if an approved study protocol does not exist.

Regulatory Compliance:

Clinical studies must follow a set of laws and regulations, which are intended to protect the rights, safety, and welfare of human subjects participating in human trials, ensure the quality, validity, and integrity of the clinical trial data, and promote the availability of new medical products to the public. These laws and regulations define the roles and responsibilities of entities, such as sponsors, clinical investigators, and institutional review boards. In addition, various guidance documents and standard operating procedures are available to clarify policies and procedures for the IND process.

Frequently Asked Questions (FAQs)

1. Q: When is an IND application required?

   A: An IND application is required when a sponsor intends to conduct a clinical study involving the administration of an investigational product to humans, and the study is not exempt from IND requirements. Unless exempted, sponsors must obtain authorization from the FDA by submitting an IND application initiating administration of the investigational product.

2. Q: How do I prepare to submit an IND application?

   A: Sponsors should compile all necessary information and documentation required for the IND submission, including preclinical data, clinical trial protocols, manufacturing information, and safety monitoring plans. The submission should be prepared according to FDA regulations and guidance documents.

3. Q: What are the key components of the IND application?

   A: The content and format of an IND submission must be complete, well-organized as per 21 CFR 312, and include all applicable FDA Forms, provided below. These forms can also be found in FDA's Form Database.

   - Form FDA 1571: Investigational New Drug Application
   - For Individual patient INDs, a licensed physician may use Form FDA 3926: Individual Patient Expanded Access IND in place of the Form FDA 1571
   - Form FDA 1572: Statement of Investigator
   - Form FDA 3454: Certification: Financial Interests and Arrangements of Clinical Investigators
   - Form FDA 3455: Disclosure: Financial Interests and Arrangements of Clinical Investigators
   - Form FDA 3674: Certification of Compliance

4. Q: What happens after I submit an IND application?

   A: After submission, the assigned CBER Regulatory Project Manager (RPM) will acknowledge receipt of the IND application to inform the sponsor that the application has been received. CBER will review the IND application to ensure compliance with regulatory requirements and assess the safety and scientific merit of the proposed clinical studies. The review process may include requests for additional information or clarification from the sponsor.

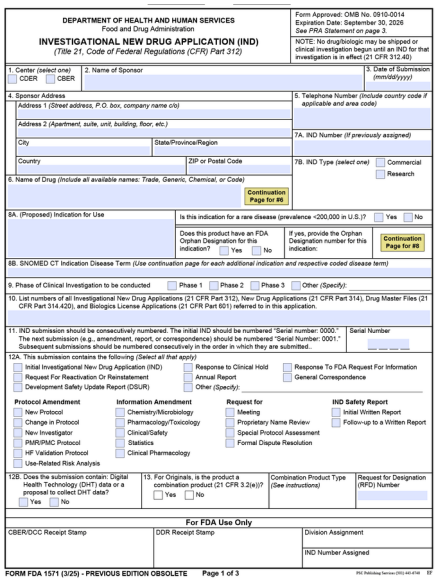
5. Q: When can I initiate a study after I submitted an IND?

   A: Only after CBER reviews the IND application within 30 days from the receipt and notifies the sponsor that the study is safe to proceed, the sponsor can initiate the clinical study. If the study is deemed not safe to proceed due to deficiencies, the IND is a subject to a clinical hold.

FDA Form 1571 Investigational New Drug Application [2]

Since FDA form 1571 Investigational New Drug Application is a project deliverable, let's take a look at what the three-page form looks like.

It will help inform what work is required and which team members will be assigned the needed work activities to prepare all the necessary supporting documents.



DEPARTMENT OF HEALTH AND HUMAN SERVICES
Food and Drug Administration

INVESTIGATIONAL NEW DRUG APPLICATION (IND)
(Title 21, Code of Federal Regulations (CFR) Part 312)

Form Approved: OMB No. 0910-0014
Expiration Date: September 30, 2026
See PRA Statement on page 3.

NOTE: No drug/biologic may be shipped or clinical investigation begun until an IND for that investigation is in effect (21 CFR 312.40)

1. Center *(select one)*: ☐ CDER ☐ CBER
2. Name of Sponsor
3. Date of Submission *(mm/dd/yyyy)*
4. Sponsor Address
Address 1 *(Street address, P.O. box, company name c/o)*
Address 2 *(Apartment, suite, unit, building, floor, etc.)*
City | State/Province/Region
Country | ZIP or Postal Code
5. Telephone Number *(Include country code if applicable and area code)*
7A. IND Number *(If previously assigned)*
7B. IND Type *(select one)*: ☐ Commercial ☐ Research
6. Name of Drug *(Include all available names: Trade, Generic, Chemical, or Code)*
Continuation Page for #6
8A. (Proposed) Indication for Use
Is this indication for a rare disease (prevalence <200,000 in U.S.)? ☐ Yes ☐ No
Does this product have an FDA Orphan Designation for this indication? ☐ Yes ☐ No
If yes, provide the Orphan Designation number for this indication:
Continuation Page for #8
8B. SNOMED CT Indication Disease Term *(Use continuation page for each additional indication and respective coded disease term)*
9. Phase of Clinical Investigation to be conducted ☐ Phase 1 ☐ Phase 2 ☐ Phase 3 ☐ Other *(Specify)*:
10. List numbers of all Investigational New Drug Applications (21 CFR Part 312), New Drug Applications (21 CFR Part 314), Drug Master Files (21 CFR Part 314.420), and Biologics License Applications (21 CFR Part 601) referred to in this application.
11. IND submission should be consecutively numbered. The initial IND should be numbered "Serial number: 0000." The next submission (e.g., amendment, report, or correspondence) should be numbered "Serial Number: 0001." Subsequent submissions should be numbered consecutively in the order in which they are submitted.
Serial Number
12A. This submission contains the following *(Select all that apply)*
☐ Initial Investigational New Drug Application (IND)
☐ Request For Reactivation Or Reinstatement
☐ Development Safety Update Report (DSUR)
☐ Response to Clinical Hold
☐ Annual Report
☐ Other *(Specify)*:
☐ Response To FDA Request For Information
☐ General Correspondence

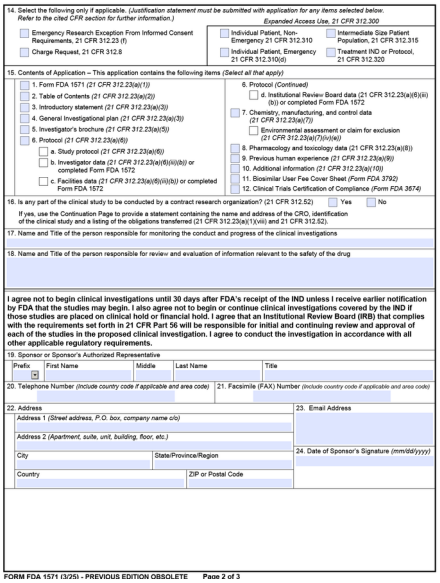
Protocol Amendment	Information Amendment	Request for	IND Safety Report
☐ New Protocol	☐ Chemistry/Microbiology	☐ Meeting	☐ Initial Written Report
☐ Change in Protocol	☐ Pharmacology/Toxicology	☐ Proprietary Name Review	☐ Follow-up to a Written Report
☐ New Investigator	☐ Clinical/Safety	☐ Special Protocol Assessment	
☐ PMR/PMC Protocol	☐ Statistics	☐ Formal Dispute Resolution	
☐ HF Validation Protocol	☐ Clinical Pharmacology		
☐ Use-Related Risk Analysis			

12B. Does the submission contain: Digital Health Technology (DHT) data or a proposal to collect DHT data? ☐ Yes ☐ No
13. For Originals, is the product a combination product (21 CFR 3.2(e))? ☐ Yes ☐ No
Combination Product Type *(See instructions)*
Request for Designation (RFD) Number

For FDA Use Only
CBER/OCC Receipt Stamp | DDR Receipt Stamp | Division Assignment | IND Number Assigned

FORM FDA 1571 (3/25) - PREVIOUS EDITION OBSOLETE Page 1 of 3

Figure 2.9.2: IND Form 1571 Page 1. (FDA [2])



14. Select the following only if applicable. *(Justification statement must be submitted with application for any items selected below. Refer to the cited CFR section for further information.)*
Expanded Access Use, 21 CFR 312.300
☐ Emergency Research Exception From Informed Consent Requirements, 21 CFR 312.23 (f)
☐ Charge Request, 21 CFR 312.8
☐ Individual Patient, Non-Emergency 21 CFR 312.310
☐ Individual Patient, Emergency 21 CFR 312.310(d)
☐ Intermediate Size Patient Population, 21 CFR 312.315
☐ Treatment IND or Protocol, 21 CFR 312.320

15. Contents of Application – This application contains the following items *(Select all that apply)*
☐ 1. Form FDA 1571 *(21 CFR 312.23(a)(1))*
☐ 2. Table of Contents *(21 CFR 312.23(a)(2))*
☐ 3. Introductory statement *(21 CFR 312.23(a)(3))*
☐ 4. General Investigational plan *(21 CFR 312.23(a)(3))*
☐ 5. Investigator's brochure *(21 CFR 312.23(a)(5))*
☐ 6. Protocol *(21 CFR 312.23(a)(6))*
☐ a. Study protocol *(21 CFR 312.23(a)(6))*
☐ b. Investigator data *(21 CFR 312.23(a)(6)(iii)(b))* or completed Form FDA 1572
☐ c. Facilities data *(21 CFR 312.23(a)(6)(iii)(b))* or completed Form FDA 1572
☐ 6. Protocol *(Continued)*
☐ d. Institutional Review Board data (21 CFR 312.23(a)(6)(iii)(b)) or completed Form FDA 1572
☐ 7. Chemistry, manufacturing, and control data *(21 CFR 312.23(a)(7))*
☐ Environmental assessment or claim for exclusion *(21 CFR 312.23(a)(7)(iv)(e))*
☐ 8. Pharmacology and toxicology data (21 CFR 312.23(a)(8))
☐ 9. Previous human experience *(21 CFR 312.23(a)(9))*
☐ 10. Additional information *(21 CFR 312.23(a)(10))*
☐ 11. Biosimilar User Fee Cover Sheet *(Form FDA 3792)*
☐ 12. Clinical Trials Certification of Compliance *(Form FDA 3674)*

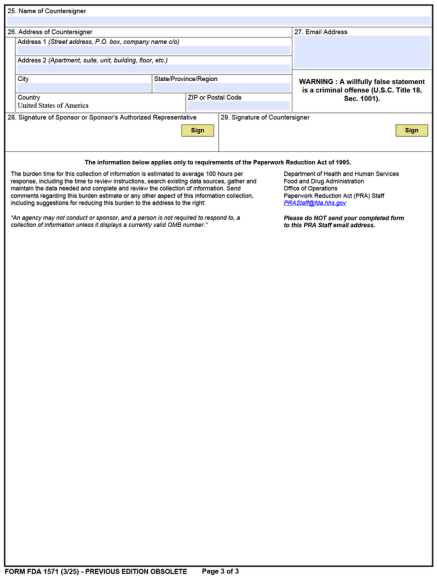
16. Is any part of the clinical study to be conducted by a contract research organization? (21 CFR 312.52) ☐ Yes ☐ No
If yes, use the Continuation Page to provide a statement containing the name and address of the CRO, identification of the clinical study and a listing of the obligations transferred (21 CFR 312.23(a)(1)(viii) and 21 CFR 312.52).
17. Name and Title of the person responsible for monitoring the conduct and progress of the clinical investigations
18. Name and Title of the person responsible for review and evaluation of information relevant to the safety of the drug

I agree not to begin clinical investigations until 30 days after FDA's receipt of the IND unless I receive earlier notification by FDA that the studies may begin. I also agree not to begin or continue clinical investigations covered by the IND if those studies are placed on clinical hold or financial hold. I agree that an Institutional Review Board (IRB) that complies with the requirements set forth in 21 CFR Part 56 will be responsible for initial and continuing review and approval of each of the studies in the proposed clinical investigation. I agree to conduct the investigation in accordance with all other applicable regulatory requirements.

19. Sponsor or Sponsor's Authorized Representative
Prefix | First Name | Middle | Last Name | Title
20. Telephone Number *(Include country code if applicable and area code)*
21. Facsimile (FAX) Number *(Include country code if applicable and area code)*
22. Address
Address 1 *(Street address, P.O. box, company name c/o)*
Address 2 *(Apartment, suite, unit, building, floor, etc.)*
City | State/Province/Region
Country | ZIP or Postal Code
23. Email Address
24. Date of Sponsor's Signature *(mm/dd/yyyy)*

FORM FDA 1571 (3/25) - PREVIOUS EDITION OBSOLETE Page 2 of 3

Figure 2.9.3: IND Form 1571 page 2. (FDA [2])


25. Name of Countersigner

26. Address of Countersigner

Address 1 (Street address, P.O. box, company name c/o)

Address 2 (Apartment, suite, unit, building, floor, etc.)

City

State/Province/Region

Country
United States of America

ZIP or Postal Code

27. Email Address

WARNING : A willfully false statement is a criminal offense (U.S.C. Title 18, Sec. 1001).

28. Signature of Sponsor or Sponsor's Authorized Representative

Sign

29. Signature of Countersigner

Sign

The information below applies only to requirements of the Paperwork Reduction Act of 1995.

The burden time for this collection of information is estimated to average 100 hours per response, including the time to review instructions, search existing data sources, gather and maintain the data needed and complete and review the collection of information. Send comments regarding this burden estimate or any other aspect of this information collection, including suggestions for reducing this burden to the address to the right:

"An agency may not conduct or sponsor, and a person is not required to respond to, a collection of information unless it displays a currently valid OMB number."

Department of Health and Human Services
Food and Drug Administration
Office of Operations
Paperwork Reduction Act (PRA) Staff
PRAStaff@fda.hhs.gov

Please do NOT send your completed form to this PRA Staff email address.

FORM FDA 1571 (3/25) - PREVIOUS EDITION OBSOLETE Page 3 of 3

Figure 2.9.4: IND Form 1571 page 3. (FDA [2])

Instructions for completing Form 1571

The images of the IND form in Figure 2.9.2, Figure 2.9.3, and Figure 2.9.4 are explained in detail below.

Field 1: SELECT APPROPRIATE CENTER

Note

CBER (FDA's Center for Biologics Evaluation and Research) is specified for a vaccine.

Field 2: NAME OF SPONSOR

- The sponsor is the person who takes responsibility for and initiates a clinical investigation. The sponsor may be an individual, pharmaceutical company, governmental agency, academic institution, private organization or other organization (21 CFR 312.3(b)).
- A Sponsor-Investigator is an individual who both initiates and conducts a clinical investigation and under whose immediate direction the investigational drug is being
- administered or dispensed (21 CFR 312.3(b)). For administrative reasons, only one individual should be designated as sponsor-investigator.
- If a pharmaceutical company will be supplying the drug, but will not itself be submitting the IND, the
- company is not the sponsor.

Field 3: DATE OF SUBMISSION

- Enter the date the submission is being submitted to the FDA. The date entered should match the date of the cover letter for the submission.

Fields 4–5: SPONSOR ADDRESS AND TELEPHONE NUMBER.

- Provide the address and telephone number of the sponsor identified in field 1 (21 CFR 312.23(a)(1)(i)). If this address is a post office box number, a street address should also be provided on line 2. The telephone number is the number where the sponsor is usually available during normal working hours.

Field 6: NAME(S) of DRUG

- For name(s) of drug (21 CFR 312.23(a)(1)(i)), list the established name(s) and trade name, if available. Also, provide the **dosage form**(s), and the unique ingredient identifier (UNII) term and code for active substances (if applicable). Use the Continuation Page if additional space is needed.

Field 7A: IND NUMBER

- Provide the IND number if it was previously assigned. If an IND number has not been assigned, leave the field blank. For IND numbers less than six digits, the IND number should be preceded using leading zeros (i.e., for IND 12345 enter 012345).

Field 7B: IND TYPE

- Select Commercial IND if the product under investigation is intended to be commercialized at a later date. (21 CFR 312.320).
- Select Research IND* if the product under investigation is not intended to be commercialized at a later date. Research INDs are generally sponsored by individual investigators, academic institutions and nonprofit entities. Includes INDs for emergency use or other expanded access. (21 CFR 312.310, 312.315 and 312.320).

Field 8A: PROPOSED INDICATION FOR USE

- Provide the proposed indication. Indicate if the proposed indication is for a rare disease (prevalence <200,000 U.S. patients). If the sponsor for the submission is the holder ofthe Orphan Designation Number, select “Yes” and provide the six-digit Orphan Designation Number in the appropriate field; if not, select “No.” Use the Continuation Page if there are more than one proposed indications for use by adding one indication per entry and providing rare disease/Orphan Drug Designation information for each entry, as applicable..

Field 8B: SNOMED CT INDICATION DISEASE TERM(S)

- For each original IND submission(including resubmissions to this submission type), provide the SNOMED
- CT coded disease term (e.g., 38341003 | Hypertensive disorder, systemic arterial (disorder) |) for the indication provided in Field 7A. To look up the indication’s SNOMED CT coded disease term:

1. Navigate to IHTSDOTOOLS.
2. Under Local Extensions, select ‘Go Browsing United States edition’.
3. Select ‘Search’ tab located in the upper left hand of page.
4. Enter the disease term in the search field.
5. Check the box ‘Group by concept’.
6. Select the single most appropriate term for the indication.
7. Select the ‘Expression’ tab located in the upper right hand of page.
8. Copy the entire text that appears under the heading ‘Pre-coordinated Expression.’
9. Paste the copied SNOMED CT disease term into Field 7B of Form FDA 1571.
10. For additional indications, use the continuation page for #7 and repeat these steps.

Field 9: PHASE(S) OF CLINICAL INVESTIGATION TO BE CONDUCTED

- Identification of the phase or phases of the clinical investigation to be conducted (21 CFR 312.23(a)(1)(ii)).

Field 10: CROSS REFERENCE

- It is necessary for the sponsor to submit certain information with an IND (such as manufacturing and controls information, pharmacology and toxicology data, or data from prior human studies) unless that information has previously been submitted to FDA. If the sponsor of the previously submitted information is not the same as the sponsor listed in field 2, the sponsor of the previously submitted information must provide a letter authorizing FDA to refer to the information (21 CFR 312.22(d)), (21 CFR 312.23 (b)). The sole exception to this requirement is when a marketed drug is used in the study, without modification to its approved packaging, in which case the marketed drug product must be identified by trade name, established name,

dosage form, strength, and lot number (21 CFR 312.23(a)(7)(d)).

Field 11: SERIAL NUMBER

- IND submissions should be consecutively numbered. The initial IND should be numbered "Serial Number: 0000." The next submission (e.g., amendment, report, or correspondence) should be numbered "Serial Number: 0001." Submissions should be numbered consecutively in the order in which they are submitted.

Field 12A: SUBMISSION INFORMATION

Checklist of what is included in the submission:

- Initial Investigational New Drug Application (IND): Should only be checked for an original IND submission. For subsequent submissions, check ALL the boxes below that apply because the submission may contain more than one type of information:
  - Response to Clinical Hold: A submission addressing deficiencies previously cited in a Clinical Hold Letter (21 CFR 312.42(e))
  - Response to FDA Request for Information: A submission containing responses to information
  - requests (21 CFR 312.41)
  - Request for Reactivation or Reinstatement: A request to resume clinical investigation under an IND placed on inactive status (21 CFR 312.45(d)); or, terminated by FDA (21 CFR 312.44(d))
- Annual Report: A brief report of the progress of the investigation submitted within 60 days of the anniversary date that the IND went into effect (21 CFR 312.33)
- General Correspondence: Any communication between the sponsor and FDA pertinent to the IND (21 CFR 312.41)
- Development Safety Update Report (DSUR): A report that provides information to assure that sponsors are adequately monitoring and evaluating the evolving safety profile of the investigational drug; may be used in place of the Annual Report
- Other: Any submission that does not fit in the other categories.

Protocol Amendment(s) check list:

- New Protocol: A protocol for a study not covered by a protocol already contained in the IND (21 CFR 312.30(a))
- Change in Protocol: A submission describing changes in a protocol (21 CFR 312.30(b)), including changes to investigators (21 CFR 312.30(e))
- New Investigator: A new investigator added to carry out a previously submitted protocol (21 CFR 312.30(c))
- HF Validation Protocol: A protocol for a human factors (HF) validation study
- Use Related Risk Analysis: Used to analyze product tasks, risks, harms, and associated mitigations for the proposed product, can be used to support design of a HF validation study, or to support a justification that an HF validation study does not need to be submitted for Agency review.

Information Amendment(s): Select the scientific review discipline(s) to which the submission applies (21 CFR 312.31)

- Chemistry/Microbiology
- Pharmacology/Toxicology
- Clinical/Safety
- Statistics
- Clinical Pharmacology

Request For: Select the type(s) of request(s) contained within the submission.

- Meeting
- Proprietary Name Review
- Special Protocol Assessment
- Formal Dispute Resolution

IND Safety Report(s):

- Initial Written Report: 21 CFR 312.32(c)
- Follow-up to a Written Report: 21 CFR 312.32(d)

Field 12B

If the submission contains Digital Health Technology (DHT) data or a proposal to collect DHT data, select 'Yes'. DHTs are systems that use computing platforms, connectivity, software, and/or sensors, (e.g., activity trackers, mobile medical applications) for remote data acquisition from participants in a clinical investigation. Although not final, FDA's guidance on DHTs represents the Agency's current thinking on the use of DHTs: FDA DHT information.

Field 13: COMBINATION PRODUCTS

Field 12A (Submission Information) should be filled out before attempting to fill out Field 13. Field 13 is only fillable for original submissions [i.e., the Initial Investigational New Drug Application (IND)]. Indicate if the product proposed within the submission is a combination product [e.g., drug-device, drug-biological product, drug-device-biological product, see 21 CFR 3.2(e)] by selecting 'Yes' and entering the number below that best identifies the type:

1. Convenience Kit or Co-Package
2. Prefilled Drug Delivery Device/System
3. Prefilled Biologic Delivery Device/System
4. Device Coated/Impregnated/Otherwise Combined with Drug
5. Device Coated or Otherwise Combined with Biologic
6. Drug/Biologic Combination
7. Separate Products Requiring Cross Labeling
8. Possible Combination Based on Cross Labeling of Separate Products
9. Other Type of Part 3 Combination Product (e.g., Drug/Device/Biological Product)

If the product in the submission is not a combination product, select 'No' and leave the Combination Product Type blank. If the submission relates to both combination product and non-combination product configurations (e.g., the submission relates to a drug product in a vial and is also in a pre-filled syringe presentation), select 'Yes' and identify the type that applies to the combination product configuration. If the submission relates to a product configuration that has multiple combination product types (e.g., a submission
is for a product that contains a prefilled syringe (Type 2) and is co-packaged with other devices and drug (Type 1)), select type '9'. If this is the initial submission for a product for which a Request for Designation (RFD) was submitted, provide the six-digit RFD number.

Field 14. SUBMISSIONS REQUIRING A JUSTIFICATION STATEMENT

Select the following only if applicable. A justification statement must be submitted with the application for any items selected. Refer to the cited CFR section for further information.
Expanded Access Use, 21 CFR 312.300: Note that Treatment INDs and Treatment Protocols are not intended for single patient use. Before checking this box, the sponsor should be thoroughly familiar with the cited regulations and contact the appropriate FDA review division to discuss the proposed treatment use (21 CFR 312.320).

Field 15: CONTENTS OF APPLICATION

This section contains items 1 through 12 which is a checklist that should be used to indicate the types of information contained within a particular application or submission. Check all that apply.

- For a Sponsor-Investigator IND, Items 2–4 may be briefly addressed in the cover letter or in a summary.
- When the investigational drug is obtained from a supplier in a final dosage form, Items 5, 7, 8 and 9 may be referenced if authorization is given by the supplier (see explanation in field 9 above). If the investigational drug is prepared or altered in any way after shipment by the supplier, complete manufacturing (or compounding) and controls information, including information on sterility and pyrogenicity testing for parenteral drugs, must be submitted for that process in Item 7.
- Item 6 **requires that a protocol be submitted**, along with information on the investigators, facilities, and Institutional Review Board. Completed Form(s) FDA 1572 with attachments would suffice for Items 6 b–d.
- Item 7 also requires submission of either a claim of categorical exclusion from the requirement to submit an environmental assessment (21 CFR 25.15(a)). When claiming a categorical exclusion, the sponsor should include the following statements: "I claim categorical exclusion (under 21 CFR 25.31(e)) for the study(ies) under this IND. To my knowledge, no extraordinary

circumstances exist."

- In certain applications, information on special topics may be needed (Item 10). Refer to the cited CFR section for further information. Additional information may also include the Generic Drug User Fee (GDUFA) Coversheet (Form FDA 3794) for GDUFA Master Files (Type II APIs), when applicable.
- Include the Biosimilar User Fee Cover Sheet (Form FDA 3792) (Item 11) and / or the Clinical Trials Certification of Compliance (Form FDA 3674) (Item 12) as applicable.

Note

The substantive items include:

- General investigational plan
- Protocol (study protocol, investigator data, and facilities data)
- Chemistry, manufacturing, and control data
- Pharmacology and toxicology data

Field 16

CONTRACT RESEARCH ORGANIZATIONS (CROs) Check the appropriate box to indicate if the clinical study will be conducted by a CRO. Use the Continuation Page to provide a statement containing the name and address of the CRO, identification of the clinical study and a listing of the obligations transferred (21 CFR 312.23(a)(1)(viii)).

Field 17

Provide the Name and Title of the person responsible for monitoring the conduct and progress of the clinical investigations (21 CFR 312.23(a)(1)(vi)). For Sponsor-Investigator INDs, the investigator has this responsibility.

Field 18

Provide the Name(s) and Title(s) of the person(s) responsible under 21 CFR 312.32 for review and evaluation of information relevant to the safety of the drug (21 CFR 312.23(a)(1)(vii)). For Sponsor-Investigator INDs, the investigator has this responsibility. Certain important commitments that the IND sponsor makes by signing Form FDA 1571 are listed below field 17. Under Section 744G (11) of the Federal Food, Drug, and Cosmetic Act (the FD&C Act), as added by the Biosimilar User Fee Act of 2012 (BsUFA), the term "financial hold" means an order issued by FDA to prohibit the sponsor of a clinical investigation from continuing the investigation if FDA determines that the
investigation is intended to support a biosimilar biological product application and the sponsor has failed to pay any required initial biosimilar biological product development (BPD) fee, annual BPD fee, or reactivation fee. The term "financial hold" does not mean that any of the bases for a clinical hold identified in section 505(i)(3) of the FD&C Act have been determined by FDA to exist concerning the investigation.

Field 19: NAME OF SPONSOR OR SPONSOR'S AUTHORIZED REPRESENTATIVE

For a sponsor-investigator IND, the sponsor-investigator should be named and must sign the form. For an IND sponsored by a pharmaceutical firm or research organization, the name of the sponsor's authorizing representative should be entered and that individual must sign the form.

Fields 20–22

- Provide the telephone number, facsimile number, and full mailing address of the individual identified in field 19.

Field 23

- Provide the email address of the person identified in field 19. For INDs submitted to the Center for Biologics Evaluation

and Research (CBER), a specific statement authorizing communication via non-secure email should be included in the cover letter as applicable.

Field 24

Provide the date the form is signed by the sponsor or sponsor's authorized representative. This date may be different from the date provided in field 3.

Field 25: NAME OF COUNTERSIGNER

If the person signing the application in field 28 does not reside or have a place of business within the United States, the submission must be countersigned by an attorney, agent, or other authorized official who resides or maintains a place of business within the United States (21 CFR 312.23(a)(1)(ix)). If applicable, provide the name of the attorney, agent or other authorized official countersigning the application in field 29.

Field 26: ADDRESS OF COUNTERSIGNER

- If applicable, provide the full mailing address of the individual identified in field 25. If the sponsor or sponsor's authorized representative identified in field 18 does not reside or have a place of business within the United States, this mailing address is the address to which written correspondence from the FDA should be directed.

Field 27: EMAIL ADDRESS

- Enter the email address of the Countersigner named in Field 25.

Field 28: SIGNATURE OF SPONSOR OR SPONSOR'S AUTHORIZED REPRESENTATIVE

- The person identified in field 19 must sign in this field (21 CFR 312.23(a)(1)(ix)).

Field 29: SIGNATURE OF COUNTERSIGNER

- If applicable, the person identified in field 25 must countersign in this field.

References

1. FDA IND. Information downloaded January 7, 2026.
2. FDA IND Form 1571 link. Form FDA 1571: Investigational New Drug Application. Information downloaded January 7, 2026.
3. FDA IND Application. Accessed January 21, 2026.
4. FDA IND Form 1571 instructions. Food and Drug Administration. "Instructions for Filling out Form FDA 1571 Investigational New Drug Application (IND). Accessed January 21, 2026.
5. Food and Drug Administration, FDA Vaccine Approval Process (for the Illustration of the Vaccine Approval Process.)

This page titled 2.9: Investigational New Drug (IND) Application is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

2.10: Biologics License Application (BLA)

Note

The information has been extracted verbatim from Reference 1, with some minor edits.

Added FDA arrow.

Biologics License Applications (BLA) Process (CBER)

The Biologics License Application (BLA) is a request for permission to introduce, or deliver for introduction, a biologic product into interstate commerce (21 CFR 601.2). The BLA is regulated under 21 CFR 600 – 680. A BLA is submitted by any legal person or entity who is engaged in manufacture or an applicant for a license who takes responsibility for compliance with product and establishment standards. Form 356h specifies the requirements for a BLA. This includes:

- Applicant information
- Product/Manufacturing information
- Pre-clinical studies
- Clinical studies
- Labeling

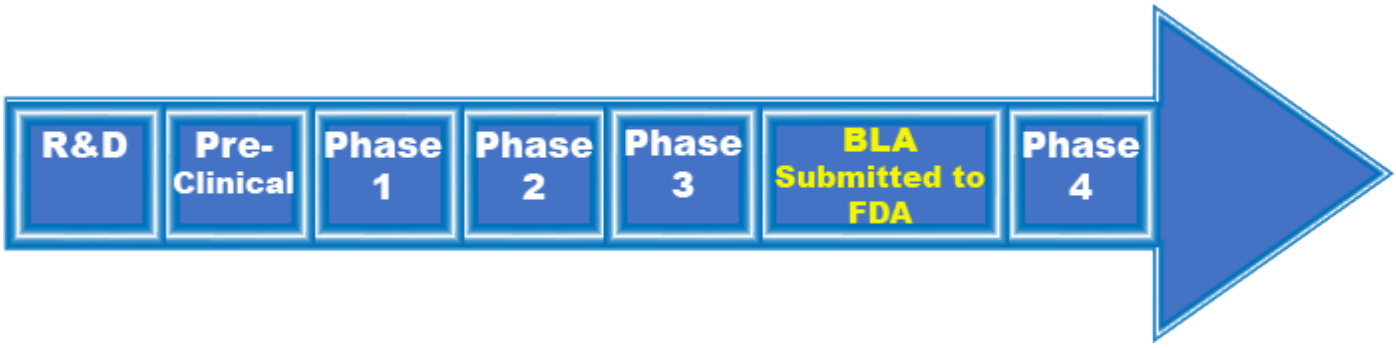


Figure 2.10.1: Vaccine Approval Process. Includes R&D, Pre-Clinical, Phase 1, Phase 2, Phase 3, BLA Submitted to FDA, and Phase 4. (FDA, public domain [4])

Note

The overall process for vaccine development and approval includes:

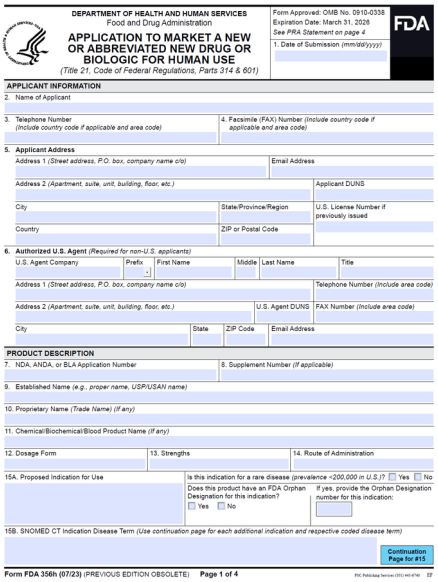
- Research and Development
- Pre-clinical trials with non-humans
- The filing of an Investigational New Drug (IND) application with the FDA at the conclusion of pre-clinical trials requesting proceeding to in-human testing
- Pre-licensure vaccine Clinical Trials (Phases 1, 2, and 3)
- Filing a Biologics License Application (BLA) at the conclusion of Clinical Trial Phase 3 with the FDA
- Inspection of the pharmaceutical company's manufacturing facility or facilities and achieving approval
- Presenting the pharmaceutical company's findings to the FDA's Vaccines and Related Biological Products Advisory Committee (VRBPAC)
- Usability testing of the vaccine product labeling.

Form 356h

On the FDA web site, one can find the instructions to fill in the form as well as the form itself.

Since the BLA is a project deliverable, let's now take a look at the actual four page form. It will provide insight into which team members need to make contributions in completing the form and listing those team members as resources in preparing the project plan.

Page 1



DEPARTMENT OF HEALTH AND HUMAN SERVICES
Food and Drug Administration

APPLICATION TO MARKET A NEW OR ABBREVIATED NEW DRUG OR BIOLOGIC FOR HUMAN USE
(Title 21, Code of Federal Regulations, Parts 314 & 601)

Form Approved: OMB No. 0910-0338
Expiration Date: March 31, 2026
See PRA Statement on page 4

FDA

1. Date of Submission (mm/dd/yyyy)

APPLICANT INFORMATION

2. Name of Applicant

3. Telephone Number (Include country code if applicable and area code)

4. Facsimile (FAX) Number (Include country code if applicable and area code)

5. Applicant Address

Address 1 (Street address, P.O. box, company name c/o)

Email Address

Address 2 (Apartment, suite, unit, building, floor, etc.)

Applicant DUNS

City

State/Province/Region

U.S. License Number if previously issued

Country

ZIP or Postal Code

6. Authorized U.S. Agent (Required for non-U.S. applicants)

U.S. Agent Company

Prefix

First Name

Middle

Last Name

Title

Address 1 (Street address, P.O. box, company name c/o)

Telephone Number (Include area code)

Address 2 (Apartment, suite, unit, building, floor, etc.)

U.S. Agent DUNS

FAX Number (Include area code)

City

State

ZIP Code

Email Address

PRODUCT DESCRIPTION

7. NDA, ANDA, or BLA Application Number

8. Supplement Number (If applicable)

9. Established Name (e.g., proper name, USP/USAN name)

10. Proprietary Name (Trade Name) (If any)

11. Chemical/Biochemical/Blood Product Name (If any)

12. Dosage Form

13. Strengths

14. Route of Administration

15A. Proposed Indication for Use

Is this indication for a rare disease (prevalence <200,000 in U.S.)? Yes No

Does this product have an FDA Orphan Designation for this indication? Yes No

If yes, provide the Orphan Designation number for this indication:

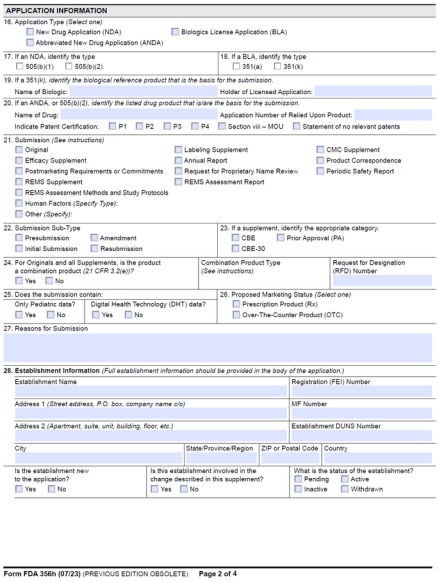
15B. SNOMED CT Indication Disease Term (Use continuation page for each additional indication and respective coded disease term)

Continuation Page for #15

Form FDA 356h (07/23) (PREVIOUS EDITION OBSOLETE) Page 1 of 4

Figure 2.10.2: FDA BLA Form 356h page 1. (FDA, public domain[2])

Page 2



APPLICATION INFORMATION

16. Application Type (Select one)
New Drug Application (NDA)
Biologics License Application (BLA)
Abbreviated New Drug Application (ANDA)

17. If an NDA, identify the type
505(b)(1) 505(b)(2)

18. If a BLA, identify the type
351(a) 351(k)

19. If a 351(k), identify the biological reference product that is the basis for the submission.
Name of Biologic
Holder of Licensed Application:

20. If an ANDA, or 505(b)(2), identify the listed drug product that is/are the basis for the submission.
Name of Drug:
Application Number of Relied Upon Product:
Indicate Patent Certification: P1 P2 P3 P4 Section viii – MOU Statement of no relevant patents

21. Submission (See instructions)
Original
Efficacy Supplement
Postmarketing Requirements or Commitments
REMS Supplement
REMS Assessment Methods and Study Protocols
Human Factors (Specify Type):
Other (Specify):
Labeling Supplement
Annual Report
Request for Proprietary Name Review
REMS Assessment Report
CMC Supplement
Product Correspondence
Periodic Safety Report

22. Submission Sub-Type
Presubmission Amendment
Initial Submission Resubmission

23. If a supplement, identify the appropriate category.
CBE Prior Approval (PA)
CBE-30

24. For Originals and all Supplements, is the product a combination product (21 CFR 3.2(e))?
Yes No

Combination Product Type (See instructions)

Request for Designation (RFD) Number

25. Does the submission contain:
Only Pediatric data? Yes No
Digital Health Technology (DHT) data? Yes No

26. Proposed Marketing Status (Select one)
Prescription Product (Rx)
Over-The-Counter Product (OTC)

27. Reasons for Submission

28. Establishment Information (Full establishment information should be provided in the body of the application.)

Establishment Name

Registration (FEI) Number

Address 1 (Street address, P.O. box, company name c/o)

MF Number

Address 2 (Apartment, suite, unit, building, floor, etc.)

Establishment DUNS Number

City

State/Province/Region

ZIP or Postal Code

Country

Is the establishment new to the application? Yes No

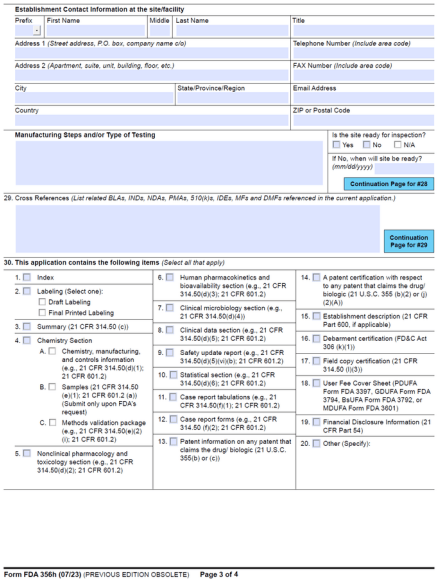
Is this establishment involved in the change described in this supplement? Yes No

What is the status of the establishment?
Pending Active
Inactive Withdrawn

Form FDA 356h (07/23) (PREVIOUS EDITION OBSOLETE) Page 2 of 4

Figure 2.10.3: FDA BLA Form 356h page 2. (FDA, public domain[2])

Page 3



Establishment Contact Information at the site/facility

Prefix | First Name | Middle | Last Name | Title

Address 1 (Street address, P.O. box, company name c/o) | Telephone Number (Include area code)

Address 2 (Apartment, suite, unit, building, floor, etc.) | FAX Number (Include area code)

City | State/Province/Region | Email Address

Country | ZIP or Postal Code

Manufacturing Steps and/or Type of Testing

Is the site ready for inspection? ☐ Yes ☐ No ☐ N/A

If No, when will site be ready? (mm/dd/yyyy)

Continuation Page for #28

29. Cross References (List related BLAs, INDs, NDAs, PMAs, 510(k)s, IDEs, MFs and DMFs referenced in the current application.)

Continuation Page for #29

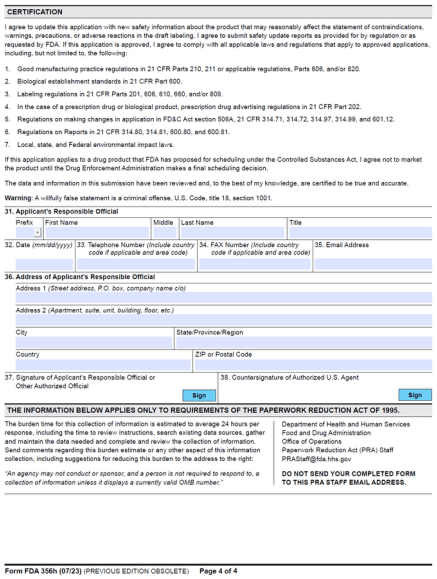
30. This application contains the following items (Select all that apply)

1. ☐ Index
2. ☐ Labeling (Select one): ☐ Draft Labeling ☐ Final Printed Labeling
3. ☐ Summary (21 CFR 314.50 (c))
4. ☐ Chemistry Section
   A. ☐ Chemistry, manufacturing, and controls information (e.g., 21 CFR 314.50(d)(1); 21 CFR 601.2)
   B. ☐ Samples (21 CFR 314.50 (e)(1); 21 CFR 601.2 (a)) (Submit only upon FDA's request)
   C. ☐ Methods validation package (e.g., 21 CFR 314.50(e)(2)(i); 21 CFR 601.2)
5. ☐ Nonclinical pharmacology and toxicology section (e.g., 21 CFR 314.50(d)(2); 21 CFR 601.2)
6. ☐ Human pharmacokinetics and bioavailability section (e.g., 21 CFR 314.50(d)(3); 21 CFR 601.2)
7. ☐ Clinical microbiology section (e.g., 21 CFR 314.50(d)(4))
8. ☐ Clinical data section (e.g., 21 CFR 314.50(d)(5); 21 CFR 601.2)
9. ☐ Safety update report (e.g., 21 CFR 314.50(d)(5)(vi)(b); 21 CFR 601.2)
10. ☐ Statistical section (e.g., 21 CFR 314.50(d)(6); 21 CFR 601.2)
11. ☐ Case report tabulations (e.g., 21 CFR 314.50(f)(1); 21 CFR 601.2)
12. ☐ Case report forms (e.g., 21 CFR 314.50 (f)(2); 21 CFR 601.2)
13. ☐ Patent information on any patent that claims the drug/ biologic (21 U.S.C. 355(b) or (c))
14. ☐ A patent certification with respect to any patent that claims the drug/ biologic (21 U.S.C. 355 (b)(2) or (j)(2)(A))
15. ☐ Establishment description (21 CFR Part 600, if applicable)
16. ☐ Debarment certification (FD&C Act 306 (k)(1))
17. ☐ Field copy certification (21 CFR 314.50 (l)(3))
18. ☐ User Fee Cover Sheet (PDUFA Form FDA 3397, GDUFA Form FDA 3794, BsUFA Form FDA 3792, or MDUFA Form FDA 3601)
19. ☐ Financial Disclosure Information (21 CFR Part 54)
20. ☐ Other (Specify):

Form FDA 356h (07/23) (PREVIOUS EDITION OBSOLETE) Page 3 of 4

Figure 2.10.4: FDA BLA Form 356h page 3. (FDA, public domain[2])

Page 4



CERTIFICATION

I agree to update this application with new safety information about the product that may reasonably affect the statement of contraindications, warnings, precautions, or adverse reactions in the draft labeling. I agree to submit safety update reports as provided for by regulation or as requested by FDA. If this application is approved, I agree to comply with all applicable laws and regulations that apply to approved applications, including, but not limited to, the following:

1. Good manufacturing practice regulations in 21 CFR Parts 210, 211 or applicable regulations, Parts 606, and/or 820.
2. Biological establishment standards in 21 CFR Part 600.
3. Labeling regulations in 21 CFR Parts 201, 606, 610, 660, and/or 809.
4. In the case of a prescription drug or biological product, prescription drug advertising regulations in 21 CFR Part 202.
5. Regulations on making changes in application in FD&C Act section 506A, 21 CFR 314.71, 314.72, 314.97, 314.99, and 601.12.
6. Regulations on Reports in 21 CFR 314.80, 314.81, 600.80, and 600.81.
7. Local, state, and Federal environmental impact laws.

If this application applies to a drug product that FDA has proposed for scheduling under the Controlled Substances Act, I agree not to market the product until the Drug Enforcement Administration makes a final scheduling decision.

The data and information in this submission have been reviewed and, to the best of my knowledge, are certified to be true and accurate.

Warning: A willfully false statement is a criminal offense, U.S. Code, title 18, section 1001.

31. Applicant's Responsible Official

Prefix | First Name | Middle | Last Name | Title

32. Date (mm/dd/yyyy) | 33. Telephone Number (Include country code if applicable and area code) | 34. FAX Number (Include country code if applicable and area code) | 35. Email Address

36. Address of Applicant's Responsible Official

Address 1 (Street address, P.O. box, company name c/o)

Address 2 (Apartment, suite, unit, building, floor, etc.)

City | State/Province/Region

Country | ZIP or Postal Code

37. Signature of Applicant's Responsible Official or Other Authorized Official — Sign

38. Countersignature of Authorized U.S. Agent — Sign

THE INFORMATION BELOW APPLIES ONLY TO REQUIREMENTS OF THE PAPERWORK REDUCTION ACT OF 1995.

The burden time for this collection of information is estimated to average 24 hours per response, including the time to review instructions, search existing data sources, gather and maintain the data needed and complete and review the collection of information. Send comments regarding this burden estimate or any other aspect of this information collection, including suggestions for reducing this burden to the address to the right:

"An agency may not conduct or sponsor, and a person is not required to respond to, a collection of information unless it displays a currently valid OMB number."

Department of Health and Human Services
Food and Drug Administration
Office of Operations
Paperwork Reduction Act (PRA) Staff
PRAStaff@fda.hhs.gov

DO NOT SEND YOUR COMPLETED FORM TO THIS PRA STAFF EMAIL ADDRESS.

Form FDA 356h (07/23) (PREVIOUS EDITION OBSOLETE) Page 4 of 4

Figure 2.10.5: FDA BLA Form 356h page 4. (FDA, public domain[2])

Two of the salient items are Field 21 and Field 30 described in detail below.[3]

Field 21: Submission Type

- Original: An application for which FDA has never issued an approval letter.
- Labeling Supplement: A supplemental application for labeling changes to an approved product as described under 21 CFR 314.70 and 21 CFR 601.12 that does not otherwise qualify as another type of supplement (e.g., Efficacy, CMC, REMS).
- CMC Supplement: A supplemental application for chemistry, manufacturing, and control (CMC) changes to an approved product as described under 21 CFR 314.70, 21 CFR 314.71, 21 CFR 314.72, and 21 CFR 601.12, including CMC supplements with corresponding labeling changes.
- Efficacy Supplement: A supplemental application for changes to an approved product from among the following changes: new or modified indication; new or revised dose or dosing regimen; new route of administration; comparative efficacy claim naming another product; significant alteration in the patient population; switch of marketing status from prescription to over-the-counter use; traditional approval of a product originally approved under Subpart H (Accelerated Approval) or Subpart I

(Animal Rule); labeling or manufacturing change requiring clinical data for approval.
- Annual Report: See 21 CFR 314.81(b)(2) for NDAs and 21 CFR 601.12(d) for BLAs.
- Product Correspondence: Any communication or general correspondence related to an application (e.g., routine administrative changes, donor re-entry requests, lot distribution reports, license reissuance requests, meeting requests) that is not an amendment to a pending application. Provide a description of the content or intent of the Product Correspondence in Field 27 (Reasons for the Submission).
- REMS Assessment Methodologies and Study Protocols: A submission containing information on methodological approaches and study protocols use to assess a REMS program.
- REMS Assessment Report: A submission containing information and data to support if the goal of each strategy is being met; modifications to the REMS or revisions to the REMS assessment plan are needed, including the timing of the REMS assessments; and whether the REMS is still necessary to ensure the benefits outweigh the risks of the drug.
- REMS Supplement: A supplemental application proposing a new Risk Evaluation and Mitigation strategy (REMS) or modifications (major and/or minor) to an approved REMS.
- Post Marketing Requirements or Commitments: A submission containing information related to post marketing requirements or commitments (e.g., nonclinical protocol, final study report).
- Periodic Safety Report: Periodic reports (Periodic Adverse Drug Experience Reports (PADERs)) of adverse drug or biological product experience as described under 21 CFR 314.80(c)(2), 21 CFR 314.98, and 21 CFR 600.80(c)(2), including those in Periodic Safety Update Report (PSUR) format.
- Request for Proprietary Name Review: A submission containing a request for proprietary name review.
- Human Factors (Specify Type): Select check box and specify what type of HF information is being submitted in “Specify Type” (e.g. HF Protocol, HF Study Report, Use-Related Risk Analysis, Justification for no HF Validation Study, Comparative Analysis, etc.).
- Other (specify): State the submission type if it is not one of the previous submission types listed above (e.g., formal dispute resolution request, REMS revisions). If this box is checked, provide the Reasons for the Submission in Field 27.

Field 30 (application contents)

This section contains items 1 through 20 which is a checklist that should be used to indicate the types of information contained within a particular application or submission.

1. Index
2. Labelling
3. Summary
4. Chemistry Section
   A. Chemistry, manufacturing, and controls information
   B. Samples
   C. Methods validation package
5. Nonclinical pharmacology and toxicology section
6. Human pharmacokinetics and bioavailability section
7. Clinical microbiology section
8. Clinical data section
9. Safety update report
10. Statistical section
11. Case report tabulations
12. Case report forms
13. Patent information
14. A patent certification
15. Establishment description
16. Debarment certification
17. Field copy certification
18. User Fee Cover Sheet
19. Financial Disclosure information
20. Other

A complete index or table of contents should immediately follow the Form FDA 356h and, if applicable, a User Fee Cover Sheet (Forms FDA 3397, 3792, or 3794). Note that the CFR references are provided for most items in order to indicate what type of information should or must be submitted in each section. For further information, the applicant may consult the guidance documents that are available from the Agency. Please note, selecting Field 30, item 15 indicates the complete Biological Products

establishment description is included at Field 28.

References

1. FDA BLA Application process. Information downloaded January 7, 2026.
2. FDA Forms BLA. Information downloaded January 7, 2026.
3. INSTRUCTIONS FOR FILLING OUT FORM FDA 356h – APPLICATION TO MARKET A NEW OR ABBREVIATED NEW DRUG OR BIOLOGIC FOR HUMAN USE, FORM FDA 356h SUPPLEMENT (05/23) – INSTRUCTIONS, FDA 356h instructions. Accessed January 7, 2026.
4. Food and Drug Administration, FDA Vaccine Approval Process arrow (for the Illustration of the Vaccine Approval Process.)

This page titled 2.10: Biologics License Application (BLA) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

2.11: Emergency Use Authorization (EUA)

Note

The material in this section has been extracted from an FDA site almost verbatim, with some edits. The material is in the public domain and this textbook provides attribution to the FDA at the end of this section.

About Emergency Use Authorizations (EUAs)

The Emergency Use Authorization (EUA) authority allows FDA to help strengthen the nation's public health protections against chemical, biological, radiological, and nuclear (CBRN) threats including infectious diseases, by facilitating the availability and use of medical countermeasures (MCMs) needed during public health emergencies.

Under section 564 of the Federal Food, Drug, and Cosmetic Act (FD&C Act), when the Secretary of HHS declares that an emergency use authorization is appropriate, FDA may authorize unapproved medical products or unapproved uses of approved medical products to be used in an emergency to diagnose, treat, or prevent serious or life-threatening diseases or conditions caused by CBRN threat agents when certain criteria are met, including there are no adequate, approved, and available alternatives.

The HHS EUA Declaration to support such use must be based on one of four types of determinations of threats or potential threats by the Secretary of HHS, Homeland Security, or Defense. HHS EUA Declarations, and associated determinations, supporting use of EUAs are described below. Please note that an HHS EUA Declaration is distinct from, and not dependent on, a Public Health Emergency (PHE) Declaration under section 319 of the PHS Act.

Note

The four types of determinations are referenced in the HHS Secretary issued 4 COVID-19 EUA Declarations cited below.

Under Section 564 of the FD&C Act, 21 U.S.C. 360bbb-3(opens in new window), the Commissioner of the Food and Drug Administration (FDA), acting under delegated authority from the Secretary of HHS, may issue an Emergency Use Authorization (EUA) authorizing:

1. the emergency use of an unapproved drug, an unapproved or uncleared device, or an unlicensed biological product; or
2. an unapproved use of an approved drug, approved or cleared device, or licensed biological product.

Before an EUA may be issued, the Secretary of HHS must declare that circumstances exist justifying the authorization based on one of four determinations:

1. a determination by the Secretary of Homeland Security that there is a domestic emergency, or a significant potential for a domestic emergency, involving a heightened risk of attack with a, chemical, biological, radiological, or nuclear ("CBRN") agent or agents;
2. the identification of a material threat by the Secretary of Homeland Security pursuant to section 319F-2 of the Public Health Service (PHS) Act [1] sufficient to affect national security or the health and security of United States citizens living abroad;
3. a determination by the Secretary of Defense that there is a military emergency, or a significant potential for a military emergency, involving a heightened risk to United States military forces, including personnel operating under the authority of title 10 or title 50, of attack with (i) a biological, chemical, radiological, or nuclear agent or agents; or (ii) an agent or agents that may cause, or are otherwise associated with, an imminently life-threatening and specific risk to United States military forces; or

4. a determination by the Secretary that there is a public health emergency, or a significant potential for a public health emergency, that affects, or has a significant potential to affect, national security or the health and security of United States citizens living abroad, and that involves a CBRN agent or agents, or a disease or condition that may be attributable to such agent or agents.[2]

Public Readiness and Emergency Preparedness Act (PREP Act)

The PREP Act amended the Public Health Service Act (PHS Act) to add section 319F-3 (42 U.S.C. 247d-6d) to allow HHS to issue a declaration to provide immunity (except for willful misconduct) for activities related to administration or use of 'covered countermeasures,' defined to include medical products authorized for emergency use. This PREP Act Declaration is distinct from, and not dependent on, HHS EUA Declarations or section 319 PHE Declarations.

Coronavirus Disease 2019 (COVID-19) EUA Information

- **Coronavirus Disease (COVID-19) updates from FDA**
- **The HHS Secretary issued 4 COVID-19 EUA Declarations:**
  - EUA Declaration (effective February 4, 2020) that circumstances exist justifying the authorization of emergency use of in vitro diagnostics for detection and/or diagnosis of the virus that causes COVID-19
  - EUA Declaration (effective March 2, 2020) that circumstances exist justifying the authorization of emergency use of personal respiratory protective devices during the COVID-19 outbreak
  - EUA Declaration (effective March 24, 2020) that circumstances exist justifying the authorization of emergency use of medical devices, including alternative products used as medical devices, due to shortages during the COVID-19 outbreak
  - EUA Declaration (effective March 27, 2020) for drugs and biologic products during the COVID-19 pandemic
- **The HHS COVID-19 EUA Declarations were based on an HHS COVID-19 Determination:**
  - Determination of a Public Health Emergency Justifying Authorizations Pursuant to Section 564(b) of the FD&C Act (February 4, 2020)
  - Amended Determination of a Public Health Emergency or Significant Potential for a Public Health Emergency Pursuant to Section 564(b) of the FD&C Act (effective March 15, 2023; Federal Register notice March 20, 2023

Note

The Federal Register covered the initial two EUAs issued for COVID-19 vaccines as follows:

> The Food and Drug Administration (FDA) is announcing the issuance of two Emergency Use Authorizations (EUAs) (the Authorizations) under the Federal Food, Drug, and Cosmetic Act (FD&C Act) for biological products for use during the COVID-19 pandemic. FDA issued one Authorization for a biological product as requested by Pfizer, Inc, and one Authorization for a biological product as requested by ModernaTX, Inc. The Authorizations contain, among other things, conditions on the emergency use of the authorized products. The Authorizations follow the February 4, 2020, determination by the Secretary of Health and Human Services (HHS) that there is a public health emergency that has a significant potential to affect national security or the health and security of U.S. citizens living abroad and that involves a novel (new) coronavirus. The virus, now named SARS-CoV-2, causes the illness COVID-19. On the basis of such determination, the Secretary of HHS declared on March 27, 2020, that circumstances exist justifying the authorization of emergency use of drugs and biological products during the COVID-19 pandemic, pursuant to the FD&C Act, subject to the terms of any authorization issued under that section. The Authorizations, which include an explanation of the reasons for issuance, are reprinted in this document.[3]
>
> The Authorization for Pfizer, Inc. is effective as of December 11, 2020; the Authorization for ModernaTX, Inc. is effective as of December 18, 2020.[3]

Attributions/References (public domain material)

1. FDA Emergency Use Authorization. Accessed January 18, 2026.
2. Federal Register Health Emergency . Accessed January 18, 2026.

3. Federal Register EUA Authorizations. January 19, 2021. Accessed January 18, 2026.

Note

FDA website policies. Downloaded January 7, 2026.

Linking to or Copying Information on the FDA Website

Unless otherwise noted, the contents of the FDA website (www.fda.gov) — both text and graphics — are not copyrighted. They are in the public domain and may be republished, reprinted and otherwise used freely by anyone without the need to obtain permission from FDA. Credit to the U.S. Food and Drug Administration as the source is appreciated but not required.

People are also free to link to any URL on FDA's site. FDA's preference is that people link to the material on the FDA site (rather than copying it to their personal websites) because the agency continuously updates the information on the website as better information becomes available. A person copying documents to another website, instead of linking to them, would then have to monitor the original documents to know when these documents were updated by FDA or else risk giving bad or incorrect advice to visitors to their website. Providing consumers or health professionals with advice that is not fully up to date can lead to serious public health consequences. Providing industry advice that is not fully up to date can lead to companies being out of compliance with regulatory requirements.

If a person, nonetheless, decides to copy content or images, FDA strongly recommends that the copied item lists the date that the material was copied and provides a link back to its source on the FDA website. Users can then see for themselves if the copied material has been updated or changed.

This page titled 2.11: Emergency Use Authorization (EUA) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

2.12: Current Good Manufacturing Practice (CGMP)

Note

Information extracted verbatim from Reference 1.

The Food and Drug Administration (FDA) regulates the quality of pharmaceuticals very carefully.

The main regulatory standard for ensuring pharmaceutical quality is the Current Good Manufacturing Practice (CGMP) regulations for human pharmaceuticals. Consumers expect that each batch of medicines they take will meet quality standards so that they will be safe and effective. Most people, however, are not aware of CGMP, or how FDA assures that drug manufacturing processes meet these basic objectives. Recently, FDA has announced a number of regulatory actions taken against drug manufacturers based on the lack of CGMP.

This section discusses some facts that may be helpful in understanding how CGMP establishes the foundation for drug product quality.

What is CGMP?

CGMP refers to the Current Good Manufacturing Practice regulations enforced by the FDA. CGMP provides for systems that assure proper design, monitoring, and control of manufacturing processes and facilities. Adherence to the CGMP regulations assures the identity, strength, quality, and purity of drug products by requiring that manufacturers of medications adequately control manufacturing operations. This includes establishing strong quality management systems, obtaining appropriate quality raw materials, establishing robust operating procedures, detecting and investigating product quality deviations, and maintaining reliable testing laboratories. This formal system of controls at a pharmaceutical company, if adequately put into practice, helps to prevent instances of contamination, mix-ups, deviations, failures, and errors. This assures that drug products meet their quality standards.

The CGMP requirements were established to be flexible in order to allow each manufacturer to decide individually how to best implement the necessary controls by using scientifically sound design, processing methods, and testing procedures. The flexibility in these regulations allows companies to use modern technologies and innovative approaches to achieve higher quality through continual improvement. Accordingly, the "C" in CGMP stands for "current," requiring companies to use technologies and systems that are up-to-date in order to comply with the regulations. Systems and equipment that may have been "top-of-the-line" to prevent contamination, mix-ups, and errors 10 or 20 years ago may be less than adequate by today's standards.

It is important to note that CGMP regulations for drugs contain the minimum requirements. Many pharmaceutical manufacturers are already implementing comprehensive, modern quality systems and risk management approaches that exceed these minimum standards.

Note

(Information in this Note is extracted verbatim from Reference 2)

Code of Federal Regulations (CFR)

FDA's portion of the *CFR* is in Title 21, which interprets the *Federal Food, Drug and Cosmetic Act* and related statutes, including the Public Health Service Act. The pharmaceutical or drug quality-related regulations appear in several parts of Title 21, including sections in parts 1-99, 200-299, 300-499, 600-799, and 800-1299.

The regulations enable a common understanding of the regulatory process by describing the requirements to be followed by drug manufacturers, applicants, and FDA.

- 21 CFR Part 314. For FDA approval to market a new drug.

- 21 CFR Part 210. Current Good Manufacturing Practice in Manufacturing Processing, packing, or Holding of Drugs.
- 21 CFR Part 211. Current Good Manufacturing Practice for Finished Pharmaceuticals.
- 21 CFR Part 212. Current Good Manufacturing Practice for Positron Emission Tomography Drugs.
- 21 CFR Part 600. Biological Products: General.

Why is CGMP so important?

A consumer usually cannot detect (through smell, touch, or sight) that a drug product is safe or if it will work. While CGMP requires testing, testing alone is not adequate to ensure quality. In most instances testing is done on a small sample of a batch (for example, a drug manufacturer may test 100 tablets from a batch that contains 2 million tablets), so that most of the batch can be used for patients rather than destroyed by testing. Therefore, it is important that drugs are manufactured under conditions and practices required by the CGMP regulations to assure that quality is built into the design and manufacturing process at every step. Facilities that are in good condition, equipment that is properly maintained and calibrated, employees who are qualified and fully trained, and processes that are reliable and reproducible, are a few examples of how CGMP requirements help to assure the safety and efficacy of drug products.

How does FDA determine if a company is complying with CGMP regulations?

FDA inspects pharmaceutical manufacturing facilities worldwide, including facilities that manufacture active ingredients and the finished product. Inspections follow a standard approach and are conducted by highly trained FDA staff. FDA also relies upon reports of potentially defective drug products from the public and the industry. FDA will often use these reports to identify sites for which an inspection or investigation is needed. Most companies that are inspected are found to be fully compliant with the CGMP regulations.

If a manufacturer is not following CGMP, are drug products safe for use?

If a company is not complying with CGMP regulations, any drug it makes is considered “adulterated” under the law. This kind of adulteration means that the drug was not manufactured under conditions that comply with CGMP. It does not mean that there is necessarily something wrong with the drug.

For consumers currently taking medicines from a company that was not following CGMP, FDA usually advises these consumers not to interrupt their drug therapy, which could have serious implications for their health. Consumers should seek advice from their health care professionals before stopping or changing medications. Regulatory actions against companies with poor CGMP are often intended to prevent the possibility of unsafe and/or ineffective drugs. In rare cases, FDA regulatory action is intended to stop the distribution or manufacturing of violative product. The impact of CGMP violations depends on the nature of those violations and on the specific drugs involved. A drug manufactured in violation of CGMP may still meet its labeled specifications, and the risk that the drug is unsafe or ineffective could be minimal. Thus, FDA’s advice will be specific to the circumstances, and health care professionals are best able to balance risks and benefits and make the right decision for their patients.

What can FDA do to protect the public when there are CGMP violations?

If the failure to meet CGMP results in the distribution of a drug that does not offer the benefit as labeled because, for example, it has too little active ingredient, the company may subsequently recall that product. This protects the public from further harm by removing these drugs from the market. While FDA cannot force a company to recall a drug, companies usually will recall voluntarily or at FDA’s request. If a company refuses to recall a drug, FDA can warn the public and can seize the drug.

FDA can also bring a seizure or injunction case in court to address CGMP violations even where there is no direct evidence of a defect affecting the drug’s performance. When FDA brings a seizure case, the agency asks the court for an order that allows federal officials to take possession of “adulterated” drugs. When FDA brings an injunction case, FDA asks the court to order a company to stop violating CGMP. Both seizure and injunction cases often lead to court orders that require companies to take many steps to correct CGMP violations, which may include repairing facilities and equipment, improving sanitation and cleanliness, performing additional testing to verify quality, and improving employee training. FDA can also bring criminal cases because of CGMP violations, seeking fines and jail time.

How would a new drug company learn about CGMP and about FDA's expectations on complying with them?

FDA publishes regulations and guidance documents for industry in the *Federal Register*. This is how the federal government notifies the public of what we are doing and why. FDA's website, www.fda.gov also contains links to the CGMP regulations, guidance documents, and various resources to help drug companies comply with the law. FDA also conducts extensive public outreach through presentations at national and international meetings and conferences, to discuss and explain the CGMP requirements and the latest policy documents.

References

1. FDA CGMP. Accessed January 21, 2026.
2. FDA CGMP Regulations. Accessed January 21, 2026.

This page titled 2.12: Current Good Manufacturing Practice (CGMP) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

2.13: CDC Advisory Committee on Immunization Practices (ACIP)

Note

Information extracted verbatim from References 1-3.

Key points

- The Advisory Committee on Immunization Practices (ACIP) is a federal advisory committee that develops recommendations on the use of vaccines in the civilian population of the United States.
- CDC sets the U.S. adult and childhood immunization schedules based on recommendations from ACIP.
- Federal laws, regulations, and other applicable HHS policies and procedures are and will be consistently applied to all CDC FACA committees.

Charter

ACIP was established under section 222 of the Public Health Service Act (42 U.S.C. §2l7a) and is governed by its charter.

Policies and procedures

The ACIP policies and procedures document serves as a reference for ACIP members and interested members of the public about ACIP's structure and functions.

The Advisory Committee on Immunization Practices (ACIP) is a federal advisory committee, composed of medical and public health experts, that provides advice and guidance to the Director of the Centers for Disease Control and Prevention (CDC) on the most effective means to prevent vaccine-preventable diseases in the United States.

ACIP's guidance includes the use of vaccines, and may also include recommendations for administration of immune globulin preparations and/or antimicrobial therapy shown to be effective in controlling a disease for which a vaccine is available. In addition, ACIP has a statutory authority for the Vaccines for Children (VFC) program; ACIP has sole responsibility and authority to determine the vaccines, number of doses, schedule and contraindications for the VFC program. ACIP develops written recommendations—subject to the approval of the CDC Director—for the routine administration of vaccines to both pediatric and adult populations. To inform its advice to the CDC Director, ACIP considers disease epidemiology and burden of disease, vaccine efficacy and effectiveness, vaccine safety, the quality of evidence reviewed, economic analyses, and implementation issues. The overall goals of ACIP are to provide advice that will assist CDC and HHS in reducing the incidence of vaccine preventable diseases and to increase the safe usage of vaccines and related biological products, including active and passive immunoprophylaxis. The target populations for ACIP recommendations are public and private health care providers who administer vaccines, public and private officials who make vaccine policy, and the general public.

We will not cover these policies and procedures in any detail, but consult Reference 3 if you are interested.[3]

ACIP Charter details [2]

Authority

The ACIP was established under Section 222 of the Public Health Service Act (42 U.S.C. §2l7a), as amended. The Committee is governed by the provisions of the Federal Advisory Committee Act, Public Law 92-463 (5 U.S.C. § 1001 et seq.), as amended. The ACIP has been given statutory roles under subsections 1928(c)(2)(B)(i) and 1928(e) of the Social Security Act (42 U.S.C. §

1396s(c)(2)(B)(i) and 1396s(e)) and subsection 2713(a)(2) of the Public Health Service Act (42 U.S.C. § 300gg-13(a)(2)). Objective and Scope of Activities. The Secretary, Department of Health and Human Services (HHS), and by delegation the Director, Centers for Disease Control and Prevention (CDC), are authorized under Section 311 and Section 317 of the Public Health Service Act, [42 U.S.C. §243 and 42 U.S.C. §247b], as amended, to assist states and their political subdivisions in the prevention and control of communicable diseases; to advise the states on matters relating to the preservation and improvement of the public's health; and to make grants to states and, in consultation with the state health authorities, to agencies and political subdivisions of states to assist in meeting the costs of communicable disease control programs.

The ACIP shall provide advice and guidance to the Director of the CDC regarding use of vaccines and related agents for effective control of vaccine-preventable diseases in the civilian population of the United States.

Recommendations made by the ACIP are reviewed by the CDC Director, and if adopted, are published as official CDC/HHS recommendations in the Morbidity and Mortality Weekly Report (MMWR).

The CDC Director informs the Secretary, HHS, and the Assistant Secretary for Health, of immunization recommendations. Upon the licensure of any vaccine or any new indication for a vaccine, the Committee shall, as appropriate, consider the use of the vaccine at its next regularly scheduled meeting. If the Committee does not make a recommendation at the Committee's first regularly scheduled meeting, the Committee shall provide an update on the status of such for the Committee's review.

Description of Duties

The Committee shall provide advice for the control of diseases for which a vaccine is licensed in the U.S. The guidance will address use of vaccines and may include recommendations for administration of immune globulin preparations and/or antimicrobial therapy shown to be effective in controlling a disease for which a vaccine is available.

Guidance for use of unlicensed vaccines may be developed if circumstances warrant.

For each vaccine, the Committee advises on population groups and/or circumstances in which a vaccine or related agent is recommended. The Committee also provides recommendations on contraindications and precautions for use of the vaccine and related agents and provides information on recognized adverse events. The Committee also may provide recommendations that address the general use of vaccines and immune globulin preparations as a class of biologic agents, use of specific antibody products for prevention of infectious diseases, and special situations or populations that may warrant modification of the routine recommendations.

Committee deliberations on use of vaccines to control disease in the U.S. shall include consideration of disease epidemiology and burden of disease, vaccine safety, vaccine efficacy and effectiveness, the quality of evidence reviewed, economic analyses, and implementation issues.

The Committee may revise or withdraw their recommendation(s) regarding a particular vaccine as new information on disease epidemiology, vaccine effectiveness or safety, economic considerations, or other data become available.

In accordance with Section 1928 of the Social Security Act, the ACIP also shall establish and periodically review and, as appropriate, revise the list of vaccines for administration to children and adolescents eligible to receive vaccines through the Vaccines for Children Program, along with schedules regarding the appropriate dose and dosing interval, and contraindications to administration of the pediatric vaccines.

The Secretary, and as delegated the CDC Director, shall use the list established by the ACIP for the purpose of the purchase, delivery, and administration of pediatric vaccines in the Vaccines for Children Program. Further, under provisions of the Affordable Care Act (Section 2713 of the Public Health Service Act, as amended), immunization recommendations of the Committee that have been adopted by the Director of the Centers for Disease Control and Prevention must be covered by applicable health plans.

References

[1] ACIP. Advisory Committee on Immunization Practices (ACIP). Information downloaded January 7, 2026.

[2] Centers for Disease Control. ACIP Charter.

[3] Centers for Disease Control. ACIP Policies and Procedures.

This page titled 2.13: CDC Advisory Committee on Immunization Practices (ACIP) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

2.14: Project Management Considerations

Introduction

In this national and global emergency, time was of the essence. As the pandemic spread, there was intense pressure to have therapeutics to help those infected by the virus, to develop vaccines to deter the spread of the virus, to supplement the medical infrastructure that needed more ventilators/respirators, and to obtain large supplies of Personal Protective Equipment (PPE) to protect first responders, medical staff, and the general populace. PPE included:

- N95 masks and respirators
- Disposable gowns
- Eye protection (goggles, face shields)
- Gloves.

In this case study, we are primarily focusing on vaccines, but Operation Warp Speed needed to address a larger scope.

In times of emergency (national health emergencies or stresses in a competitive marketplace, it is not unusual for executives to press for seemingly impossible project completion dates. The project team may be asked to cut corners, “bend the laws of physics”, etc. in times of war and other emergencies (e.g., “Manhattan Project”). Project managers that are accustomed to following a standard process for decades may have trouble “tailoring the process” to accommodate the emergencies.

We will be looking at the following in subsections of this section:

- Risk management,
- Project activity sequencing, and
- Project acceleration techniques.

Relevance of the Project Management Triangle

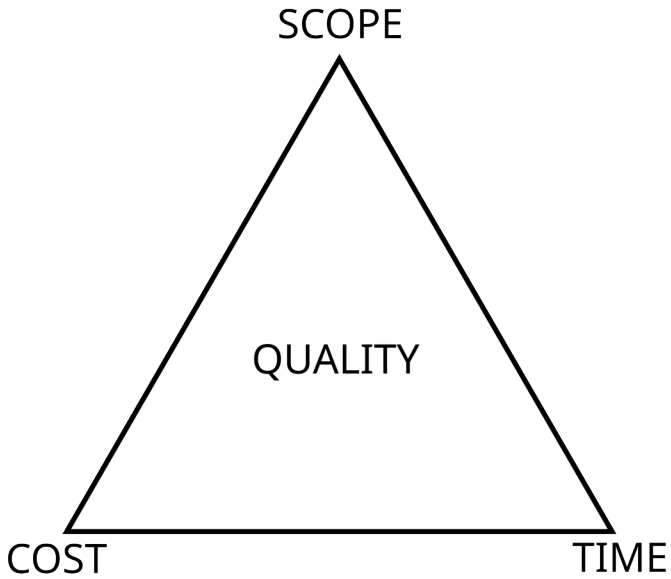


Figure 2.14.1: Project Management Triangle. (Public domain[8])

The **project management triangle** is a model of the constraints experienced when conducting project management. Its premises are:

1. The quality (depicted in the center of the triangle) of work is constrained by the project's cost (budget), time (deadlines), and

scope (features) that are shown at the vertices of the triangle.
2. Scope, time, and cost are interdependent and the project manager can trade between the constraints.
3. Changes in one constraint necessitate changes in others to compensate, otherwise quality will suffer.[5]

The time constraint refers to the amount of time available to complete a project. The cost constraint refers to the budgeted amount of money available for the project. The scope constraint refers to what must be done to produce the project's end result(s) or deliverables. These three constraints are often competing constraints. For example, an increase in project scope typically means an elongated schedule (increase in time) and and increase in cost (project budget). A tight time constraint could mean increased costs to try to accomplish all the work within that amount of time) and potentially a reduction in project scope to more comfortably complete the work within the time constraint. A tight budget could result in a reduction of project scope or increased time.[5]

In the case of Operation Warp Speed time was likely deemed the paramount constraint due to the loss of life caused by the spread of the virus. There was clearly flexibility on the cost side as the government had virtually unlimited ability to fund what was necessary and hedged its bets by investing in several vaccines concurrently. There was flexibility in scope in terms of the order in which vaccines would be provided to segments of the population, as discussed below in the subsection on phased rollout.

In the case of a pharmaceutical company developing and manufacturing a vaccine, time was similarly the major constraint, so the earlier that therapeutics and vaccines could be made available, the better. There was flexibility in cost due to the US government stepping forward and executing contracts, as we will see in detail in the section on Legal Considerations. There was some flexibility in scope that would allow addressing subsets of the overall population in initial human clinical trials, and cover the remainder subsequently (for example, younger people).

OPERATION WARP SPEED

Let's take a quick look at how tiered project management might have been handled.

- **Portfolio Management**
  - Which COVID-19 vaccine candidates to fund to hedge bets with different technologies with different risk profiles. BARDA managed the contracts with the pharmaceutical companies.
- **Program Management**
  - Vaccines (oversight of)
  - Ventilators (accelerate production by existing vendors, pursue other options).
  - Personal Protective Equipment (PPE): ramp up production of N95 masks, protective clothing, gloves.
- **Project Management**
  - Oversight of individual projects
- **Operations**
  - Vaccine volume manufacturing scale-up (oversight of)
  - Vaccine distribution logistics, including refrigeration equipment, air and ground transportation.
  - Management of cold chain logistics

Moderna Contract overview

The Moderna contract will be examined in more detail in the "Legal Considerations" section.

A review of the US government's contract with one mRNA vaccine provider provides insight onto some of the project management techniques that might have been employed.

> "The Department of Health and Human Services (HHS), Office of the Assistant Secretary for Preparedness and Response (ASPR), Biomedical Advanced Research and Development Authority (BARDA) requires the contractor(s) to develop a mRNA vaccine to licensure for the prevention of COVID-19. The project will entail pre-clinical and Phase 2 and Phase 3 clinical studies sufficient to demonstrate the safety and efficacy of the proposed vaccine(s); CMC development, scale-up, scale-out and validation of manufacturing capacities, including bulk drug substance and fill and finished drug product, with a capacity of 100 million doses by 2021 and all program management and regulatory activities necessary to achieve FDA licensure of the vaccine. The project shall be accomplished on an **accelerated timeline**, with **parallel activity WBS**, **aggressive manufacturing scale-up**, **risk management**, and **taking advantage of any regulatory flexibilities**. Contract terms include a requirement for domestic production of vaccine and assurance of material sourcing for vaccine production

during execution of the project."[4, page 4]

Note

The statement "The project will entail pre-clinical and Phase 2 and Phase 3 clinical studies..." does not require Phase 1 clinical studies. Phase 1 was already underway at the time the contract was executed.

Note

The phrase "**parallel activity WBS"** likely means that project management "fast-tracking" techniques were used. There will be more on this later.

In this section of Project Management Considerations (and subsequent subsections) we will focus on some of the content of the statement: "The project shall be accomplished on an accelerated timeline, with parallel activity WBS, aggressive manufacturing scale-up, risk management, and taking advantage of any regulatory flexibilities."[4, page 4]

The full WBS and project schedule are redacted from the publicly available contract, so we can only infer the character of what was done rather than what was actually done.

Use of Earned Value Management (EVM)

The contract also calls for the use of Earned Value Management "Lite" as stated in Section B.4.3 of the contract and elaborated on in Attachment 11 of the contract. [4]

> B.4.3 Earned Value Management (EVM) Lite Requirements
>
> The Contractor shall use an Earned Value Management (EVM) System for all retrofit and development activities of the anticipated requirement, that is consistent with the "7 Principles of Earned Value Management Tier 2 System Implantation [sic, should say Implementation] Intent Guide" attached to this contract. Alternative systems may be submitted to the Contracting Officer for consideration and approval.[4]

The "7 Principles of Earned Value Management Tier 2 System Implementation Intent Guide"[6] states the seven EVM principles (with supplemental information integrated) as follows:

- Principle 1: Plan all Work Scope
  - This includes a Statement of Work (SOW), a Work Breakdown Structure (WBS), and a WBS dictionary.
- Principle 2: Break Work into Finite Pieces and Define Person/Organization Responsible for Work
  - For Tier 2 the control account and activities/work packages are at at least at level 4 and potentially level 5. It requires running critical path analysis.
  - The schedule needs to include at least the following:
    - WBS number
    - Control Account number
    - Work package number
    - Task name
    - Duration
    - Baseline Start and Finish Dates
    - Actual Start and Finish Dates
    - Forecast Start and Finish Dates
    - Predecessors/Successors
    - Activity Percent Complete
- Principle 3a: Integrate Scope. Schedule and Budget into a Performance Measurement Baseline
  - This can be done by resource loading in a project management tool such as Microsoft Project (resource names, loaded rates for the resources).
  - Reporting needs to include Earned Value (EV) or Budgeted Cost of Work Performed (BCWP) and Actual Costs of Work Performed (ACWP).

- Principle 3b: Control Changes to the Baseline
- Principle 4: Use Actual Costs Incurred and Recorded in Accomplishing the Work Performed
- Principle 5: Objectively Assess Accomplishments at the Work Performance Level
- Principle 6a: Analyze Significant Variances From the Plan
  - This includes calculating:
    - Cost Variance (CV): Budgeted Cost of Work Performed (BCWP) minus Actual Cost of Work Performed (ACWP)
    - Schedule Variance (SV): Budgeted Cost of Work Performed (BCWP) minus Budgeted Cost of Work Schedule (BCWS)
    - Cost Performance Index (CPI): BCWP divided by ACWP
    - Schedule Performance Index (SPI): BCWP divided by BCWS
    - To-Complete Performance Index (TCPI) that calculates the cost performance at which the control account activities need to perform in order to complete the work within the current reported Estimate at Completion (EAC). EAC is ACWP plus the Estimate to Complete (ETC). ETC is an estimate of the cost of the work that still needs to be performed to complete a task.
  - Reporting should include the current Budget at Completion (BAC) and the current Estimate at Completion (EAC).
- Principle 6b: Prepare an Estimate at Completion Based on Performance to Date and Work to be Performed
- Principle 7: Use EVMS Information in the Company's Management Processes.

In BARDA's tiered approach to EVM, Tier 2 is defined as:[6]

> For countermeasure research and development contracts that have a [sic] total acquisition costs greater than or equal to $25 million and have a Technical Readiness Level (TRL) of less than 7 will apply EVM principle for tracking cost, schedule and technical performance that comply with the 7 Principles of EVM Implementation.[6]

Note

BARDA defines Technical Readiness levels.[7]

Medical Countermeasures | ABOUT THE TECHNOLOGY READINESS LEVELS (TRLs)

Risk Management

There was significant scientific risk in the use of mRNA vaccines against COVID-19.

> “...despite extensive previous research attempts to produce safe, effective vaccines against coronaviruses, such as SARS and MERS, all vaccine candidates for coronavirus infections have failed during clinical research, and no vaccine exists to prevent any coronavirus infection.”[1]

Another large risk was in overlapping volume manufacturing ramp up with clinical trials. This will be covered in a subsequent section.

Project Acceleration Techniques

We will examine in a subsequent section some potential schedule improvement techniques:

- Crashing
- Fast tracking
- Combining Clinical Trials Phase 1 and 2
- Overlapping Clinical Trials Phase 2 and 3
- Overlapping Clinical Trial Phase 3 and Manufacturing volume production ramp
- Adaptive clinical trial design: if the clinical trial testing data is excellent, adjustments in schedule are possible. For example, trial durations can be reduced. Moderna's Phase 3 was eight weeks. Additionally, one might use fewer participants.

Phased-Rollout (CDC Playbook)[3]

Since it was obvious that over 330 million doses of COVID-19 vaccines were not going to be available initially, advance planning

guidance from the Centers for Disease Control (CDC) called for a three phase approach for allocating vaccine doses as they became available from manufacturers.

A key point to consider is that vaccine supply will be limited at the beginning of the program, so the allocation of doses must focus on vaccination providers and settings for vaccination of limited critical populations as well as outreach to these populations. The vaccine supply is projected to increase quickly over the proceeding months, allowing vaccination efforts to be expanded to additional critical populations and the general public.[3]

Populations of focus for initial COVID-19 vaccination may include:

- Healthcare personnel likely to be exposed to or treat people with COVID-19.
- People at increased risk for severe illness from COVID-19, including those with underlying medical conditions and people 65 years of age and older
- Other essential workers

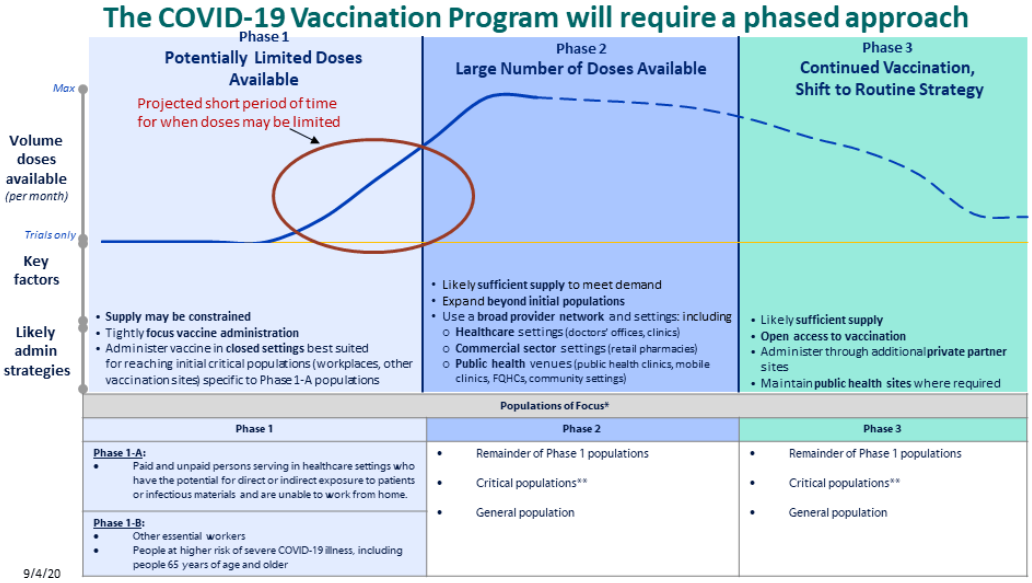


Figure 2.14.1: CDC Vaccination Playbook. (Centers for Disease Control and Prevention (CDC), COVID-19 Vaccination Program Interim Playbook for Jurisdiction Operations, September 16, 2020, p11)

The CDC advises that jurisdictions should be planning in terms of three phases:[3]

1. Phase 1: Potentially limited supply of COVID-19 vaccine doses available
    - Focus initial efforts on reaching the critical populations listed above. Ensure vaccination locations selected can reach populations, manage cold chain requirements, and meet reporting requirements for vaccine supply and uptake.
2. Phase 2: Large number of vaccine doses available
    - Focus on ensuring access to vaccine for members of Phase 1 critical populations who were not yet vaccinated as well as for the general population; expand provider network.
3. Phase 3: Sufficient supply of vaccine doses for entire population (surplus of doses)
    - Focus on ensuring equitable vaccination access across the entire population. Monitor vaccine uptake and coverage; reassess strategy to increase uptake in populations or communities with low coverage.

Critical Populations

The CDC identified critical populations as follows (not in any particular order):[3. page 14]

- Critical infrastructure workforce
    - Healthcare personnel (i.e., paid and unpaid personnel working in healthcare settings, which may include vaccinators, pharmacy staff, ancillary staff, school nurses, and EMS personnel)
    - Other essential workers (see additional guidance from the Cybersecurity and Infrastructure Security Agency [CISA])
    - Note: The critical infrastructure workforce varies by jurisdiction. Each jurisdiction must decide which groups to focus on

when vaccine supply is limited by determining key sectors that may be within their populations (e.g., port-related workers in coastal jurisdictions)
- People at increased risk for severe COVID-19 illness
  - LTCF residents (i.e., nursing home, assisted living, independent living facility residents)
  - People with underlying medical conditions that are risk factors for severe COVID-19 illness
  - People 65 years of age and older
- People at increased risk of acquiring or transmitting COVID-19
  - People from racial and ethnic minority groups
  - People from tribal communities
  - People who are incarcerated/detained in correctional facilities
  - People experiencing homelessness/living in shelters
  - People attending colleges/universities
  - People who work in educational settings (e.g., early learning centers, schools, and colleges/universities)
  - People living and working in other congregate settings
- •People with limited access to routine vaccination services
  - People living in rural communities
  - People with disabilities
  - People who are under- or uninsured

Vaccine Allocation

The federal government will determine the amount of COVID-19 vaccine designated for each jurisdiction. The jurisdiction's immunization program will then be responsible for managing and approving orders from enrolled providers within their jurisdiction using this allotment.[3, page 24]

Ordering

COVID-19 vaccination providers enrolled by the jurisdiction will order COVID-19 vaccine through their jurisdiction's immunization program. Most jurisdictions will ask COVID-19 vaccination providers to place orders using systems and procedures routinely used for ordering publicly funded vaccines (e.g., IIS/ExIS upload to CDC's VTrckS for provider direct order entry), though some jurisdictions may have augmented systems.[3, page 24]

Ancillary supplies will be packaged in kits and will be automatically ordered in amounts to match vaccine orders in VTrckS. Each kit will contain supplies to administer 100 doses of vaccine, including:

- Needles, 105 per kit (various sizes for the population served by the ordering vaccination provider)
- Syringes, 105 per kit
- Alcohol prep pads, 210 per kit
- 4 surgical masks and 2 face shields for vaccinators, per kit
- COVID-19 vaccination record cards for vaccine recipients, 100 per kit

Vaccine Distribution[3, pages 25 and 26]

COVID-19 vaccines and ancillary supplies will be procured and distributed by the federal government at no cost to enrolled COVID-19 vaccination providers. CDC will use its centralized distribution contract to fulfill orders for most vaccine products and associated ancillary supplies. Some vaccine products, such as those with ultra-cold temperature requirements, will be shipped directly from the manufacturer to the vaccination provider site.

Jurisdictions should ensure accurate and complete shipping information (e.g., shipment address, provider contact information, shipping hours) is available in the VTrckS system for all vaccine shipments to enrolled vaccination providers.

COVID-19 vaccine (and diluent or adjuvant, if required) will be shipped to vaccination provider sites enrolled by the jurisdiction's immunization program within 48 hours of order approval. Because of cold chain requirements, ancillary supply kits (and diluent, if applicable) will ship separately from vaccine but should arrive before or on the same day as vaccine.

The federally contracted vaccine distributor uses validated shipping procedures to maintain COVID-19 vaccine cold chain and minimize the likelihood of vaccine loss or damage during shipment. Once a vaccine product has been shipped to a COVID-19

vaccination provider site, the federal government will neither redistribute the product nor take financial responsibility for its redistribution.

Whenever possible, vaccine should be shipped to the location where it will be administered to minimize potential breaks in the cold chain. However, there may be circumstances where COVID-19 vaccine needs to be redistributed beyond the identified primary CDC ship-to sites (i.e., for orders smaller than the minimum order size or for large organizations whose vaccine is shipped to a central depot and requires redistribution to additional clinic locations). In these instances, vaccination provider organizations/facilities, third-party vendors, and other vaccination providers may be allowed, if approved by the jurisdiction's immunization program, to redistribute COVID-19 vaccine, if validated cold-chain procedures are in place in accordance with the manufacturer's instructions and CDC's guidance on COVID-19 vaccine storage and handling. These entities must sign and agree to conditions in the CDC COVID-19 Vaccine Redistribution Agreement for the sending facility/organization and have a fully completed and signed CDC COVID-19 Vaccination Provider Profile form for each receiving location. Jurisdictions should be extremely judicious in allowing redistribution and limit any redistribution to refrigerated vaccines only.

Vaccine Storage and Handling[3, page 27]

COVID-19 vaccine products are temperature-sensitive and must be stored and handled correctly to ensure efficacy and maximize shelf life. Proper storage and handling practices are critical to minimize vaccine loss and limit risk of administering COVID-19 vaccine with reduced effectiveness. Jurisdictions should work with staff at each COVID-19 vaccination provider site to ensure appropriate vaccine storage and handling procedures are established and followed.

It is expected that cold chain storage and handling requirements for COVID-19 vaccine products will vary in temperature from refrigerated (2°C to 8°C) to frozen (-15 to -25°C) to ultra-cold (-60°C to -80°C in the freezer or within the dry ice shipping container in which product was received). Ongoing stability testing may impact these requirements. Note: These temperatures are based on information available as of 9/04/2020. Updated information will be provided as it becomes available.

For a reliable cold chain, three elements must be in place:

- Well-trained staff
- Reliable storage and temperature monitoring equipment
- Accurate vaccine inventory management

The cold chain begins at the COVID-19 vaccine manufacturing plant, includes delivery to and storage at the COVID-19 vaccination provider site, and ends with administration of COVID-19 vaccine to a person. Jurisdictions and vaccination providers are responsible for maintaining vaccine quality from the time a shipment arrives at a vaccination provider site until the dose is administered. To minimize opportunities for breaks in the cold chain, most COVID-19 vaccine will be delivered from CDC's centralized distributor directly to the location where the vaccine will be stored and administered, although some vaccine may be delivered to secondary depots for redistribution. Certain COVID-19 vaccine products, such as those with ultra-cold temperature requirements, will be shipped directly from the manufacturer to the vaccination provider site. If redistributing vaccine, jurisdictions must adhere to all cold chain requirements and should limit transport of frozen or ultra-cold vaccine products.

Emergency Use Authorization[3, page 39]

Initially available COVID-19 vaccines may be authorized for use under an EUA issued by FDA or approved as licensed vaccines.

> **Note**
>
> Please remember these items as project deliverables when preparing your WBS and project schedule.

Emergency Use Authorization Fact Sheets

The EUA authority allows FDA to authorize either (a) the use of an unapproved medical product (e.g., drug, vaccine, or diagnostic device) or (b) the unapproved use of an approved medical product during an emergency based on certain criteria. The EUA will outline how the COVID-19 vaccine should be used and any conditions that must be met to use the vaccine. FDA will coordinate with CDC to confirm these "conditions of authorization." Vaccine conditions of authorization are expected to include distribution

requirements, reporting requirements, and safety and monitoring requirements. The EUA will be authorized for a specific time period to meet response needs (i.e., for the duration of the COVID-19 pandemic).

- Product-specific **EUA fact sheet for COVID-19 vaccination providers** will be made available that will include information on the specific vaccine product and instructions for its use.
- An **EUA fact sheet for vaccine recipients** will also be developed, and both will likely be made available on the FDA website and through the CDC website. Jurisdictions should ensure providers know where to find both the provider and recipient fact sheets, have read and understand them, and are clear on the requirement to provide the recipient fact sheet to each client/patient prior to administering vaccine.

Caution

Students sometimes confuse the use of **Clinical Trials phase names**: Phase 1, Phase 2, Phase 3, and Phase 4 with the project life cycle that employs phase gate reviews with a project steering committee after the following **project phase gate names**: Phase 0, Phase 1, Phase 2, Phase 3, and Phase 4.

References

1. Wikipedia - Operation Warp Speed. August 24, 2020.
2. Jenni Fink, Newsweek, "Fauci calls naming of Trump administration's Operation Warp Speed unfortunate, says it implies reckless speed." August 14, 2020.
3. Centers for Disease Control and Prevention (CDC), COVID-19 Vaccination Program Interim Playbook for Jurisdiction Operations, September 16, 2020. Page 11 for the allocation of vaccines in phases.
4. HHS-BARDA-Moderna-Contract-75A50120C00034-16April2020-Version8Feb2021. HHS/ASPR/BARDA-Moderna Contract 75A50120C00034. Moderna Contract. April 16, 2020. (Less redacted version obtained via FOIA lawsuit on February 8, 2021).
5. Wikipedia - Project Management Triangle. Accessed January 23, 2026.
6. HHS-BARDA-Moderna-Contract-75A50120C00034-16April2020-Version8Feb2021. HHS/ASPR/BARDA-Moderna Contract 75A50120C00034. Moderna Contract. April 16, 2020. (Less redacted version obtained via FOIA lawsuit on February 8, 2021). Specifically, Attachment#11: Department of Health & Human Services HHS, Office of the Assistant Secretary for Preparedness and Readiness ASPR, Biomedical Advanced Research and Development Authority BARDA. "7 Principles of Earned Value Management Tier 2 System Implementation Intent Guide", 21 December 2011.
7. Technology Readiness Levels. Accessed January 31, 2026.
8. Wikimedia Illustration of Project Management Triangle.

This page titled 2.14: Project Management Considerations is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

2.14.1: Project Schedule Preparation

Project schedule preparation involves the following:

- Activity identification
- Activity sequencing
- Activity duration estimation
- Activity resource assignments.

Activity identification

The starting point is the project Work Breakdown Structure (WBS). The objective is to identify what specific actions the project needs to take to take to produce the project's deliverables. The project's deliverables are translated into work packages and then unto specific activities. Since the activities describe work that the project needs to perform, the format of an activity that is placed in the project schedule is often "Verb...object."

The list of required activities is well known in pharmaceutical development.

In the case of the Moderna mRNA vaccine, we can get partial visibility into a WBS by inspecting the contract (that has many redactions) and assembling information from several sections and other sources.

WBS, recreated from Ref 1, Section C.2 page 9ff and page 34.

WBS item	Item Description	Deliverable
1.0	mRNA Vaccine Development	
1.1	Program Management	
1.1.1	mRNA Program Management; Program and Alliance Management	• Management Plans; Routine Reporting Deliverables
1.2	Nonclinical Toxicology	
1.2.2	Safety	
1.2.2.1	Development and Reproductive Toxicology of mRNA (to assess the risk of vaccinating pregnant women, do a complete GLP rat development and reproductive toxicology (DART) study).	• Final Study Report
1.3	Nonclinical (continue to support non-clinical activities to develop murine (relating to rodents) and non-human primate efficacy studies and animal models to assess the potential of vaccine-enhanced disease)	
1.3.1	Model Development (reserved)	
1.3.1.2	NHP Efficacy Study	• Final Study Report
1.3.1.3	Mouse Efficacy Study	• Final Study Report
1.3.3.1	Assess Disease Enhancement (assess the theoretical risk of vaccine-induced disease enhancement triggered by COVID infection after vaccination with the mRNA vaccine)	

WBS item	Item Description	Deliverable
1.3.3.2	Establish a Surrogate of Protection	
1.4	Clinical	
1.4.2	Phase 2	
1.4.2.1	Phase 2 Safety and Immunogenicity Study	• Clinical Study Protocol • Final Clinical Study Report
1.4.3	Phase 3	
1.4.3.1	Phase 3 Efficacy or Safety and Imrnunogenicity or Phase 3 Pivotal Study (Scenario I in which COVID-19 virus is circulating and Scenario II in the absence of the virus)	• Clinical Study Protocol • Final Clinical Study Report
1.4.3.2	Phase 3 Lot-to-Lot (all information was redacted)	• Clinical Study Protocol • Final Clinical Study Report
1.4.3.3	Phase 3 Adolescents or Pediatrics (all information was redacted)	• Clinical Study Protocol • Final Clinical Study Report
1.5	Regulatory	
1.5.1	IND	
1.5.1.1	IND Filing	• To support Phase 2 onward
1.5.1.2	IND Maintenance	• Record of FDA Communications
1.5.2	BLA	
1.5.1.1	BLA Submission	• NA
1.6 CMC		
1.6.3	Pilot Scale Manufacturing	

WBS item	Item Description	Deliverable
1.6.3.2	CTM Manufacture for P201 (for Phase 2)	• CoA for Clinical Lots
1.6.3.3	mRNA Process Development (optimize process parameters for mRNA manufacturing) (All information was redacted.)	
1.6.3.4	CTM Manufacture for P301	• CoA for Clinical Lots
1.6.3.6	CTM Manufacture for P302/P303	• CoA for Clinical Lots
1.6.3.8	BLA Readiness	
1.6.4.1	Process Development for Full Commercial Scale	
1.6.5.2	Analytical Method Development and Validation	
1.6.5.3	Characterization Assay Development and Implementation (All information was redacted.)	
1.6.5.4	Stability Studies (to help determine product shelf life)	

Note

It should be noted that this is not a complete WBS. For example, it does not include items such as a document specifying cold-chain instructions; the instructions for administering the vaccine and list of side effects that ships with the vaccine vial; and other items.

Additionally, Clinical Trial Phase 1 was not included in the BARDA contract with Moderna because it was already in progress under another contract.

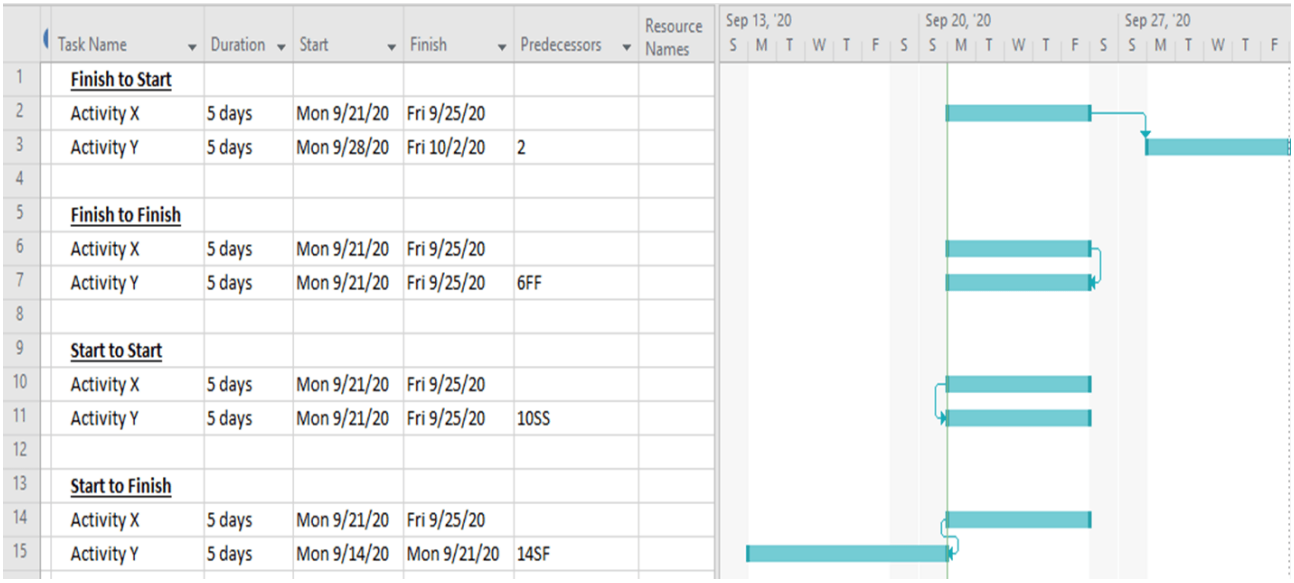
Activity sequencing

Activity sequencing involves arranging predecessor and successor activities to reflect their order or sequence and to reflect task dependencies.

The four activity sequencing relationships are:

- **Finish to Start (FS).** This is used when starting a successor activity depends on finishing a predecessor activity or activities. This relationship is the most common.
- **Finish to Finish (FF).** This is used when completion of a successor depends on the completion of a predecessor activity or activities.
- **Start to Start (SS).** This is used when starting a successor activity depends on starting a predecessor activity.
- **Start to Finish (SF).** This is used when the completion of an activity depends on the start of another activity.

These sequencing relationships can be easily instrumented using commonly available project management software tools such as Microsoft Project and ProjectLibre. The "Predecessors" field is used to declare for a successor which activity or activities are its predecessors, and state the nature of the relationship. Finish to Start (FS) is the default since it is the most common activity sequencing relationship.



	Task Name	Duration	Start	Finish	Predecessors	Resource Names
1	**Finish to Start**					
2	Activity X	5 days	Mon 9/21/20	Fri 9/25/20		
3	Activity Y	5 days	Mon 9/28/20	Fri 10/2/20	2	
4						
5	**Finish to Finish**					
6	Activity X	5 days	Mon 9/21/20	Fri 9/25/20		
7	Activity Y	5 days	Mon 9/21/20	Fri 9/25/20	6FF	
8						
9	**Start to Start**					
10	Activity X	5 days	Mon 9/21/20	Fri 9/25/20		
11	Activity Y	5 days	Mon 9/21/20	Fri 9/25/20	10SS	
12						
13	**Start to Finish**					
14	Activity X	5 days	Mon 9/21/20	Fri 9/25/20		
15	Activity Y	5 days	Mon 9/14/20	Mon 9/21/20	14SF	

Figure 2.14.1.1: Examples of Activity Sequencing Relationships. (CC BY NC-SA 4.0; Robert Slate)

Figure 2.14.1.1 illustrates these four relationships.

- **Finish to Start (FS).** Activity Y can start after Activity X completes. Since the FS relationship is the most common, "FS" does not need to be explicitly designated in the Predecessors column. Here "2" indicates the line on which Predecessor ActiVity X resides.
- **Finish to Finish (FF).** Completion of successor Activity Y depends on the completion of a predecessor Activity X. "6FF" in the Predecessors column states that the activity in line 6 is in the FF relationship.
- **Start to Start (SS).** Starting successor Activity Y on line 11 depends on starting predecessor Activity X on line 10 (10SS).
- **Start to Finish (SF).** The completion of Activity Y on line 15 depends on the start of Activity X on line 14 (14SF).

Leads and Lags

Leads and lags can be used for overlapping activities and delaying activities.

A lead is used to start an activity before a predecessor activity has completed. The amount of lead time is specified.

Note

Example of the use of a lead are:

1. Start coding n days before design completion.
2. Start writing documentation before coding is complete.
3. A relay runner starts running before the previous runner reaches the baton handoff zone.

This would be useful in starting volume production manufacturing before the full completion of Phase 3 clinical trials.

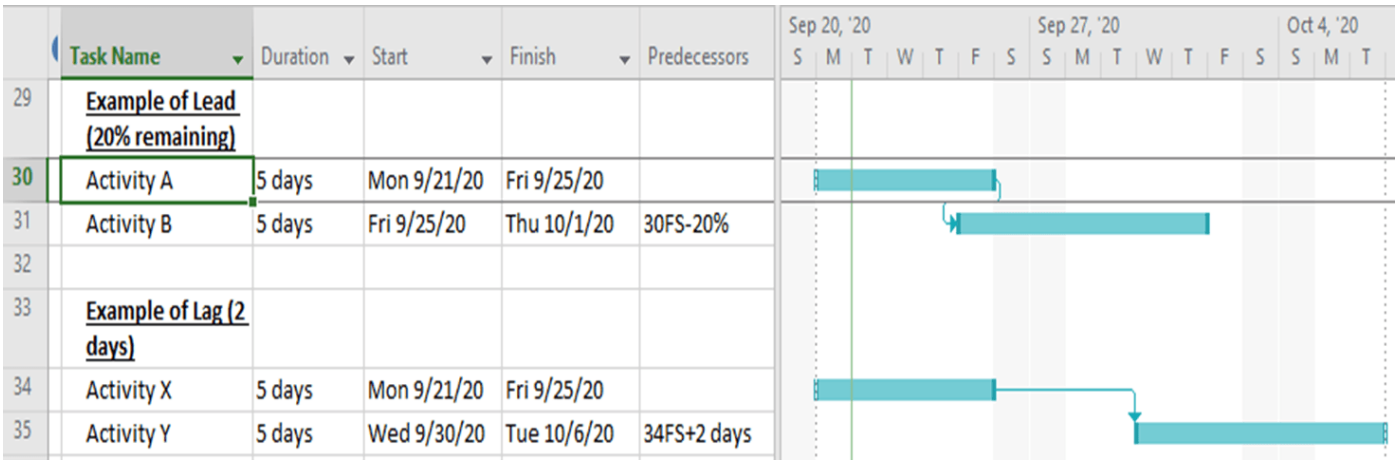
A lag is used to delay the start of a successor activity in relation to a predecessor activity. It allows a project manager to insert waiting time between activities.

Note

Examples of lags are:

1. Wait n days after pouring the concrete for a building's foundation before starting framing. This is done to ensure the concrete has hardened adequately.
2. Wait 20 minutes after receiving a vaccine before leaving the vaccine administration area. This is done to monitor for potential side-effects of the vaccine.

Leads and Lags can be instrumented using commonly available project management software tools such as Microsoft Project and ProjectLibre.



	Task Name	Duration	Start	Finish	Predecessors
29	Example of Lead (20% remaining)				
30	Activity A	5 days	Mon 9/21/20	Fri 9/25/20	
31	Activity B	5 days	Fri 9/25/20	Thu 10/1/20	30FS-20%
32					
33	Example of Lag (2 days)				
34	Activity X	5 days	Mon 9/21/20	Fri 9/25/20	
35	Activity Y	5 days	Wed 9/30/20	Tue 10/6/20	34FS+2 days

Figure 2.14.1.2: Leads and Lags. (CC BY NC-SA 4.0; Robert Slate, CIS 79 project materials)

In summary:

- **Lead:** Provides a means for overlapping activities (allows for starting Activity B before Activity A completes)
- **Lag:** Allows for introducing a delay (introduces a delay before Activity Y starts)

In Figure 2.14.1.2, an example of each is provided:

- **Lead:** Start Activity B before Activity A completes, with 20% remaining to go on Activity A. "30FS-20%" means to use the predecessor activity on line 30 (Activity A) in the FS relationship, 20% before its completion
- **Lag:** Introduces a delay of 2 days from the completion of Activity X (on line 34) before Activity Y starts (34FS+2days).

Activity duration estimation

There are several ways to estimate activity durations and there are covered in CIS 79 lectures. The estimates may be more difficult in the development of the COVID-19 vaccines because of breaking new scientific ground with mRNA vaccines.

Activity resource assignments

Figure 2.14.1.1 shows a "Resource Names" field. This is used to specify the resource or resources working on the activity.

Typically these resources are entered on a Resource Table in the project management software tool specifying the resource by name and the resource's labor rate.

For a pharmaceutical company working on a COVID-19 vaccine, the resource professionals might include:

- Toxicology specialists
- Medical doctors (dose safety and efficacy)
- Clinical pharmacologists (drug interactions, effects on organs such as liver and kidneys)
- Clinical trial volunteer management
- Statisticians
- Vaccine manufacturing personnel

- Quality control professionals
- Regulatory support
- Legal support

References

1. HHS-BARDA-Moderna-Contract-75A50120C00034-16April2020-Version8Feb2021. HHS/ASPR/BARDA-Moderna Contract 75A50120C00034. BARDA-Moderna Contract. April 16, 2020. (Less redacted version obtained via FOIA lawsuit on February 8, 2021).

This page titled 2.14.1: Project Schedule Preparation is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

2.14.2: Project Acceleration Techniques

In this section we will cover several project acceleration techniques that could be used on any project. We then look at several possibilities for a COVID-19 vaccine project.

General Project Acceleration Techniques

- **Crashing** involves adding additional resources to activities on the critical path in an attempt to improve schedule.
- **Fast tracking** is doing activities in parallel that normally would be done sequentially.

Vaccine Project Acceleration Techniques

The technique employed on a pharmaceutical company's project to develop a COVID-19 vaccine is fast-tracking, by doing activities in parallel that normally would be done sequentially, accepting the associated risks.

Recall the extract from the BARDA contract with Moderna that stated:

> "The project shall be accomplished on an accelerated timeline, with parallel activity WBS, aggressive manufacturing scale-up, risk management, and taking advantage of any regulatory flexibilities."[1]

The reference to "parallel activity WBS" is a clear reference to the use of fast-tracking. We can look for potential activities that could be performed in parallel, albeit with risk.

Some possibilities for parallel activity could conclude:

- Combining Clinical Trials Phase 1 and 2
- Overlapping Clinical Trials Phase 2 and 3. Some portions of Phase 3 clinical trials could be overlapped with Phase 2 clinical trials.
- Overlapping Clinical Trial Phase 3 and Manufacturing ramp. The volume production manufacturing ramp-up could be started in parallel with Phase 3 clinical trials (Moderna Phase 3 was 8 weeks). Without an emergency situation, it would be extremely risky to ramp up production before knowing the full results of Phase 3 clinical trials. The full results could indicate that the production formulation needs to be scrapped, with major financial impact. The government is helping to de-risk the situation.

Additionally, one could make use of "Adaptive clinical trial design." For example, if the clinical trial testing data is excellent, adjustments in schedule are possible. These adjustments might include reducing the durations of trials. It might be possible to use fewer clinical trial participants.

Let's now return to the Figure shown earlier in this chapter from the General Accountability Office's analysis that is replicated here for convenience.[2] We can see extensive use of overlapping of Clinical Trial Phases. For example, in the case of the Moderna vaccine, we can see that Clinical Trial Phases 1, 2, and 3 are overlapping. Figure $2.14.2.1$ is only addressing the specifics of clinical trails, so it is not showing any overlapping of clinical trials with manufacturing ramp up.

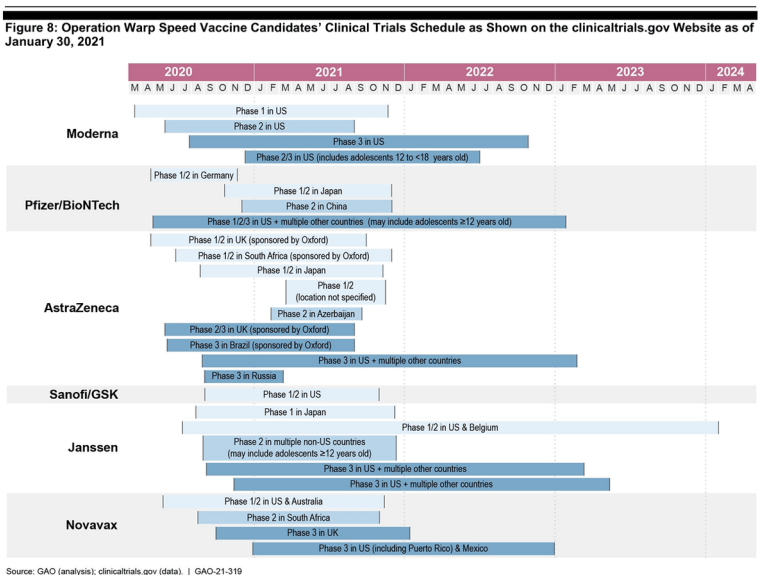


Figure 2.14.2.1: Vaccine clinical trials schedule summaries. (GAO-21-319, Accessible Version, OPERATION WARP SPEED: COVID-19 Vaccine Development and Manufacturing [2,p28])

Accessibility Note

The following is a recap of the synopsis of the Clinical trial phase overlapping as employed by various pharmaceutical companies from a project management perspective:

- Moderna: Phase 2 started and completed during Phase 1 in the Unites States. Phase 3 in the United States started while Phase 1 and Phase 2 were in progress.
- Pfizer/BioNTech: Conducted Phase 1 and Phase 2 in Germany, and started Phase 1/Phase 2 in Japan toward the end of Phase 1/2 in Germany. It initiated Phase 2 in China shortly after initiating Phase 1/Phase 2 in Japan. It also initiated Phase 1/2/3 in the US and other countries at the same time as it started the Phase 1/2 in Germany.
- AstraZeneca: After initiating Phase 1/2 in the UK with Oxford, it initiated an overlapping Phase 1/2 in South Africa, also with Oxford. While the Phase 1/2 in South Africa was in progress, it initiated a Phase 1/2 in Japan and a Phase 2 in Azerbaijan. A Phase 2/3 in the UK overlapped with the Phase 1/2 in the UK, as did a Phase 3 in Brazil, followed by the initiation of a Phase 3 in the United States and other countries, including Russia.
- Sanofi/GSK: Conducted a Phase 1/2 in the United States.
- Janssen: Conducted a Phase 1/2 in the United States and Belgium, with an overlapping Phase 1 in Japan, and a Phase 2 in multiple non-US countries,
- Novavax: It initiated a Phase 1/2 in the United States and Australia, followed by an overlapping Phase 2 in South Africa, which in turn overlapped with a Phase 3 in the UK and later by a overlapping Phase 3 in the United States.

Note

Please keep Figure 2.14.2.1 in mind when you are doing your CIS 79 vaccine schedule assignment.

Early volume manufacturing

“As part of Operation Warp Speed, doses of a vaccine will be **manufactured before it proves to be safe and effective**. The goal is to have 300 million doses ready to be administered to the American public, and the administration has entered agreements with at least three companies.”[3]

“It's a process that requires a significant financial investment but one that Fauci has praised time and time again. On August 5, he told Bloomberg's Balance of Power show that preparing sites for clinical trials and **proactively producing the vaccine can shave off months of time if a vaccine candidate works out**. And if the candidate isn't viable, "the only thing you've lost is

money," he noted."[3]

> **Note**
>
> Bolding of text was added for emphasis.

This will be discussed further on the section on risk management.

References

1. HHS-BARDA-Moderna-Contract-75A50120C00034-16April2020-Version8Feb2021. BARDA-Moderna Contract. April 16, 2020. (Less redacted version obtained via FOIA lawsuit on February 8, 2021).
2. United States Government Accountability Office. GAO-21-319 Operation Warp Speed Accelerated COVID 19 Vaccine Development Status and Efforts to Address Manufacturing Challenges. Accessible Version. GAO-21-319 OWS. Accessed February 2, 2026.
3. Jenni Fink, Newsweek, "Fauci alls naming of Trump administration's Operation Warp Speed unfortunate, says it implies reckless speed." August 14, 2020.

This page titled 2.14.2: Project Acceleration Techniques is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

2.14.3: Risk Management

We will cover the following areas of Risk Management:

- Scientific risk
- Volume manufacturing ramp risk
- Potential vaccine vial shortage risk

Scientific Risk

At the start of the project, there were considerable scientific risks associated with the development of COVID-19 vaccines:

> "...despite extensive previous research attempts to produce safe, effective vaccines against coronaviruses, such as SARS and MERS, all vaccine candidates for coronavirus infections have failed during clinical research, and no vaccine exists to prevent any coronavirus infection."[1]

Note

However, the government contract [2, Section C.1 Background, page 9] with Moderna does reference their prior success with other pathogens.

Coronaviruses are potential epidemic threats and have been recognized on the World Health Organization's list of top emerging diseases likely to cause major epidemics and Coalition for Epidemic Preparedness Innovations priority pathogens list.

In early 2020, there were no vaccines to prevent Coronavirus infection that were available.

The emergence of the SARS-CoV-2 virus created an urgent need to rapidly develop vaccines to prevent COVID-19 disease.

Developing and delivering a vaccine for highly transmissible, emerging diseases such as the SARS-CoV-2 virus requires breaking from traditional approaches.

It requires:

- parallel development activities
- aggressive manufacturing scale-up
- risk management and
- implementation of regulatory flexibilities.

Many of these requirements are met by manufacturing platforms that are capable of producing vaccines against different agents using essentially the same manufacturing systems. Suitable platforms are constituted by defined product production processes that allow significant planning for manufacturing scale and time to vaccine availability and will be supported by human safety and immunogenicity data targeting one or more infectious agents.

The BARDA contract with Moderna[2] states that is critical that the vaccine be produced in the United States. Domestic production of the vaccine is the only assurance that Americans will have access to the finished vaccine product.

Moderna's mRNA-based vaccine platform has been used in the past to rapidly prepare vaccine candidates against:

- Cytomegalovirus
- Zika
- Respiratory Syncytial Virus (RSV)
- Influenza
- Human Metapneumovirus and
- Parainfluenza virus.

Four of these candidates have been evaluated in Phase 1 clinical studies and shown to be safe and immunogenic.

Moderna collaborated with the Vaccine Research Center, NIAID ("VRC/NIAID") to design a lead SARS-CoV-2 vaccine candidate encoding a stabilized pre-fusion SARS-CoV-2 Spike protein, which is more immunogenic than wild-type or subunit proteins.

At the time of BARDA's issuance of the contract with Moderna [2], Moderna's mRNA vaccine was being evaluated in pre-clinical and Phase 1 trials sponsored by NIAID. For the purposes of this contract[2], Moderna will perform all work required to support:

- the advanced development,
- scale-up manufacturing, and
- FDA licensure of their lead SARS-CoV-2 vaccine candidate(s).

This work included:

- pre-clinical development of mRNA vaccines to demonstrate safety and efficacy against COVID-19,
- mRNA vaccine process and manufacturing scale-up development,
- product lot release assay development and process validation,
- production of clinical material and consistency lots,
- clinical evaluation studies for safety, immunogenicity and efficacy; and
- fill/finish capacity evaluation, expansion, and validation.[2]

Manufacturing Risk

No commercial enterprise would be willing to assume the risk of ramping up production manufacturing before normal approvals. This concern was mitigated by the US government stepping in and assuming financial responsibility.

This was briefly discussed in the section on Schedule acceleration techniques:

> “As part of Operation Warp Speed, doses of a vaccine will be **manufactured before it proves to be safe and effective**. The goal is to have 300 million doses ready to be administered to the American public, and the administration has entered agreements with at least three companies.”[1]
>
> “It's a process that **requires a significant financial investmen**t but one that Fauci has praised time and time again. On August 5, he told Bloomberg's Balance of Power show that preparing sites for clinical trials and **proactively producing the vaccine can shave off months of time if a vaccine candidate works out**. And if the candidate isn't viable, "the only thing you've lost is money," he noted.”[1]

The US government and the pharmaceutical companies working on vaccines needed to be working very closely with one another to make such a major decision. No project would have sufficient management reserves to handle the downside risk.

Handling potential vaccine vial shortage

As the pandemic progressed, there were concerns that there might be a shortage of vials to contain the vaccine given the large numbers of doses for people in the United States. This concern was heightened when it became apparent that the leading mRNA vaccines would require a second dose.

As a result, BARDA executed a contract with a vial supplier to mitigate that risk, discussed further in the section on "Legal Considerations."

References

1. Jenni Fink, Newsweek, "Fauci alls naming of Trump administration's Operation Warp Speed unfortunate, says it implies reckless speed." August 14, 2020.
2. HHS-BARDA-Moderna-Contract-75A50120C00034-16April2020-Version8Feb2021. Moderna Contract. April 16, 2020. (Less redacted version obtained via FOIA lawsuit on February 8, 2021).

This page titled 2.14.3: Risk Management is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

2.15: Legal Considerations

In this section on Legal Considerations we will cover the following topics:

- The legal underpinning for declaring National Emergencies (National Emergencies Act).
- The actual invocation of an emergency by President Donald Trump for the COVID-19 emergency (Proclamation 9944).
- Contracts in general (including contract types and when to use them, such as possibly using incentive clauses for accelerated delivery due to the COVID-19 emergency).
- BARDA contracts for COVID-19 (contracts between the United Stated government and vaccine developers and for procuring vials).
- Mitigating risks (ordering of hundreds of millions of vaccines in advance, support for starting volume production early).
- Handling the potential liability issues associated with the COVID-19 vaccines that were developed under a highly accelerated timetable.

This page titled 2.15: Legal Considerations is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

2.15.1: National Emergency Powers

Legal Basis for Emergency Powers

"The President of the United States has available certain powers that may be exercised in the event that the nation is threatened by crisis, exigency, or emergency circumstances (other than natural disasters, war, or near-war situations). Such powers may be stated explicitly or implied by the Constitution, assumed by the Chief Executive to be permissible constitutionally, or inferred from or specified by statute."[2,unnumbered page "Summary"]

Some emergency powers may be implicitly granted in the Constitution under Article I, Section 8 ("provide for the common Defense and general Welfare").[2, page 1]

"The National Emergencies Act (50 U.S.C. §§1601-1651) eliminated or modified some statutory grants of emergency authority, required the President to formally declare the existence of a national emergency and to specify what statutory authority activated by the declaration would be used, and provided Congress a means to countermand the President's declaration and the activated authority being sought."[2,unnumbered page "Summary"]

As a result of the severity of the COVID-19 pandemic in the spring of 2020, utilizing the authority granted by the National Emergencies Act (50 U.S.C. §§1601-1651), President Trump issued Proclamation 9944 on March 13, 2020, "Declaring a National Emergency Concerning the Novel Coronavirus Disease (COVID-19) Outbreak" (see 85 Federal Register 15337, as shown in Declared National Emergencies Under the National Emergencies Act in Effect. [2, Table 2, page 14]

Proclamation 9944

The content of the proclamations is as follows (quoted verbatim):

Administration of Donald J. Trump, 2020

Proclamation 9994—Declaring a National Emergency Concerning the Novel Coronavirus Disease (COVID–19) Outbreak

March 13, 2020

By the President of the United States of America A Proclamation

In December 2019, a novel (new) coronavirus known as SARS–CoV–2 ("the virus") was first detected in Wuhan, Hubei Province, People's Republic of China, causing outbreaks of the coronavirus disease COVID–19 that has now spread globally. The Secretary of Health and Human Services (HHS) declared a public health emergency on January 31, 2020, under section 319 of the Public Health Service Act (42 U.S.C. 247d), in response to COVID–19. I have taken sweeping action to control the spread of the virus in the United States, including by suspending entry of foreign nationals seeking entry who had been physically present within the prior 14 days in certain jurisdictions where COVID–19 outbreaks have occurred, including the People's Republic of China, the Islamic Republic of Iran, and the Schengen Area of Europe. The Federal Government, along with State and local governments, has taken preventive and proactive measures to slow the spread of the virus and treat those affected, including by instituting Federal quarantines for individuals evacuated from foreign nations, issuing a declaration pursuant to section 319F–3 of the Public Health Service Act (42 U.S.C. 247d–6d), and releasing policies to accelerate the acquisition of personal protective equipment and streamline bringing new diagnostic capabilities to laboratories. On March 11, 2020, the World Health Organization announced that the COVID–19 outbreak can be characterized as a pandemic, as the rates of infection continue to rise in many locations around the world and across the United States.

The spread of COVID–19 within our Nation's communities threatens to strain our Nation's healthcare systems. As of March 12, 2020, 1,645 people from 47 States have been infected with the virus that causes COVID–19. It is incumbent on hospitals and medical facilities throughout the country to assess their preparedness posture and be prepared to surge capacity and capability. Additional measures, however, are needed to successfully contain and combat the virus in the United States.

Now, Therefore, I, Donald J. Trump, President of the United States, by the authority vested in me by the Constitution and the laws of the United States of America, including sections 201 and 301 of the National Emergencies Act (50 U.S.C. 1601 et seq.)

and consistent with section 1135 of the Social Security Act (SSA), as amended (42 U.S.C. 1320b–5), do hereby find and proclaim that the COVID–19 outbreak in the United States constitutes a national emergency, beginning March 1, 2020. Pursuant to this declaration, I direct as follows:

Section 1. Emergency Authority. The Secretary of HHS may exercise the authority under section 1135 of the SSA to temporarily waive or modify certain requirements of the Medicare, Medicaid, and State Children's Health Insurance programs and of the Health Insurance Portability and Accountability Act Privacy Rule throughout the duration of the public health emergency declared in response to the COVID–19 outbreak.

Sec. 2. Certification and Notice. In exercising this authority, the Secretary of HHS shall provide certification and advance written notice to the Congress as required by section 1135(d) of the SSA (42 U.S.C. 1320b–5(d)).

Sec. 3. General Provisions.

(a) Nothing in this proclamation shall be construed to impair or otherwise affect:

(i) the authority granted by law to an executive department or agency, or the head thereof; or

(ii) the functions of the Director of the Office of Management and Budget relating to budgetary, administrative, or legislative proposals.

(b) This proclamation shall be implemented consistent with applicable law and subject to the availability of appropriations.

(c) This proclamation is not intended to, and does not, create any right or benefit, substantive or procedural, enforceable at law or in equity by any party against the United States, its departments, agencies, or entities, its officers, employees, or agents, or any other person.

In Witness Whereof, I have hereunto set my hand this thirteenth day of March, in the year of our Lord two thousand twenty, and of the Independence of the United States of America the two hundred and forty-fourth.

DONALD J. TRUMP

[Filed with the Office of the Federal Register, 8:45 a.m., March 17, 2020] NOTE: This proclamation was published in the Federal Register on March 18. Categories: Proclamations : Novel coronavirus disease (COVID–19) outbreak, national emergency declaration. Subjects: Diseases : Coronavirus, domestic prevention efforts. DCPD Number: DCPD202000156.

Federal Declarations for the COVID-19 Pandemic

The following summarizes the declarations of emergency authorized under the Public Health Service Act (PHSA), the National Emergencies Act (50 U.S.C. §§1601-1651), the Robert T. Stafford Disaster Relief and Emergency Assistance Act (Stafford Act), and the Small Business Act.

Figure 1. Chronology of Federal Declarations for the COVID-19 Pandemic

Under the Public Health Service Act, the Stafford Act, the National Emergencies Act, and the Small Business Act, as of June 1, 2021

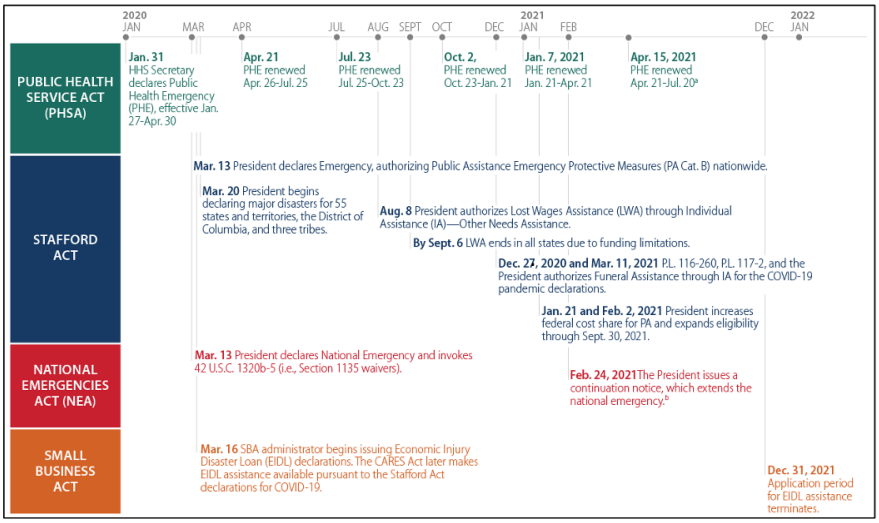


Figure 2.15.1.1: Federal Declarations for COVID-19. (Halchin, L. Elaine; Lee, Erica A.; Lindsay, Bruce R.; Lister, Sarah A.; Nagel, Jared C.; Sheikh, Hassan Z.; Webster, Elizabeth M. Congressional Research Service, "Federal Emergency and Major Disaster Declarations for the COVID-19 Pandemic", June 4, 2021. (report R46809).[3])

Accessibility Note

Figure 2.15.1.1 (Figure 1 from [3]) describes the chronology of federal declarations for the COVID-19 Pandemic in the following areas:

- Public Health Service Act (PHSA)
  - January 31, 2020: HHS Secretary declares Public Health Emergency (PHE)
  - PHE renewed essentially every three months thereafter throughout 2021
- Robert T. Stafford Disaster Relief and Emergency Assistance Act (Stafford Act)
  - March 13, 2020 President Trump declared Emergency, authorizing Public Assistance Emergency Protective Measures
  - March 20, 2020 President Trump begins declaring major disasters for 55 states and territories, the District of Columbia, and three tribes.
  - August 8, 2020 President Trump authorizes Lost Wages Assistance (LWA) through Individual Assistance (IA)-Other Needs Assistance
- National Emergencies Act (50 U.S.C. §§1601-1651)
  - March 31, 2020: President Trump declares National Emergency
  - February 21, 2021: President issues continuation notice
- Small Business Act.
  - March 16, 2020: SBA administrator begins issuing Economic Injury Disaster Loan (EIDL) declarations.

References

1. Congressional Research Service. National Emergency Powers, Updated March 23, 2020. CRS National Emergency Powers.
2. Elizabeth M. Webster, Congressional Research Service, "National Emergency Powers", Updated November 19, 2021. (report 98-505). CRS National Emergency Powers 98-505.
3. Halchin, L. Elaine; Lee, Erica A.; Lindsay, Bruce R.; Lister, Sarah A.; Nagel, Jared C.; Sheikh, Hassan Z.; Webster, Elizabeth M. Congressional Research Service, "Federal Emergency and Major Disaster Declarations for the COVID-19 Pandemic", June 4, 2021. (report R46809). Federal Emergency Declarations.
4. Proclamation 9994—Declaring a National Emergency Concerning the Novel Coronavirus Disease (COVID–19) Outbreak. March 13, 2020. COVID National Emergency.

This page titled 2.15.1: National Emergency Powers is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

2.15.2: Contracts

This section covers two major topics:

- An overview of contract types that are generally used. Some contract types can employ incentives, which might be useful in emergency situations such as the COVID-19 pandemic.
- Biomedical Advanced Research and Development Authority (BARDA) contracts in particular that were used for COVID-19.

This page titled 2.15.2: Contracts is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

2.15.2.1: Selecting the Type of Contract

Learning Objectives

1. Identify factors that determine which type of contract to select.
2. Describe the types of fixed cost contracts.
3. Describe the types of cost reimbursable contracts.

An agreement between the organization and an outside provider of a service or materials is a legal contract. To limit misunderstandings and make them more enforceable, contracts are usually written documents that describe the obligations of both parties.

Because legal agreements often create risk for the parent organization, procurement activities are often guided by the policies and procedures of the parent organization. After the project management team develops an understanding of what portions of the project work will be outsourced and defines the type of relationships that are needed to support the project execution plan, the procurement team begins to develop the contracting plan. On smaller, less complex projects, the contract development and execution is typically managed through the parent company or by a part-time person assigned to the project. On larger, more complex projects, the procurement team can consist of work teams within the procurement function with special expertise in contracting. The contract plan defines the relationship between the project and the subcontractors (supplier, vendor, or partner) and also defines a process for making changes in the agreement to accommodate changes that will occur on the project. This change management process is similar to the change management process used with the project agreement with the project client.

The contracting plan of the project supports the procurement approach of the project. The following are some factors to consider when selecting the type of contract:

- The uncertainty of the scope of work needed
- The party assuming the risk of unexpected cost increases
- The importance of meeting the scheduled milestone dates
- The need for predictable project costs

There are several types of contracting approaches and each supports different project environments and project approaches. The legal contracts that support the procurement plan consist of two general types of contract: the fixed price and the cost reimbursable contracts, with variations on each main type.

Fixed Price Contracts

The fixed price contract is a legal agreement between the project organization and an entity (person or company) to provide goods or services to the project at an agreed-on price. The contract usually details the quality of the goods or services, the timing needed to support the project, and the price for delivering goods or services. There are several variations of the fixed price contract. For commodities and goods and services where the scope of work is very clear and not likely to change, the fixed price contract offers a predictable cost. The responsibility for managing the work to meet the needs of the project is focused on the contractor. The project team tracks the quality and schedule progress to assure the contractors will meet the project needs. The risks associated with fixed price contracts are the costs associated with project change. If a change occurs on the project that requires a change order from the contractor, the price of the change is typically very high. Even when the price for changes is included in the original contract, changes on a fixed price contract will create higher total project costs than other forms of contracts because the majority of the cost risk is transferred to the contractor, and most contractors will add a contingency to the contract to cover their additional risk.

Fixed price contracts require the availability of at least two or more suppliers that have the qualifications and performance histories that assure the needs of the project can be met. The other requirement is a scope of work that is most likely not going to change. Developing a clear scope of work based on good information, creating a list of highly qualified bidders, and developing a clear contract that reflects that scope of work are critical aspects of a good fixed priced contract.

Fixed Total Cost Contract

If the service provider is responsible for incorporating all costs, including profit, into the agreed-on price, it is a fixed total cost contract. The contractor assumes the risks for unexpected increases in labor and materials that are needed to provide the service or materials and in the quantity of time and materials needed.

Fixed Price with Price Adjustment

The fixed price contract with price adjustment is used for unusually long projects that span years. The most common use of this type of contract is the inflation-adjusted price. In some countries, the value of its local currency can vary greatly in a few months, which affects the cost of local materials and labor. In periods of high inflation, the client assumes risk of higher costs due to inflation, and the contract price is adjusted based on an inflation index. The volatility of certain commodities can also be accounted for in a price adjustment contract. For example, if the price of oil significantly affects the costs of the project, the client can accept the oil price volatility risk and include a provision in the contract that would allow the contract price adjustment based on a change in the price of oil.

Fixed Price with Incentive Fee Contract

Fixed price with incentive fee is a contract type that provides an incentive for performing on the project above the established baseline in the contract. The contract might include an incentive for completing the work on an important milestone for the project. Often contracts have a penalty clause if the work is not performed according to the contract. For example, if the new software is not completed in time to support the start-up of a new plant, the contract might penalize the software company a daily amount of money for every day the software is late. This type of penalty is often used when the software is critical to the project and the delay will cost the project significant money.

Incentive Fee on Copper Mine Project

A project in South America to design and construct a copper mine would supply copper to several companies throughout the world. The copper that would be produced by the mine was sold before the mine was complete and ships were scheduled to make the delivery dates to processing plants. Any delay in the project would mean a delay in shipping and significant loss to the mine, the shipping company, and the plants that were expecting the copper. Including an incentive fee for completing the project on time and including the important subcontracts increased the likelihood that the mine would make copper deliveries on time.

Fixed Unit Price

If the service or materials can be measured in standard units, but the amount needed is not known accurately, the price per unit can be fixed—a fixed unit price contract. The project team assumes the responsibility of estimating the number of units used. If the estimate is not accurate, the contract does not need to be changed, but the project will exceed the budgeted cost.

Fixed Unit Price Contract for Concrete

An example of a fixed price contract is a contract for the concrete needed for the foundation of a building. The project contracted for the concrete company to supply 1,000 cubic yards (CY) at 5,000 PSI (hardness standard) of concrete at the project site according to, and in support of, the project schedule for $70 per square yard. This is an example of a unit price contract. If the project only uses 970 CY, then the total costs will be lower. If the project uses 1,050 CY, then the costs will be higher. An alternative pricing would be to establish a fixed price of $70,000 (1,000 CY × $70.00). Both the unit price approach and the total costs approach are fixed price contracts.

Table 2.15.2.1.1: Table of Fixed Price Contracts and Characteristics

Type	Known Scope	Share of Risk	Incentive for Meeting Milestones	Predictability of Cost

Fixed Total Cost	Very High	All Contractor	Low>	Very High
Fixed Unit Price	High	Mostly Project	Low	High
Fixed price with Incentive Fee	High	Mostly Project	High	Medium-high
Fixed Fee with Price Adjustment	High	Mostly Project	Low	Medium

Cost Reimbursable Contracts

In a cost reimbursable contract, the organization agrees to pay the contractor for the cost of performing the service or providing the goods. Cost reimbursable contracts are also known as cost plus contracts. Cost reimbursable contracts are most often used when the scope of work or the costs for performing the work are not well known. The project uses a cost reimbursable contract to pay the contractor for allowable expenses related to performing the work. Since the cost of the project is reimbursable, the contractor has much less risk associated with cost increases. When the costs of the work are not well known, a cost reimbursable contract reduces the amount of money the bidders place in the bid to account for the risk associated with potential increases in costs.

The contractor is also less motivated to find ways to reduce the cost of the project unless there are incentives for supporting the accomplishment of project goals. Cost reimbursable contracts also require good documentation of the costs that occurred on the project to assure that the contractor gets paid for all the work performed and to assure that the organization is not paying for something that was not completed.

Cost Reimbursable Contract to Drill Wells

A project to build a new plant in an area that did not have sufficient water included the drilling of water wells to produce several thousand gallons of water a day for the new plant. Although geological surveys indicated there was sufficient water to meet the plant's requirements, the number of wells needed was unknown. The project developed a cost reimbursable contract that paid the well drilling contractor for allowable costs associated with drilling the wells.

The contractor is paid an additional amount above the costs. There are several ways to compensate the contractor.

Cost Reimbursable Contract with Fixed Fee

A cost reimbursable contract with a fixed fee provides the contractor with a fee or profit amount that is determined at the beginning of the contract and does not change.

Fixed Fee for Providing Water

On the new water plant project, the project accountant reviewed each bill, including time cards for labor, invoices for materials, and other documents that supported the invoice. The contractor was then reimbursed for the allowable costs plus the administrative fee and a fixed amount for his profit.

Cost Reimbursable Contract with Percentage Fee

A cost reimbursable contract with a percentage fee pays the contractor for costs plus a percentage of the costs, such as 5 percent of total allowable costs. The contractor is reimbursed for allowable costs and is paid a fee.

Percentage Fee to Evaluate Dam in West Virginia

A small community in West Virginia was worried about the structural integrity of a dam above the town. The county council was worried the dam would break and cause loss of life and property. They contracted with a civil engineering firm to evaluate the dam structure and attest to the structural soundness. The firm hired an expert from outside the area and paid the expert $1,000.00 per day plus expenses such as meals, travel, and lodging. The civil engineering firm billed the community for the expert's fees and expenses plus 10 percent of the total.

Cost Reimbursable Contract with an Incentive Fee

A cost reimbursable contract with an incentive fee is used to encourage performance in areas critical to the project. Often the contract attempts to motivate contractors to save or reduce project costs. The use of the cost reimbursable contract with an incentive fee is one way to motivate cost reduction behaviors.

Incentive Fee for Road Project

A road construction company won a contract to build a small road to the new county courthouse. The estimate to complete the road was $10 million. The contract received a cost reimbursable contract that would pay all costs plus a 3 percent fee. The contactor could also earn an incentive by performing the work for less than $10 million. The contract might include a fee that would pay the contract 20 percent of all savings below the estimated $10 million. In this case, the county got the road at a lower cost, and the contractor made more money.

The contract could have focused on schedule and paid a bonus for completing ahead of schedule. This type of contract requires that the project management team has the capability to assure the quality of work performed meets project specifications and the savings was not generated through reducing the quality of the work.

Cost Plus Contract with Award Fee

A cost plus contract with award fee reimburses the contractor for all allowable costs plus a fee that is based on performance criteria. The fee is typically based on goals or objectives that are more subjective. An amount of money is set aside for the contractor to earn through excellent performance, and the decision on how much to pay the contractor is left to the judgment of the project team. The amount is sufficient to motivate excellent performance.

The following Reuters story is about the use of an award fee to incentivize the contractor's performance in maintaining the ship's performance during transfer to other owners.

> VSE Corporation (NASDAQ GS: VSEC) reported today that it has been awarded a $249 million cost-plus award fee contract option modification by the Naval Sea Systems Command that can be exercised by the Navy to provide one additional year of continued support to NAVSEA PMS 326 and 333 for ex-U.S. Navy ships that are sold, leased or otherwise transferred through the Foreign Military Sales (FMS) program to FMS clients.
> This contract provides for services supporting U.S. ships that are sold, leased or otherwise transferred to FMS clients by providing engineering, technical, procurement, logistics, test, inspection, calibration, repair, maintenance and overhaul support services, including reactivation and modernization.
> Since 1995, VSE's International Group, GLOBAL Division (formerly BAV Division) has transferred 42 ships to foreign governments. VSE is currently reactivating EX-USNS Andrew J. Higgins (TAO-190) for transfer to Chile. Additionally, VSE actively supports various countries through the follow-on technical support requirements of the contract, providing training, maintenance, repair, and in-country infrastructure improvement assistance in support of transferred ships. Countries currently supported by VSE include Bahrain, Egypt, Japan, Mexico, Taiwan, Turkey, Poland, Philippines, Italy and Romania.
> "We are extremely pleased to have won this award. It is a testament to the confidence and trust we have earned from the U.S. Navy and their foreign clients for the past 14 years," said VSE CEO/COO/President Maurice "Mo" Gauthier. "We look forward to continuing to deliver excellence and innovative technology solutions to the world's navies for years to come" (VSE Corporation, 2009).[1]

Time and Materials Contracts

On small activities that have a high uncertainty, the contractor might charge an hourly rate for labor, plus the cost of materials, plus a percentage of the total costs. This type of contract is called time and materials (T&M). Time is usually contracted on an hourly rate basis and the contractor usually submits time sheets and receipts for items purchased on the project. The project reimburses the contractor for the time spent based on an agreed-on rate and the actual cost of the materials. The fee is typically a percent of the total cost.

Time and materials contracts are used on projects for work that is smaller in scope and has uncertainty or risk, and the project rather than the contractor assumes the risk. Since the contractor will most likely include contingency in the price of other types of contracts to cover the high risk, T&M contracts provide lower total cost to the project.

Table 2.15.2.1.2: Table of Contract Types and Characteristics

Cost Reimbursable (CR)	Known Scope	Share of Risk	Incentive for Meeting Milestones	Predictability of Cost
CR with Fixed Fee	Medium	Mostly Project	Low	Medium-high
CR with Percentage Fee	Medium	Mostly Project	Low>	Medium-high
CR with Incentive Fee	Medium	Mostly Project	High	Medium
CR with Award Fee	Medium	Mostly Project	High	Medium
Time and Materials	Low	All Project	Low	Low

To minimize the risk to the project, the contract typically includes a not-to-exceed amount, which means the contract can only charge up to the agree amount. The T&M contract allows the project to make adjustments as more information is available. The final cost of the work is not known until sufficient information is available to complete a more accurate estimate.

Archeological Site Evaluation

On a road construction project, the survey team discovers an archeological site. Contractors are required to preserve archeological sites and the project team explores ways to advance the schedule while a decision is made on how to handle the site. The project issues a T&M contract to an archeologist to determine the nature of the site and develop a plan to preserve the integrity of the site. A T&M contract is awarded because the size and character of the site is unknown and the amount of time and the type of equipment needed to explore the site is also unknown. An archeologist from the local university was issued a T&M contract to cover the labor and expenses to explore the site and develop a plan. An hourly rate was established for each member of the five-person archeological team. Equipment rental fees plus 15 percent was paid on all equipment rented and supplies used. The archeological team's profit was incorporated into the labor rates. A not-to-exceed amount was also included in the contract to capture the team's estimate of the amount of work. A contract change order would increase the not-to-exceed number when more information was available.

Progress Payments and Change Management

Vendors and suppliers usually require payments during the life of the contract. On contracts that last several months, the contractor will incur significant cost and will want the project to pay for these costs as early as possible. Rather than wait until the end of the contract, a schedule of payments is typically developed as part of the contract and is connected to the completion of a defined amount of work or project milestones. These payments made before the end of the project and based on the progress of the work are called progress payments. For example, a concrete supplier on a construction project may bill the contract for the amount of concrete poured the previous month plus the profit earned during that period. On a training project, the contract might develop a payment schedule that pays for the development of the curriculum, and payment is made when the curriculum is completed and accepted. In each case, there is a defined amount of work to be accomplished, a time frame for accomplishing that work, and a quality standard the work must achieve before the contractor is paid for the work.

Just as the project has a scope of work that defines what is included in the project and what work is outside the project, vendors and suppliers have a scope of work that defines what they will produce or supply to the company. (Partners typically share the project scope of work and may not have a separate scope of work.) Often changes occur on the project that require changes in

the contractor's scope of work. How these changes will be managed during the life of the project is typically documented in the contract. Capturing these changes early, documenting what changed and how the change impacted the contract, and developing a change order (a change to the contract) are important to maintaining the progress of the project. Conflict among team members will often arise when changes are not documented or when the team cannot agree on the change. Developing and implementing an effective change management process for contractors and key suppliers will minimize this conflict and the potential negative effect on the project.

Key Takeaways

- Contract selection is based on uncertainty of scope, assignment of risk, need for predictable costs, and the importance of meeting milestone dates.
- Total fixed cost is a single price where the scope is well defined. A fixed price with incentive contract offers a reward for finishing early or under budget or a penalty for being late. A fixed price with adjustment allows for increases in cost of materials or changes in currency values. A fixed unit price contract sets a price per unit, but the exact number of units is not known.
- In a cost reimbursable contract, the project pays for costs. A cost plus fixed fee contract assures the contractor of a known fee. A cost plus percentage fee calculates the fee as a percentage of the costs. A cost plus incentive fee sets goals for the contractor to achieve that would result in a bonus. A cost plus award fee is similar, but the goals are more subjective. Time and materials contracts pay for costs plus an hourly rate for the contractor's time.

Exercises

Internalize your learning experience by preparing to discuss the following.

If you were a contractor, which type of contract would you prefer most and which would you like least? Explain your choices. Your explanation should demonstrate that you are familiar with the definitions of the contracts you chose and at least one similar type of contract.

1. A key factor in choosing the type of contract is the uncertainty of the _____, risk, cost, or schedule of the activity.
2. A contract with an _____ fee might reward the contractor for finishing early.
3. Contracts that pay the contractor's costs are _____ contracts.
4. Which type of contract is most appropriate to use if the scope is extremely well known, and which type is most appropriate if the scope is very uncertain? Explain your choices.
5. Why would a water well drilling company prefer a cost reimbursable contract versus a fixed cost contract?

References

1. VSE Corporation, "VSE Awarded $249 Million Cost-Plus Award Fee Contract Option Modification by Navy," Business Wire, August 24, 2009, VSE Cost plus Award Contract. (accessed October 5, 2009).

This page titled 2.15.2.1: Selecting the Type of Contract is shared under a CC BY-NC-SA 4.0 license and was authored, remixed, and/or curated by Anonymous via source content that was edited to the style and standards of the LibreTexts platform.

2.15.2.2: BARDA Contracts

Introduction to BARDA's COVID-19 Vaccine Portfolio for COVID-19

Biomedical Advanced Research and Development Authority (BARDA) in the Department of Health and Human Services (HHS) handled the major contracts since it is "responsible for the procurement and development of medical counter measures , principally against bioterrorism, including chemical, biological, radiological and nuclear (CBRN) threats, as well as pandemic influenza and emerging diseases."[1]

When the threat of COVID-19 emerged, BARDA quickly responded beginning in January 2020. BARDA's COVID-19 medical countermeasure strategy focused on pivoting existing, proven technologies to expedite medical countermeasure availability, leveraging flexible partnerships, de-risking development through advanced purchase agreements, and establishing domestic-based manufacturing capabilities.[3]

BARDA rapidly executed new partnerships and built a robust COVID-19 response focused on medical countermeasures, support services, and medical countermeasure manufacturing.[3]

As of June 2020, "seven companies had been chosen for funding from Operation Warp Speed to expedite development and preparation for manufacturing their respective vaccine candidates: Johnson & Johnson (Janssen Pharmaceutical), AstraZeneca-University of Oxford, Pfizer-BioNTech, Moderna, Merck, Vaxart, and Inovio. Funding from BARDA totaled more than $2 billion by the end of June, with the largest awards of $1.2 billion given to AstraZeneca and $483 million to Moderna."[1]

"In June 2020, BARDA and the U.S. Department of Defense signed a $143 million contract with SiO_2 Materials Science to ramp up production of vials and syringes used for COVID drugs and vaccines."[1] This was done because there were concerns during the early part of the pandemic response that there likely would be a shortage of vaccine vials due to the magnitude of ramp up for hundreds of mullions of doses. In this section we will look at the BARDA contract with Moderna and the BARDA contract with SiO_2 Materials Science.

Awarding Contracts

Instead of entering into contracts directly with vaccine pharmaceutical companies, more than $6 billion in Operation Warp Speed funding would be routed through Advanced Technology International, which then awarded contracts to companies working on COVID-19 vaccines. The contracts could be more tailored to the company's wants and needs and include fewer standard taxpayer protections because Advanced Technology International specializes in nontraditional government contracts.[2]



Figure 2.15.2.2.1: Operation Warp Speed Contract flow. (Robert Slate, CIS 79 course materials)

Accessibility Note

Figure 2.15.2.2.1 shows a blue box containing "Operation Warp Speed" with an arrow to a blue box containing "Advanced Technology International" with an arrow to a blue box containing "COVID-19 Vaccine companies."

BARDA's COVID-19 Vaccine Portfolio

The following are a few extracts from BARDA's COVID-19 Vaccine Portfolio:[3]

Pfizer Inc. USA

Product Name: COMIRNATY (COVID-19 Vaccine, mRNA)

Description:
Pfizer has conducted large-scale manufacturing and fill-finish of its COVID-19 mRNA vaccine, developed in collaboration with BioNTech, for distribution in the U.S. after Emergency Use Authorization and licensure were granted by the FDA.

Unique Impact:
Pfizer's COVID-19 vaccine is an mRNA vaccine that is delivered in lipid nanoparticles. RNA delivered by these protein-free lipid nanoparticles is believed to eliminate the risk of anti-vector immunity while boosting the level and duration of immune responses. Additionally, the use of mRNA technology allowed for rapid manufacturing scale-up that was critical for prevention of COVID-19.

Status: Licensed.

Contract Award Amounts

Base Award Amount:	$1,950,000,000.00 (July 21, 2020)
Mod/Option 1 Amount:	$2,011,282,500.00 (December 22, 2020)
Mod/Option 2 Amount:	$2,100,000,000.00 (February 11, 2021)
Mod/Option 3 Amount:	$4,869,750,000.00 (July 21, 2021)
Mod/Option 4 Amount:	$1,230,000,000.00 (October 22, 2021)

Mod/Option 5 Amount:	$3,200,000,000.00 (June 29, 2022)

Manufacturing Status: Commercial Scale Manufacturing

ModernaTX, Inc. USA

Product Name: Spikevax (COVID-19 Vaccine, mRNA)

Description:
FDA-approved mRNA COVID-19 vaccine that is indicated for active immunization to prevent COVID-19 in individuals 18 years of age or older.

Unique Impact:
Spikevax is the second fully approved COVID-19 vaccine. FDA-approved Spikevax (COVID-19 Vaccine, mRNA) and the EUA-authorized Moderna COVID-19 vaccine have the same formulation and can be used interchangeably.

Status: Licensed.

Contract Award Amounts

Base Award Amount:	$430,298,520.00 (April 16, 2020)
Mod/Option 1 Amount:	$53,000,000.00 (May 24, 2020)
Mod/Option 2 Amount:	$471,596,459.00 (July 25, 2020)
Mod/Option 3 Amount:	$1,525,000,000.00 (August 11, 2020)
Mod/Option 4 Amount:	$1,666,598,000.00 (December 11, 2020)
Mod/Option 5 Amount:	$1,750,000,000.00 (February 11, 2021)
Mod/Option 6 Amount:	$62,705,357.00 (March 12, 2021)
Mod/Option 7 Amount:	$236,364,615.00 (April 18, 2021)
Mod/Option 8 Amount:	$144,140,941.00 (June 15, 2021)
Mod/Option 9 Amount:	$3,303,993,662.60 (June 15, 2021)

Manufacturing Status: Commercial Scale Manufacturing

AstraZeneca USA

Product Name: AZD1222 (formerly ChAdOx1 nCoV-19 vaccine)

Description:
The AZD1222 vaccine consists of the replication-deficient simian adenovirus vector ChAdOx1, containing the structural surface glycoprotein (Spike protein) antigen of the SARS CoV-2 (nCoV-19)

Unique Impact:
The late stage clinical testing of the vaccine candidate in parallel with developing large scale manufacturing is intended to help enable the USG goal of making doses available by the end of 2020 and additional doses by the end of 1Q2021.

Status: Phase 3

Contract Award Amounts

Base Award Amount:	$1.6M (October 28, 2020)

Cost Share: Yes

Manufacturing Status: Commercial Scale Manufacturing

Janssen Pharmaceuticals, Inc. USA

Product Name: AD26.COV2.S - Viral Vector Vaccine for COVID-19

Description:
Support development to licensure of a vaccine to combat COVID-19 using Janssen's established Adenovirus 26 platform, plus demonstration of Janssen's capacity to perform large scale manufacturing domestically of up to 100 million doses of its SARS-CoV-2 vaccine, Ad26.COV2.S

Unique Impact:
Janssen has significant nonclinical and clinical experience and investment in the Ad26-based vaccine technology platform and will have significant domestic and global vaccine manufacturing capability, to maximize the probability of a successful vaccine and rapid deployment within the US and globally.

Updates:
Successful development of the prototype will demonstrate Janssen's ability to rapidly stand up and seamlessly transition to large scale manufacturing to meet surge requirements, and demonstrate capability to stockpile and distribute large quantities of the vaccine to respond when needed under EUA or FDA licensure.

Status: No Longer Authorized

Contract Award Amounts

Base Award Amount:	$20,634,721 (February 11, 2020)
Mod/Option 1 Amount:	$435,602,360.00 (March 27, 2020)
Mod/Option 2 Amount:	$1,001,650,000 (August 5, 2020)
Mod/Option 3 Amount:	$85,304,775.00 (August 21, 2020)
Mod/Option 4 Amount:	$454,310,546.00 (November 13, 2020)
Mod/Option 5 Amount:	$31,998,326 (March 25, 2021)

Manufacturing Status: Manufacturing Demonstration Project

Novavax Inc. USA

Product Name: NVX-CoV-2373 Vaccine for SARS-CoV-2

Description:
Manufacture and delivery of SARS-CoV-2 vaccine doses, NVX-CoV2373, which is suitable for use in humans under a sufficiently informed deployment strategy, and the advanced positioning of a stockpile of critical long lead raw materials for the Matrix-M adjuvant. The NVX-CoV-2373 vaccine is comprised of the Matrix-M™ adjuvant, and antigen (SARS-CoV-2 spike protein).

Unique Impact:
Successful development of the prototype will demonstrate Novavax's ability to rapidly stand up and seamlessly transition to large scale manufacturing to meet surge requirements, and demonstrate capability to stockpile and distribute large quantities of the vaccine to respond when needed under EUA or FDA licensure.

Updates:
BARDA supported FDA Emergency Use Authorization of this product.

Manufacturing Status: Demonstration Project

Project closed. Information current as of March 2024.

Note: No contract dollar amounts.

Merck and IAVI USA

Product Name: rVSVΔG-CoV2

Description:
Recombinant vesicular stomatitis virus (VSVΔG)-vectored vaccine expressing the spike protein of SARS-CoV-2.

Unique Impact:
This live, virus-vectored vaccine has the potential to deliver rapid onset to immunity with a single dose. The same vector was used in Merck's Ebola vaccine (ERVEBO®) which was licensed by the FDA in December 2019.

Status: Product no longer supported

Vaxart, Inc. USA

Product Name: Ad-5 Vectored COVID-19 Vaccine

Description:
Vaxart's vaccine uses an adenovirus serotype 5 (Ad-5) vector to express the spike protein of SARS-CoV-2. This project supports planning and execution of a Phase 2b clinical trial to evaluate this novel COVID-19 vaccine candidate, which will be administered orally to induce a mucosal immune response.

Unique Impact:
This Phase 2b trial will evaluate the capacity of an oral vaccine immune response to protect against symptomatic COVID-19. If successful, an oral pill vaccine could provide a needle-free, easier to administer option that has potential to improve vaccine coverage.

Status: Phase 2

Contract Award Amounts

Base Award Amount:	$452,884,477 (June 13, 2024)

Cost Share: No

Summary of BARDA funding

See the spreadsheet (summary tab, detail tab) in Reference 5.[5]

Summary (last updated Nov 2, 2020)

t	Count	Total Awarded	Average Award	Lowest Award	Highest Award	Proportion of Tota
	44	157,819,208	4,046,646	111,090	26,120,000	1.1%
	10	1,806,556,899	180,655,690	7,200,000	617,724,874	12.5%
	9	412,453,989	51,556,749	432,990	204,000,000	2.8%
	7	10,799,025,492	1,542,717,927	38,033,570	2,479,894,979	74.5%
	6	1,269,622,467	211,603,745	31,190,000	658,250,000	8.8%
	5	37,926,044	7,585,209	592,791	34,893,033	0.3%
able	5	8,216,736	1,643,347	699,583	3,833,363	0.1%
	86	14,491,620,835	181,145,260	111,090	2,479,894,979	

Table 2.15.2.2.1: BARDA Funding Summary (last updated Nov 2, 2020). (Zain Rizvi, BARDA Funding Tracker - Public Citizen [5])

References

1. Wikipedia - BARDA. Text is available under the Creative Commons Attribution-ShareAlike 4.0 License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.
2. NPR Pfizer. NPR. The U.S. Paid Billions To Get Enough COVID Vaccines Last Fall. What Went Wrong? August 25, 2021 (adapted)
3. BARDA COVID Vaccines. This site lists BARDA's COVID-19 vaccine portfolio.
4. BARDA. Medical Countermeasures for COVID.
5. Zain Rizvi. Public Citizen. Downloadable spreadsheet is available. Information updated through November 2, 2020. BARDA Funding tracker.
6. COVID-19 Contracts. This site has links to all COVID-19 related contracts.
7. HHS-BARDA-Moderna-Contract-75A50120C00034-16April2020-Version8Feb2021. April 16, 2020. (Less redacted version obtained via FOIA lawsuit on February 8, 2021). Moderna contract.
8. Transparency International Global Health. COVID-19 contracts by country. This site has links to COVID-19 contracts by country.
9. Advanced Technology International (ATI). ATI.

This page titled 2.15.2.2: BARDA Contracts is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

2.15.2.3: BARDA contract with Moderna

This section provides an overview of the BARDA contract with Moderna. The publicly available version(s) of the contract are scans of redacted pages in the contract obtained via Freedom of Information requests.

Section B.1 Brief Description of Supplies/Services describes the scope of the project.[1,page 4, extracted verbatim]

> "The Department of Health and Human Services (HHS), Office of the Assistant Secretary for Preparedness and Response (ASPR), Biomedical Advanced Research and Development Authority (BARDA) requires the contractor(s) to develop a mRNA vaccine to licensure for the prevention of COVID-19. The project will entail pre-clinical and Phase 2 and Phase 3 clinical studies sufficient to demonstrate the safety and efficacy of the proposed vaccine(s); CMC development, scale-up, scale-out and validation of manufacturing capacities, including bulk drug substance and fill and finished drug product, with a capacity of 100 million doses by 2021 and all program management and regulatory activities necessary to achieve FDA licensure of the vaccine. The project shall be accomplished on an accelerated timeline, with parallel activity WBS, aggressive manufacturing scale-up, risk management, and taking advantage of any regulatory flexibilities. Contract terms include a requirement for domestic production of vaccine and assurance of material sourcing for vaccine production during execution of the project."

There is a discussion in the project management section of this textbook chapter that addresses the considerations of "accelerated timeline, with parallel activity WBS, aggressive manufacturing scale-up, risk management, and taking advantage of any regulatory flexibilities."

CPFF Contract

Section B.3 states that it is a **Cost-Plus-Fixed-Fee (CPFF)** contract.[1,page 4]

In contracting with the United States government, the following information is useful to know (extracted verbatim from the Federal Acquisition Regulation (FAR) [3], edited as needed).

FAR 16.306 Cost-plus-fixed-fee contracts

a. *Description*. A cost-plus-fixed-fee contract is a cost-reimbursement contract that provides for payment to the contractor of a negotiated fee that is fixed at the inception of the contract. The fixed fee does not vary with actual cost, but *may* be adjusted as a result of changes in the work to be performed under the contract. This contract type permits *contracting* for efforts that might otherwise present too great a risk to contractors, but it provides the contractor only a minimum incentive to control costs.
b. Application.
    1. A cost-plus-fixed-fee contract is suitable for use when the conditions of 16.301-2 are present and, for example-
        i. The contract is for the performance of research or preliminary exploration or study, and the level of effort required is unknown; or
        ii. The contract is for development and test, and using a cost-plus-incentive-fee contract is not practical.
    2. A cost-plus-fixed-fee contract normally *should* not be used in development of *major systems* (see part 34) once preliminary exploration, studies, and risk reduction have indicated a high degree of probability that the development is achievable and the Government has established reasonably firm performance objectives and schedules.
c. *Limitations*. No cost-plus-fixed-fee contract *shall* be awarded unless the *contracting officer* complies with all limitations in 15.404-4(c)(4)(i) and 16.301-3.
d. *Completion and term forms*. A cost-plus-fixed-fee contract *may* take one of two basic forms-completion or term.
    1. The completion form describes the scope of work by stating a definite goal or target and specifying an *end product*. This form of contract normally requires the contractor to complete and deliver the specified *end product* (*e.g.,*a final report of research accomplishing the goal or target) within the estimated cost, if possible, as a condition for payment of the entire fixed fee. However, in the event the work cannot be completed within the estimated cost, the Government *may* require more effort without increase in fee, provided the Government increases the estimated cost.

2. The term form describes the scope of work in general terms and obligates the contractor to devote a specified level of effort for a stated time period. Under this form, if the performance is considered satisfactory by the Government, the fixed fee is payable at the expiration of the agreed-upon period, upon contractor statement that the level of effort specified in the contract has been expended in performing the contract work. Renewal for further periods of performance is a new *acquisition* that involves new cost and fee arrangements.
3. Because of the differences in obligation assumed by the contractor, the completion form is preferred over the term form whenever the work, or specific milestones for the work, can be defined well enough to permit development of estimates within which the contractor can be expected to complete the work.
4. The term form *shall* not be used unless the contractor is obligated by the contract to provide a specific level of effort within a definite time period.

Contract Periods

The contract defines several periods, where CLIN stands for Contract Line Item Number:

1. Base period
    - CLIN 0001: pre-award cost
    - CLIN 0002: development of the mRNA vaccine to BLA
2. Option period CLIN 0003: Domestic Manufacturing Scale-out

The total contract value for the above items is $483,298,520.00.

Moderna Contract Deliverables/WBS

Deliverables, recreated from Ref 1, page 34.

WBS item	Item Description	Deliverable
1	mRNA Vaccine Development	
1.1	Program Management	
1.1.1	Program and Alliance Management	• Management Plans; Routine Reporting Deliverables
1.2	Nonclinical Toxicology	
1.2.2	Safety	
1.2.2.1	Development and Reproductive Toxicology	• Final Study Report
1.3	Nonclinical	
1.3.1	Model Development (reserved)	
1.3.1.2	NHP Efficacy Study	• Final Study Report
1.3.1.3	Mouse Efficacy Study	• Final Study Report
1.4	Clinical	
1.4.2	Phase 2	

WBS item	Item Description	Deliverable
1.4.2.1	Phase 2 Safety and Immunogenicity Study	• Clinical Study Protocol • Final Clinical Study Report
1.4.3	Phase 3	
1.4.3.1	Phase 3 Efficacy or Safety and Imrnunogenicity	• Clinical Study Protocol • Final Clinical Study Report
1.4.3.2	Phase 3 Lot-to-Lot	• Clinical Study Protocol • Final Clinical Study Report
1.4.3.3	Phase 3 Adolescents	• Clinical Study Protocol • Final Clinical Study Report
1.5	Regulatory	
1.5.1	IND	
1.5.1.1	IND Filing	• NA
1.5.1.2	IND Maintenance	• Record of FDA Communications
1.5.2	BLA	
1.5.1.1	BLA Submission	• NA
1.6 CMC		
1.6.3	Pilot Scale Manufacturing	
1.6.3.2	CTM Manufacture for P201	• CoA for Clinical Lots
1.6.3.4	CTM Manufacture for P301	• CoA for Clinical Lots
1.6.3.6	CTM Manufacture for P302/P303	• CoA for Clinical Lots

Section C.2 Statement of Work[1] lists WBS elements as well, and is covered in the Project Management Considerations section of this textbook.

Intellectual Property

Who owns the intellectual property produced under the contract? Is there a section of the contract that addresses this? Let's investigate this.

Section B.4.6 only addresses Facility, Equipment, and Product (e.g., vaccines, validated lots) ownership.

It might be addressed under Section B.4.13 Limited Rights Data, but the types of information covered are all redacted.[1, page 8]

> Section 1 Intellectual Property [1, page 15] states (included verbatim): "The parties agree that that data generated prior to entering into or outside the scope of the agreement will, when delivered to the USG, be considered to be limited rights data subject to restrictions covered under FAR Clause 52.227-14 Alt II paragraph(g)(3). The government will obtain unlimited rights to data funded under this contract pursuant tp FAR Clause 52.227-14. The parties rights to subject inventions developed during performance of this contract will be governed by the terms of FAR Clause 52.227-11."

Federal Acquisition Regulation (FAR) Clause 52.227-11 [4] rendered for accessibility

52.227-11 Patent Rights-Ownership by the Contractor.

Patent Rights-Ownership by the Contractor (*May* 2014)

a. As used in this clause-

Invention means any invention or discovery that is or *may* be patentable or otherwise protectable under title 35 of the U.S. Code, or any variety of plant that is or *may* be protectable under the Plant Variety Protection Act (7 U.S.C. 2321, *et seq.*)

Made means-

1. When used in relation to any invention other than a plant variety, the conception or first actual reduction to practice of the invention; or
2. When used in relation to a plant variety, that the Contractor has at least tentatively determined that the variety has been reproduced with recognized characteristics.

Nonprofit organization means a university or other institution of higher education or an organization of the type described in section 501(c)(3) of the Internal Revenue Code of 1954 (26 U.S.C. 501(c)) and exempt from taxation under section 501(a) of the Internal Revenue Code (26 U.S.C. 501(a)), or any nonprofit scientific or educational organization qualified under a State *nonprofit organization* statute.

"Practical application" means to manufacture, in the case of a composition of product; to practice, in the case of a process or method; or to operate, in the case of a machine or system; and, in each case, under such conditions as to establish that the invention is being utilized and that its benefits are, to the extent permitted by law or Government regulations, available to the public on reasonable terms.

Subject invention means any invention of the Contractor *made* in the performance of work under this contract.

b. *Contractor's rights.*

    1. *Ownership*. The Contractor *may* retain ownership of each *subject invention* throughout the world in accordance with the provisions of this clause.
    2. *License.*

        i. The Contractor *shall* retain a nonexclusive royalty-free license throughout the world in each *subject invention* to which the Government obtains title, unless the Contractor fails to disclose the invention within the times specified in paragraph (c) of this clause. The Contractor's license extends to any domestic subsidiaries and *affiliate*s within the corporate structure of which the Contractor is a part, and includes the right to grant sublicenses to the extent the Contractor was legally obligated to do so at contract award. The license is transferable only with the written approval of the agency, except when transferred to the successor of that part of the Contractor's business to which the invention pertains.

        ii. The Contractor's license *may* be revoked or modified by the agency to the extent necessary to achieve expeditious practical application of the *subject invention* in a particular country in accordance with the procedures in FAR 27.302(i)(2) and 27.304-1(f).

c. *Contractor's obligations.*

1. The Contractor *shall* disclose *in writing* each *subject invention* to the *Contracting Officer* within 2 months after the inventor discloses it *in writing* to Contractor personnel responsible for patent matters. The disclosure *shall* identify the inventor(s) and this contract under which the *subject invention* was *made*. It *shall* be sufficiently complete in technical detail to convey a clear understanding of the *subject invention*. The disclosure *shall* also identify any publication, on sale (*i.e.*, sale or *offer* for sale), or public use of the *subject invention*, or whether a manuscript describing the *subject invention* has been submitted for publication and, if so, whether it has been accepted for publication. In addition, after disclosure to the agency, the Contractor *shall* promptly notify the *Contracting Officer* of the acceptance of any manuscript describing the *subject invention* for publication and any on sale or public use.
2. The Contractor *shall* elect *in writing* whether or not to retain ownership of any *subject invention* by notifying the *Contracting Officer* within 2 years of disclosure to the agency. However, in any case where publication, on sale, or public use has initiated the 1-year statutory period during which valid patent protection can be obtained in the *United States*, the period for election of title *may* be shortened by the agency to a date that is no more than 60 days prior to the end of the statutory period.
3. The Contractor *shall* file either a provisional or a nonprovisional patent application or a Plant Variety Protection Application on an elected *subject invention* within 1 year after election. However, in any case where a publication, on sale, or public use has initiated the 1-year statutory period during which valid patent protection can be obtained in the *United States*, the Contractor *shall* file the application prior to the end of that statutory period. If the Contractor files a provisional application, it *shall* file a nonprovisional application within 10 months of the filing of the provisional application. The Contractor *shall* file patent applications in additional countries or international patent offices within either 10 months of the first filed patent application (whether provisional or nonprovisional) or 6 months from the date permission is granted by the Commissioner of Patents to file foreign patent applications where such filing has been prohibited by a Secrecy Order.
4. The Contractor *may* request extensions of time for disclosure, election, or filing under paragraphs (c)(1), (c)(2), and (c)(3) of this clause.

d. *Government's rights*-

1. *Ownership*. The Contractor *shall* assign to the agency, on written request, title to any *subject invention*-

   i. If the Contractor fails to disclose or elect ownership to the *subject invention* within the times specified in paragraph (c) of this clause, or elects not to retain ownership; provided, that the agency *may* request title only within 60 days after learning of the Contractor's failure to disclose or elect within the specified times.
   ii. In those countries in which the Contractor fails to file patent applications within the times specified in paragraph (c) of this clause; provided, however, that if the Contractor has filed a patent application in a country after the times specified in paragraph (c) of this clause, but prior to its receipt of the written request of the agency, the Contractor *shall* continue to retain ownership in that country.
   iii. In any country in which the Contractor decides not to continue the prosecution of any application for, to pay the maintenance fees on, or defend in reexamination or opposition proceeding on, a patent on a *subject invention*.

2. *License*. If the Contractor retains ownership of any *subject invention*, the Government *shall* have a nonexclusive, nontransferable, irrevocable, paid-up license to practice, or have practiced for or on its behalf, the *subject invention* throughout the world.

e. *Contractor action to protect the Government's interest*.

1. The Contractor *shall* execute or have executed and promptly deliver to the agency all instruments necessary to-

   i. Establish or confirm the rights the Government has throughout the world in those *subject inventions* in which the Contractor elects to retain ownership; and
   ii. Assign title to the agency when requested under paragraph (d) of this clause and to enable the Government to obtain patent protection and plant variety protection for that *subject invention* in any country.

2. The Contractor *shall* require, by written agreement, its employees, other than clerical and nontechnical employees, to disclose promptly *in writing* to personnel identified as responsible for the administration of patent matters and in the Contractor's format, each *subject invention* in order that the Contractor can comply with the disclosure provisions of paragraph (c) of this clause, and to execute all papers necessary to file patent applications on *subject inventions* and to establish the Government's rights in the *subject inventions*. The disclosure format *should*

require, as a minimum, the information required by paragraph (c)(1) of this clause. The Contractor *shall* instruct such employees, through employee agreements or other suitable educational programs, as to the importance of reporting inventions in sufficient time to permit the filing of patent applications prior to U.S. or foreign statutory bars.

3. The Contractor *shall* notify the *Contracting Officer* of any decisions not to file a nonprovisional patent application, continue the prosecution of a patent application, pay maintenance fees, or defend in a reexamination or opposition proceeding on a patent, in any country, not less than 30 days before the expiration of the response or filing period required by the relevant patent office.
4. The Contractor *shall* include, within the specification of any *United States* nonprovisional patent or plant variety protection application and any patent or plant variety protection certificate issuing thereon covering a *subject invention*, the following statement, "This invention was *made* with Government support under (identify the contract) awarded by (identify the agency). The Government has certain rights in the invention."

f. *Reporting on utilization of subject inventions*. The Contractor *shall* submit, on request, periodic reports no more frequently than annually on the utilization of a *subject invention* or on efforts at obtaining utilization of the *subject invention* that are being *made* by the Contractor or its licensees or assignees. The reports *shall* include information regarding the status of development, date of first commercial sale or use, gross royalties received by the Contractor, and other data and information as the agency *may* reasonably specify. The Contractor also *shall* provide additional reports as *may* be requested by the agency in connection with any march-in proceeding undertaken by the agency in accordance with paragraph (h) of this clause. The Contractor also *shall* mark any utilization report as confidential/proprietary to help prevent inadvertent release outside the Government. As required by 35 U.S.C. 202(c)(5), the agency will not disclose that information to persons outside the Government without the Contractor's permission.
g. *Preference for United States industry*. Notwithstanding any other provision of this clause, neither the Contractor nor any assignee *shall* grant to any person the exclusive right to use or sell any *subject invention* in the *United States* unless the person agrees that any *products* embodying the *subject invention* or produced through the use of the *subject invention* will be manufactured substantially in the *United States*. However, in individual cases, the requirement for an agreement *may* be waived by the agency upon a showing by the Contractor or its assignee that reasonable but unsuccessful efforts have been *made* to grant licenses on similar terms to potential licensees that would be likely to manufacture substantially in the *United States*, or that under the circumstances domestic manufacture is not commercially feasible.
h. *March-in rights*. The Contractor acknowledges that, with respect to any *subject invention* in which it has retained ownership, the agency has the right to require licensing pursuant to 35 U.S.C. 203 and 210(c), and in accordance with the procedures in 37 CFR 401.6 and any supplemental regulations of the agency in effect on the date of contract award.
i. Special provisions for contracts with *nonprofit organizations*. If the Contractor is a *nonprofit organization*, it *shall*-

1. Not assign rights to a *subject invention* in the *United States* without the written approval of the agency, except where an assignment is *made* to an organization that has as one of its primary functions the management of inventions, *provided*, that the assignee *shall* be subject to the same provisions as the Contractor;
2. Share royalties collected on a *subject invention* with the inventor, including Federal employee co-inventors (but through their agency if the agency deems it appropriate) when the *subject invention* is assigned in accordance with 35 U.S.C. 202(e) and 37 CFR 401.10;
3. Use the balance of any royalties or income earned by the Contractor with respect to *subject inventions*, after payment of expenses (including payments to inventors) incidental to the administration of *subject inventions* for the support of scientific research or education; and
4. Make efforts that are reasonable under the circumstances to attract licensees of *subject inventions* that are small business concerns, and give a preference to a small business concern when licensing a *subject invention* if the Contractor determines that the small business concern has a plan or proposal for marketing the invention which, if executed, is equally as likely to bring the invention to practical application as any plans or proposals from applicants that are not small business concerns; *provided*, that the Contractor is also satisfied that the small business concern has the capability and resources to carry out its plan or proposal. The decision whether to give a preference in any specific case will be at the discretion of the Contractor.
5. Allow the Secretary of Commerce to review the Contractor's licensing program and decisions regarding small business applicants, and negotiate changes to its licensing policies, procedures, or practices with the Secretary of Commerce when the Secretary's review discloses that the Contractor could take reasonable steps to more effectively implement the requirements of paragraph (i)(4) of this clause.

j. *Communications. [Complete according to agency instructions.]*
k. *Subcontracts.*

1. The Contractor *shall* include the substance of this clause, including this paragraph (k), in all subcontracts for experimental, developmental, or research work to be performed by a small business concern or *nonprofit organization*.
2. The Contractor *shall* include in all other subcontracts for experimental, developmental, or research work the substance of the patent rights clause required by FAR subpart 27.3.
3. At all tiers, the patent rights clause *must* be modified to identify the parties as follows: references to the Government are not changed, and the subcontractor has all rights and obligations of the Contractor in the clause. The Contractor *shall* not, as part of the consideration for awarding the subcontract, obtain rights in the subcontractor's *subject inventions*.
4. In subcontracts, at any tier, the agency, the subcontractor, and the Contractor agree that the mutual obligations of the parties created by this clause constitute a contract between the subcontractor and the agency with respect to the matters covered by the clause; provided, however, that nothing in this paragraph is intended to confer any jurisdiction under the Contract Disputes statute in connection with proceedings under paragraph (h) of this clause.

(End of clause)

Alternate I (Jun 1989). As prescribed in 27.303 (b)(3), add the following sentence at the end of paragraph (d)(2) of the basic clause:

The license *shall* include the right of the Government to sublicense foreign governments, their nationals and international organizations pursuant to the following treaties or international agreements: _____*

[**Contracting Officer complete with the names of applicable existing treaties or international agreements. The above language is not intended to apply to treaties or agreements that are in effect on the date of the award but are not listed.*]

Alternate II (Dec 2007). As prescribed in 27.303 (b)(4), add the following sentence at the end of paragraph (d)(2) of the basic clause:

The agency reserves the right to unilaterally amend this contract to identify specific treaties or international agreements entered into by the Government before or after the effective date of the contract and effectuate those license or other rights that are necessary for the Government to meet its obligations to foreign governments, their nationals, and international organizations under the treaties or international agreements with respect to *subject inventions made* after the date of the amendment.

Alternate III (Jun 1989). As prescribed in 27.303 (b)(5), substitute the following paragraph (i)(3) in place of paragraph (i)(3) of the basic clause:

3. After payment of patenting costs, licensing costs, payments to inventors, and other expenses incidental to the administration of *subject inventions*, the balance of any royalties or income earned and retained by the Contractor during any fiscal year on *subject inventions* under this or any successor contract containing the same requirement, up to any amount equal to 5 percent of the budget of the facility for that fiscal year, *shall* be used by the Contractor for the scientific research, development, and education consistent with the research and development mission and objectives of the facility, including activities that increase the licensing potential of other inventions of the facility. If the balance exceeds 5 percent, 75 percent of the excess above 5 percent *shall* be paid by the Contractor to the Treasury of the *United States* and the remaining 25 percent *shall* be used by the Contractor only for the same purposes as described above. To the extent it provides the most effective technology transfer, the licensing of *subject inventions shall* be administered by Contractor employees on location at the facility.

Alternate IV (Jun 1989). As prescribed in 27.303 (b)(6), include the following paragraph (e)(5) in paragraph (e) of the basic clause:

5. The Contractor *shall* establish and maintain active and effective procedures to ensure that *subject inventions* are promptly identified and timely disclosed, and *shall* submit a description of the procedures to the *Contracting Officer* so that the *Contracting Officer may* evaluate and determine their effectiveness.

Alternate V (Dec 2007). As prescribed in 27.303 (b)(7), include the following paragraph (d)(3) in paragraph (d) of the basic clause:

(d)(3) *CRADA licensing.* If the Contractor performs services at a Government owned and operated laboratory or at a Government owned and Contractor operated laboratory directed by the Government to fulfill the Government's obligations under a Cooperative Research and Development Agreement (CRADA) authorized by 15 U.S.C. 3710a, the Government *may* require the Contractor to negotiate an agreement with the CRADA collaborating party or parties regarding the allocation of rights to any *subject invention* the Contractor makes, solely or jointly, under the CRADA. The agreement *shall* be negotiated prior to the Contractor undertaking the CRADA work or, with the permission of the Government, upon the identification of a *subject invention*. In the absence of such an agreement, the Contractor agrees to grant the collaborating party or parties an *option* for a license in its inventions of the same scope and terms set forth in the CRADA for inventions *made* by the Government.

Use of Earned Value Management

Section B.4.3 Earned Value Management (EVM) Lite Requirements states: "The contractor shall use an Earned Value Management (EVM) System for all retrofit and development activities of the anticipate requirement, that is consistent with the '7 Principles of Earned Value Management, Tier 2 System Implementation Intent Guide' attached to this contract. Alternative systems may be submitted to the Contracting Officer for consideration and approval." [1, page 6]

Attachment 11 [1, page 83ff] is entitled "7 Principles of Earned Value Management, Tier 2 System Implementation Intent Guide." December 21, 2011. Its origins are the Department of Health & Human Services HHS Office of the Assistant Secretary for Preparedness ASPR, Biomedical Advanced Research and Development Authority BARDA.

BARDA's plan to implement Earned Value Management (for project management) is for three tiers, and tier 2 would apply to the BARDA vaccine work. Tier 2 is for countermeasure R&D work contracts that have total acquisition costs greater than or equal to $25 million and have a Technical Readiness Level (TRL) of less than 7. Those contracts will apply EVM principles for tracking cost, schedule and technical performance that comply with the 7 Principles of EVM Implementation (extracted verbatim below):

1. Plan all work scope to completion.
2. Break down the program work scope into finite pieces that can be assigned to a responsible person or organization for control of tech ical, schedule and cost objectives.
3. Integrate program work scope, schedule, and cost objectives into a performance measurement baseline plan against which accomplishments can be measured. Control changes to the baseline.
4. Use actual costs incurred and recorded in accomplishing the work performed.
5. Objectively assess accomplishments at the work performance level.
6. Analyze significant variances from the plan, forecast impacts, and pepare an estimate at completion based on performance to date and work to be performed.
7. Use earned value information in the company's management processes.

References

1. HHS-BARDA-Moderna-Contract-75A50120C00034-16April2020-Version8Feb2021. April 16, 2020. (Less redacted version obtained via FOIA lawsuit on February 8, 2021). BARDA Moderna contract. Note: Converted numbering to use formal numbered list or lettered lists for accessibility reasons.
2. Transparency International Global Health. COVID-19 contracts. This site has links to COVID-19 contracts by country.
3. Cost-plus-fixed-fee contracts. Information pulled January 4, 2026. Section 16.306 is about Cost-plus-fixed-fee contracts. FAR is "Federal Acquisition Regulation."
4. Patent Rights Ownership by Contractor. Accessed January 4, 2026.

This page titled 2.15.2.3: BARDA contract with Moderna is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

2.15.2.4: BARDA Contract with SiO2 Medical Products, Inc. for Vials

Caution

The contract does not seem to state the number of vials being ordered, perhaps due to redactions.

"In June 2020, BARDA and the U.S. Department of Defense signed a $143 million contract with SiO_2 Materials Science to ramp up production of vials and syringes used for COVID drugs and vaccines."[1] This was done because there were concerns during the pandemic response that there likely would be a shortage of vaccine vials due to the magnitude of ramp up for hundreds of mullions of doses.

Relevant extracts from the contract[2] appear below.

This is a TECHNOLOGY INVESTMENT AGREEMENT contract between SIO2 MEDICAL PRODUCTS INC, Auburn AL and the U.S. Army Contracting Command –Aberdeen Proving Ground, Research Triangle Park Division (ACC-APG-RTP) on behalf of the Biomedical Advanced Research and Development Authority (BARDA) for Developing a Domestic Continental U.S.-Based Hybrid Pharmaceutical Vial.

Cost/Amount

Cost Sharing 8.1 Total value of the TIA means the total amount of costs that are currently expected to be charged to the award over its life, which includes amounts for the Federal share and any non-Federal cost sharing or matching required under the award; and any options, even if not yet exercised, for which the costs have been established in the award. The Total Amount of Government Funding for the Agreement is $143,000,000.00; the Total Cost Share for the Agreement is $128,438,496.00, and; the Total Estimated Value of the Agreement is $271,438,496.00

Principal Purpose: The Government and the Recipient agree that the principal purpose of this Agreement is for Government investment into the expansion of Recipient's existing capacities to support research efforts in pursuit of domestic development and distribution of a vaccine in response to the worldwide COVID-19 pandemic. SIO2 Medical Products Inc.'s proposal titled "Developing a Domestic Continental U.S.-Based Hybrid Pharmaceutical Vial.," dated 26 May 2020, including all revisions and appendices, is fully incorporated (the "Proposal" or "Project"). This Agreement is not intended to be, nor shall it be construed as, by implication or otherwise, a partnership, a corporation, or other business organization.

This Agreement is a fixed-support type Technology Investment Agreement (TIA) as described in 32 CFR §37.300. Fixed-support means that the amount of assistance established at the time of award is not meant to be adjusted later if the research project is carried out to completion. Payments will be based on Payable Milestones as defined in 32 CFR 37.815. These are payments made according to a schedule that is based on predetermined measures of technical progress or other payable milestones.

Intellectual Property

11. Intellectual Property & Patent Rights

11.1 Background IP and Materials. The Recipient and the Government each retain any intellectual property (IP) rights to their own materials, data, technology, information, documents, or know-how—or potential rights, such as issued patents, patent applications, invention disclosures, or other written documentation—that exist prior to execution of this Agreement or are developed outside the scope of this Agreement (Background IP).

11.2 Authorization and Consent for Non-commercial Products. The Government authorizes and consents to all use and manufacture, in performance of this Agreement, of any invention described in and covered by a United States patent, except for deliverables under this Agreement that are commercially available to the public by the Recipient.

11.3 Ownership. Ownership of any invention, regardless of whether it is not patentable, held as a trade secret or is patentable under U.S. patent law that is conceived or first reduced to practice under this Agreement will follow inventorship in accordance with U.S. patent law. The Parties represent and warrant that each inventor will assign his or her rights in any such inventions to his or her employing organization.

11.4 Patent Applications. Irrespective of any Disclosure of Information clauses in this Agreement the Parties will respectively have the option to file a patent application claiming any Invention made solely by their respective employees. The Parties will consult with each other regarding the options for filing a patent application claiming a joint Invention. Within two months of being notified of the discovery of an invention or filing a patent application covering an Invention, each Party will provide notice of such discovery or filing to the other Party. The Parties will reasonably cooperate with each other in the preparation, filing, and prosecution of any patent application claiming an Invention. Any Party filing a patent application will bear expenses associated with filing and prosecuting the application, as well as maintaining any patents that issue from the application, unless otherwise agreed by the Parties.

11.5 Licenses. Upon the Recipient's request, the Government agrees to enter into good faith negotiations regarding the receipt of a nonexclusive commercialization license covering the Government's interest in any Invention made in whole or in part by a Government employee. Any Invention made by a Recipient employee is subject to a nonexclusive, nontransferable, irrevocable, paid-up license for the Government to practice and have practiced the Invention.

11.6 Executive Order No. 9424 of 18 February 1944 requires all executive Departments and agencies of the Government to forward through appropriate channels to the Commissioner of Patents and Trademarks, for recording, all Government interests in patents or applications for patents. If there is any conflict between this section and the Bayh-Dole Act, the statute prevails.

References

1. Wikipedia - BARDA. Wikipedia. BARDA.
2. SIO2 Contract. Downloaded January 4, 2026.

This page titled 2.15.2.4: BARDA Contract with SiO2 Medical Products, Inc. for Vials is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

2.15.3: Risk Mitigation and Liability

To lower the risk to private firms of making and marketing vaccines, and thereby increase their supply, governments sometimes make laws intended to reduce or eliminate vaccine suppliers' legal liability. One such law is the Public Readiness and Emergency Preparedness Act (or PREP Act) of 2005, codified as part of the federal law dealing with public health emergencies (mostly at 42 U.S.C. § 247d-6d and 247d-6e).

The Public Readiness and Emergency Preparedness Act (PREP Act) authorizes the Secretary of the Department of Health and Human Services to issue a PREP Act declaration. The declaration provides immunity from liability (except for willful misconduct) for claims:

- of loss caused, arising out of, relating to, or resulting from administration or use of countermeasures to diseases, threats and conditions
- determined by the Secretary to constitute a present, or credible risk of a future public health emergency
- to entities and individuals involved in the development, manufacture, testing, distribution, administration, and use of such countermeasures[1]

A PREP Act declaration is specifically for the purpose of providing immunity from liability, and is different from, and not dependent on, other emergency declarations.[1]

COVID-19 PREP Act Declarations

The Declaration Under the Public Readiness and Emergency Preparedness Act for Medical Countermeasures Against COVID-19 was published on March 17, 2020 to provide liability immunity for activities related to medical countermeasures against COVID-19.[2]

The Public Readiness and Emergency Preparedness Act (PREP Act) authorizes the Secretary of Health and Human Services (the Secretary) to issue a Declaration to provide liability immunity to certain individuals and entities (Covered Persons) against any claim of loss caused by, arising out of, relating to, or resulting from the manufacture, distribution, administration, or use of medical countermeasures (Covered Countermeasures), except for claims involving “willful misconduct” as defined in the PREP Act. This Declaration is subject to amendment as circumstances warrant.[2]

The PREP Act was enacted on December 30, 2005, as Public Law 109-148, Division C, Section 2. It amended the Public Health Service (PHS) Act, adding Section 319F-3, which addresses liability immunity, and Section 319F-4, which creates a compensation program. These sections are codified at 42 U.S.C. 247d-6d and 42 U.S.C. 247d-6e, respectively.[2]

The Pandemic and All-Hazards Preparedness Reauthorization Act (PAHPRA), Public Law 113-5, was enacted on March 13, 2013. Among other things, PAHPRA added sections 564A and 564B to the Federal Food, Drug, and Cosmetic (FD&C) Act to provide new authorities for the emergency use of approved products in emergencies and products held for emergency use. PAHPRA accordingly amended the definitions of “Covered Countermeasures” and “qualified pandemic and epidemic products” in Section 319F-3 of the Public Health Service Act (PREP Act provisions), so that products made available under these new FD&C Act authorities could be covered under PREP Act Declarations. PAHPRA also extended the definition of qualified pandemic and epidemic products that may be covered under a PREP Act Declaration to include products or technologies intended to enhance the use or effect of a drug, biological product, or device used against the pandemic or epidemic or against adverse events from these products.[2]

Section IV of this Declaration covers Limited Immunity as follows:The Secretary must also state that liability protections available under the PREP Act are in effect with respect to the Recommended Activities. These liability protections provide that, “[s]ubject to other provisions of [the PREP Act], a covered person shall be immune from suit and liability under federal and state law with respect to all claims for loss caused by, arising out of, relating to, or resulting from the administration to or use by an individual of a covered countermeasure if a Declaration has been issued with respect to such countermeasure.” In Section IV of the Declaration, the Secretary states that liability protections are in effect with respect to the Recommended Activities.[2]

Section V of the Declaration describes Covered Persons, including Qualified Persons. The PREP Act defines Covered Persons to include, among others, the United States, and those that manufacturer, distribute, administer, prescribe or use Covered

Countermeasures. This Declaration includes all persons and entities defined as Covered Persons under the PREP Act (PHS Act 317F-3(i)(2)) as well as others set out in paragraphs (3), (4), (6), (8)(A) and (8)(B).[2]

The PREP Act's liability immunity applies to “Covered Persons” with respect to administration or use of a Covered Countermeasure. The term “Covered Persons” has a specific meaning and is defined in the PREP Act to include manufacturers, distributors, program planners, and qualified persons, and their officials, agents, and employees, and the United States.[2]

Section XIV describes the Countermeasures Injury Compensation Program. Section 319F-4 of the PHS Act, 42 U.S.C. 247d-6e, authorizes the Countermeasures Injury Compensation Program (CICP) to provide benefits to eligible individuals who sustain a serious physical injury or die as a direct result of the administration or use of a Covered Countermeasure. Compensation under the CICP for an injury directly caused by a Covered Countermeasure is based on the requirements set forth in this Declaration, the administrative rules for the Program, and the statute. To show direct causation between a Covered Countermeasure and a serious physical injury, the statute requires “compelling, reliable, valid, medical and scientific evidence.” The administrative rules for the Program further explain the necessary requirements for eligibility under the CICP. Please note that, by statute, requirements for compensation under the CICP may not align with the requirements for liability immunity provided under the PREP Act. Section XIV of the Declaration, “Countermeasures Injury Compensation Program,” explains the types of injury and standard of evidence needed to be considered for compensation under the CICP.

The PREP Act authorizes the Countermeasures Injury Compensation Program (CICP) to provide benefits to certain individuals or estates of individuals who sustain a covered serious physical injury as the direct result of the administration or use of the Covered Countermeasures, and benefits to certain survivors of individuals who die as a direct result of the administration or use of the Covered Countermeasures. The causal connection between the countermeasure and the serious physical injury must be supported by compelling, reliable, valid, medical and scientific evidence in order for the individual to be considered for compensation. The CICP is administered by the Health Resources and Services Administration, within the Department of Health and Human Services.[2]

Extension of Liability Shield

The United Stated Department of Health and Human Services (HHS) extended liability protections through December 31$\mathbb{2}$029 through Amendment 12.[4]

Compensation for COVID-19 Vaccine Injuries

The following example is extracted verbatim from the Summary section of Congressional Research Service report R46982 [3]:

> More than 26$\mathbb{0}$ million Americans, and billions of people worldwide, have received one or more doses of a vaccine to protect against COVID-19. Studies estimate that COVID-19 vaccinations have prevented millions of deaths from the disease.
>
> Most common side effects of COVID-19 vaccines are mild and generally resolve in a few days. In rare instances, COVID-19 vaccines can cause serious adverse events. Individuals who believe they have been injured by a COVID-19 vaccine may seek compensation for those injuries and associated harms or costs. Absent an applicable federal law, individuals allegedly injured by a vaccine might seek redress by filing a state tort law claim against the vaccine manufacturer. However, federal law has two distinct compensation regimes that limit legal liability for vaccine manufacturers and provide potential compensation—without requiring a showing of fault—for individuals harmed by adverse reactions to vaccines.
>
> For injuries and deaths associated with most vaccines recommended by the Centers for Disease Control and Prevention (CDC) for routine administration in the United States, the National Vaccine Injury Compensation Program (VICP) may provide compensation. During public health emergencies declared under the Public Readiness and Emergency Preparedness Act (PREP Act), the Countermeasures Injury Compensation Program (CICP) may provide compensation for injuries and deaths resulting from the administration of “covered countermeasures,” which may include vaccines.

The VICP and CICP regimes are similar in some ways, but the programs serve distinct purposes. Compensation through CICP is generally less comprehensive than compensation through VICP. CICP is a regulatory process administered by the Health Resources and Services Administration (HRSA), a division of the U.S. Department of Health and Human Services (HHS). CICP compensation is available only for death or serious injuries resulting from a covered countermeasure. A claimant must generally file a request form and associated documentation with HRSA within one year of the date that the covered countermeasure was administered. For injuries not listed by the Secretary of HHS on a Countermeasure Injury Table, the claimant must demonstrate that the injury was a direct result of the countermeasure's administration based on compelling medical and scientific evidence. HRSA makes decisions regarding eligibility and compensation; judicial review is not available. CICP compensation is limited to reasonable medical expenses, loss of employment income, and a death benefit when the claimant's death is a direct result of the administration of a covered countermeasure

Under the Secretary of HHS's current PREP Act Declaration for COVID-19, FDA-authorized or -licensed COVID-19 vaccines are covered countermeasures. While a PREP Act declaration is in effect, CICP is the sole remedy available for injuries related to covered countermeasures, so CICP—and not VICP—will apply to injuries resulting from COVID-19 vaccinations while the Declaration remains in effect

When coverage under the PREP Act Declaration for COVID-19 ends, COVID-19 vaccine injuries could be compensated through VICP, contingent on additional regulatory and statutory changes. For a vaccine to be included in VICP, (1) the vaccine must be recommended by the CDC for routine administration to children or pregnant women (which the CDC did in 2023); (2) the vaccine must be made subject by an act of Congress to the excise tax that funds VICP; and (3) the Secretary of HHS must add the vaccine to the Vaccine Injury Table, which lists injuries and conditions associated with vaccines covered by VICP. If the remaining two changes occur, COVID-19 vaccines would be covered by VICP.

To receive compensation through VICP for a vaccine-related injury or death, the injured person or their estate must file a petition with the U.S. Court of Federal Claims, generally within three years of the onset of the first symptom or significant aggravation of the injury, or within two years of death or four years of the first symptom resulting in death. To receive compensation, petitioners must either show that they experienced an injury listed in the Vaccine Injury Table within the time frame specified in the Table, or prove that the vaccine was the "but-for" cause of their injury. Special masters determine eligibility and compensation; their decisions may be appealed to the U.S. Court of Federal Claims and the U.S. Court of Appeals for the Federal Circuit. Successful petitioners may receive medical expenses, lost income, a set death benefit, and reasonable attorneys' fees and costs. Petitioners who are dissatisfied with the compensation they receive, or whose claims are delayed, may opt to pursue civil actions in court, subject to certain limitations on vaccine manufacturer liability.

References

1. Administration for Strategic Preparedness and Response, U.S. Department of Health and Human Services. Public Readiness and emergency Preparedness (PREP) Act. PREP Act. Accessed July 7, 2026.
2. Federal Register, Declaration Under the Public Readiness and Emergency Preparedness Act for Medical Countermeasures Against COVID-19. Published March 17, 2020. COVID-19 PREP Act Declaration. Accessed July 7, 2026.
3. Kevin J. Hickey and Hannah-Alise Rogers. Congressional Research Service. Compensation for COVID-19 Vaccine Injuries. Updated April 16, 2025. Report R46982. Compensation for COVID Injuries. Downloaded January 4, 2026. Note: CRS Reports, as a work of the United States Government, are not subject to copyright protection in the United States. Any CRS Report may be reproduced and distributed in its entirety without permission from CRS. However, as a CRS Report may include copyrighted images or material from a third party, you may need to obtain the permission of the copyright holder if you wish to copy or otherwise use copyrighted material.
4. U.S. Department of Health and Human Services, 12th Amendment to Declaration Under the Public Readiness and Emergency Preparedness Act for Medical Countermeasures Against COVID–19, December 11, 2024. PREP ACT Declaration Amendment. Accessed July 7, 2026.

This page titled 2.15.3: Risk Mitigation and Liability is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

Terms

This lexicon or glossary alphabetically defines technology specific vocabulary.

- **ACIP:** Advisory Committee on Immunization Practices
- **BARDA:** Biomedical Advanced Research and Development Authority
- **BLA:** A Biologics License Application is filed with the FDA
- **CBER:** The Food and Drug Administration's Center for Biologics Evaluation and Research handles vaccines
- **CDC:** Centers for Disease Control and Prevention
- **CEPI:** UK Coalition for Epidemic Preparedness Innovations
- **CFR:** Code of Federal Regulations
- **CGMP:** Current Good Manufacturing Practices
- **Clinical Trial (phase 0, 1, 2, 3, 4):** Clinical trials are prospective biomedical or behavioral research studies on human participants designed to answer specific questions about biomedical or behavioral interventions, including new treatments (such as novel vaccines, drugs, dietary choices, dietary supplements, and medical devices) and known interventions that warrant further study and comparison. Clinical trials generate data on dosage, safety and efficacy.[6]
- **CMC:** Chemistry, Manufacturing and Controls
- **Cold Chain:** A cold chain is a supply chain that uses refrigeration to maintain perishable goods, such as pharmaceuticals, produce or other goods that are temperature-sensitive.
- **COVID-19:** Coronavirus Disease 2019
- **DOD:** Department of Defense
- **Dose:** In clinical pharmacology, *dose* refers to the amount of drug administered to a person, and *dosage* is a fuller description that includes not only the dose (e.g., "500 mg") but also the frequency and duration of the treatment (e.g., "twice a day for one week").[4]
- **Efficacy:** The word *efficacy* is used in pharmacology and medicine to refer both to the maximum response achievable from a pharmaceutical drug in research settings, and to the capacity for sufficient therapeutic effect or beneficial change in clinical settings.[3]
- **Epidemiology:** Epidemiology is the study and analysis of the distribution (who, when, and where), patterns and determinants of health and disease conditions in a defined population, and application of this knowledge to prevent diseases.[2]
- **EUA:** Emergency Use Authorization
- **FAR:** Federal Acquisition Regulation governing federal government procurement
- **FDA:** Food and Drug Administration
- **GLP:** Good Laboratory Practice
- **HHS:** Department of Health and Human Services
- **IND:** An Investigational New Drug application is filed with the Unites States Food and Drug Administration to support clinical development.
- **mRNA:** Messenger RNA. **Messenger ribonucleic acid** (**mRNA**) is a single-stranded molecule of RNA that corresponds to the genetic sequence of a gene, and is read by a ribosome in the process of synthesizing a protein.[7]
- **MTD:** Maximally Tolerated Dose. This is determined in Clinical Trials Phase 1. It is essentially the dose level before "unacceptable toxicity is experienced."[1]
- **NIH:** National Institutes of Health
- **Murine:** related to rodents
- **OWS:** Operation Warp Speed
- **Pharmacodynamics:** the science dealing with actions of drugs on the body (what the drug does to the body)
- **Pharmacokinetics:** the science of what the body does to the drug
- **PPE:** Personal Protective Equipment includes items for personal protection against a virus, such as N95 masks/respirators and face shields, protective gowns, and gloves.
- **R&D:** Research and Development
- **Therapeutics:** a branch of medical science dealing with the application of remedies to diseases[8]
- **Toxicity:** Toxicity is the degree to which a chemical substance or a particular mixture of substances can damage an organism. [5]
- **TRL:** BARDA's Technology Readiness Level
- **VRBPAC:** FDA's Vaccines and Related Biological Products Advisory Committee
- **VAERS:** Vaccine Adverse Event Reporting System

References

1. Friedman, Lawrence M., Furberg, Curt D, and DeMets, David L. Fundamentals of Clinical Trials. 1998.
2. Wikipedia - Eipidemiology. Accessed January 31, 2026.
3. Wikipedia - Efficacy. Accessed January 31, 2026.
4. Wikipedia - Dose. Accessed January 31, 2026.
5. Toxicity - Wikipedia. Accessed January 31, 2026.
6. Clinical trial - Wikipedia. Accessed January 31, 2026.
7. Messenger RNA - Wikipedia. Accessed January 31, 2026.
8. THERAPEUTICS Definition & Meaning - Merriam-Webster. Accessed February 3, 2026.

This page titled 2.16: Glossary is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

3: EU GDPR COMPLIANCE

Learning Objectives

- Understand what General Data Protection Regulation (GDPR) compliance requires.
- Consider the specifics of an organization's compliance implementation.

This case study covers the European Union (EU) General Data Protection Regulation (GDPR)[1] that enhances data privacy law. It is an example of a compliance project that would be put in place to ensure that an organization complies with published regulations or standards.

From the frame of reference of an American company, this may at first blush appear to be an intellectual curiosity, as data privacy compliance efforts may be focused on things closer to home such as the California Consumer Privacy Act of 2018 (CCPA, amended in 2020). However, the reach of European Union Directives and Regulations can extend well beyond the EU. This was dramatically seen with the EU Restriction of Hazardous Substances (RoHS) Directive that affected American companies that wanted to sell ("place on the market") electronic products into the EU. Similarly, the GDPR extends its reach beyond the EU by including organizations that collect information about, or process data about, people within the EU.

This case study looks at the provisions of the regulation and the specific duties of data controllers (those that decide what will be done with information about people) and data processors (those that actually process that data). An organization that wants to be compliant needs to incorporate these responsibilities into its compliance framework.

Note

It is important to understand that you may be designated as the project manager of a compliance effort like this with little prior subject matter knowledge. You may feel a lot of pressure to attain compliance in a short time because the executive team got a late start in realizing the need for compliance. As with other compliance efforts associated with new directives and regulations, there may be very few knowledgeable consultants available to help you initially because the consultants are learning on the job as well. Any useful tools and expert learnings may only be available to those in the second wave of compliance efforts.

As a result, this case study is presented in a manner to provide the salient aspects of the regulation itself, as that may be all that is available when your project is initiated. This is done deliberately to simulate what it is like on Day 0 of your project efforts.

The text of the regulation is provided with the **European spellings of words** as they appear in the regulation, rather than the American English spellings which do not appear in the regulation. If you see any American English spellings of regulation language, it is because somewhere in the editing process of this textbook someone tried to be "helpful."

This case study is structured as follows:

- Introduction
- The Regulation: This section describes the structure of the Regulation that is divided into a "Recitals" section that provides context and an "Articles" section that itemizes the specific rules. After itemizing each category of Articles and then the particular Articles comprising that category, the next section extracts the specifics of several salient Articles that are key to a compliance implementation. In a compliance project, you need to know with what you are being compliant to understand scope and the requirements.
- Implementing GDPR Compliance: This section discusses what an organization might consider in assessing what it needs to do to attain GDPR compliance based on its current compliance level with other prior data privacy initiatives.
- Project Management Considerations: This section looks at GDPR from the frame of reference of a project manager that has responsibility for running the GDPR compliance project.
- Legal Considerations: This section provides analysis from a legal frame of reference and looks at penalties that some organizations have incurred due to non-compliance.

- Glossary: The Glossary defines salient terms.

> **Note**
>
> The necessary terms are defined when discussing the various Articles (especially Article 4) or in the Glossary at the end of the chapter.

References

1. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 1, corrigendum 2018 O.J. (L127) 2 (GDPR).

The Commission's reuse policy is implemented by the Commission Decision of 12 December 2011 on the reuse of Commission documents.

Unless otherwise indicated (e.g. in individual copyright notices), content owned by the EU on this website is licensed under the Creative Commons Attribution 4.0 International (CC BY 4.0) licence. This means that reuse is allowed, provided appropriate credit is given and changes are indicated.

Chapter thumbnail: GDPR for Europe by Dooffy is made available under the Creative Commons CC0 1.0 Universal Public Domain Dedication.

3.1: Introduction

3.2: The Regulation

3.2.1: Article 1 Subject-matter and objectives

3.2.2: Article 3 Territorial Scope

3.2.3: Article 4 Definitions

3.2.4: Article 5 Principles related to processing of data

3.2.5: Article 6 Lawfulness of processing

3.2.6: Article 7 Conditions for consent

3.2.7: Article 9 Processing of special categories of personal data

3.2.8: Article 13 Information to be provided where personal data are collected from the data subject

3.2.9: Article 15 Right of access by the data subject

3.2.10: Article 16 Right to rectification

3.2.11: Article 17 Right to erasure ('right to be forgotten')

3.2.12: Article 18 Right to restriction of processing

3.2.13: Article 20 Right to data portability

3.2.14: Article 21 Right to object

3.2.15: Article 22 Automated individual decision-making, including profiling

3.2.16: Article 24 Responsibility of the controller

3.2.17: Article 25 Data protection by design and by default

3.2.18: Article 28 Processor

3.2.19: Article 32 Security of processing

3.2.20: Article 33 Notification of a personal data breach to the supervisory authority

3.2.21: Article 34 Communication of a personal data breach to the data subject
3.2.22: Articles 35-36 Data Protection Impact Assessments
3.2.23: Article 37 Designation of the data protection officer
3.2.24: Article 38 Position of the data protection officer
3.2.25: Article 39 Tasks of the data protection officer
3.2.26: Article 42 Certification
3.2.27: Article 83 General conditions for imposing administrative fines

3.3: Implementing GDPR Compliance
3.3.1: Implementation Considerations

3.4: Project Management Considerations
3.5: Legal Considerations
3.6: Glossary

This page titled 3: EU GDPR Compliance is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

3.1: Introduction

Learning Objectives

- Understand which countries are in the EU.
- Understand which countries are in the European Economic Area (EEA).

Scope/Applicability

At a high level, this Regulation applies to companies doing business with people in the EU, including organizations that collect information about, or processes data about, people within the EU as well as the European Economic Area (EEA).

It is important to note that the constituency of the member states of the EU changes. From the time the GDPR case study was first used for CIS 79 in the fall of 2020, the United Kingdom (UK) exited from the EU (Brexit). After exiting the EU, the UK did put in place an equivalent UK GDPR, which has its own, separate compliance rules and regulatory authority, the Information Commisioner's Office (the UK ICO).

The 27 countries in the EU as of the writing of this textbook include (according to EU countries | European Union):

1. Austria
2. Belgium
3. Bulgaria
4. Croatia
5. Cyprus
6. Czechia (Czech Republic)
7. Denmark
8. Estonia
9. Finland
10. France
11. Germany
12. Greece
13. Hungary
14. Ireland
15. Italy
16. Latvia
17. Lithuania
18. Luxembourg
19. Malta
20. Netherlands
21. Poland
22. Portugal
23. Romania
24. Slovakia
25. Slovenia
26. Spain
27. Sweden

Additionally, other countries are trying to join the EU. As of the writing of this textbook, current candidates include: Albania, Bosnia and Herzegovina, Georgia, Moldova, Montenegro, North Macedonia, Serbia, Turkey, and Ukraine. Any implementation of the GDPR would need to track developments in the entry or exit of member states.

What is the European Economic Area (EEA)?

The EEA essentially extends the EU single market and ensures that all members are abiding by EU laws. The EEA includes all of the EU countries mentioned above plus the following three countries:

1. Iceland
2. Liechtenstein
3. Norway

Short History of EU Privacy Initiatives

We will briefly look at the following data protection initiatives:

- EU Data Protection Directive (95/46/EC)
- General Data Protection Regulation (2016/679).

EU Data Protection Directive (DPD)

The EU Data Protection Directive (DPD) was enacted on October 24, 1995 and went into force on December 13, 1995. It was a major initial step in regulating the processing of personal data associated with an identifiable person (the legal term used is "data subject"). The responsibility for ensuring compliance is with the "controller", the entity that determines how the data will be processed. As with the GDPR which later replaced the DPD, if a controller was outside the EU but processed data in the EU, they were subject to the DPD.

Three key principles applied:

1. **Transparency** (the user must be told when the user's data is going to be processed). This assumed that the user provided consent or that the data needed to be processed due to legal obligation or for the public good.
2. **Legitimate Purpose** (the user's data could be processed for specific legitimate purposes).
3. **Proportionality** (deals with concepts such as data retention for only as long as needed to process the data that was collected for a specific purpose).

These principles persisted into later EU data protection law.

GDPR

GDPR was adopted on April 14, 2016 and went into effect (binding, enforceable) on May 25, 2018. As an EU Regulation, the GDPR is consistently applied in EU member countries. By its enactment, it essentially superseded/repealed the Data Protection Directive. It is a Regulation on data protection and privacy in the European Union and European Economic Area. The GDPR is a law, not a “minimum requirement”. An EU regulation has more force than a directive (which is why the GDPR is stronger than its predecessor, the DPD).

The GDPR applies to any entity processing personal information of “data subjects” inside the EU/EEA (regardless of a company’s location or data subject’s citizenship/residence). One of its objectives is to provide users (e.g, Internet users) more input and control over how their data is used. The GDPR lays down binding guidelines upon companies regarding how user data is collected, stored, transferred, and used. The GDPR applies both to “data controllers” (such as organizations that collect information for various purposes) and “processors” of that data. The GDPR's scope includes international companies and organizations with data processing centers in the EU.

EU Data Act

The EU Data Act is a relatively new EU regulation that went into effect September 12, 2025. It requires that users be granted access to their data, and by September 12, 2026, it requires that products in the EU need to provide direct access to a user's data. This will not be discussed any further in this case study.

Note

In the US, there have been similar privacy initiatives such as the 2018 California Consumer Privacy Act (CCPA, as amended by the California Privacy Rights Act of 2020), but these US state information privacy laws are not the subject of this case study. As of this writing, there is no single comprehensive federal privacy law in the United States comparable to the GDPR, requiring organizations to track carefully what is happening in the states in which they operate and do business and the way that legal requirements can vary from state to state.

Without looking at the textbook, name the three member states in the EEA.

Answer available at the end of this section.

Exercise 3.1.1

Which of the following countries is not in the European Union?

1. Slovenia
2. United Kingdom
3. Greece
4. Cyprus

Answer available at the end of this section.

Answers

Norway, Iceland, and Liechtenstein.

Exercise 3.1.1

United Kingdom

This page titled 3.1: Introduction is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

3.2: The Regulation

Overview of the GDPR's structure

The GDPR is structured in two major sections, **Recitals** and **Articles**, as follows:

RECITALS

- **Recitals** provide context for interpreting the articles along with guidance and direction on implementation of the GDPR.
- Recitals are similar to the "whereas" part of a contract.
- There are 173 numbered Recitals representing 31 of the GDPR's 88 pages.

ARTICLES

- **Articles** contain the specific requirements for compliance (the rules).
- There are 99 numbered Articles in the GDPR.

Definition: Recital

In law, a **recital** (from Latin: *recitare*, "to read out") consists of an account or repetition of the details of some act, proceeding or fact. Particularly, in law, that part of a legal document—such as a lease, which contains a statement of certain facts—contains the purpose for which the deed is made.

In European Union law, a recital is a text that sets out reasons for the provisions of an enactment, while avoiding normative language and political argumentation.

Source: Recital (law) - Wikipedia

The GDPR organizes the 99 Articles into eleven chapters, covering such things as the rights of data subjects, the responsibilities of data controllers and data processors, as well as penalties.

We will not be discussing all of the GDPR's 99 Articles in detail. We will concentrate on those Articles that are important to an organization's compliance efforts.

This section only provides an overview of the general organization of the Articles. The next section will provide extracts from the GDPR of the Articles that will be relevant for many compliance implementations.

Note

Not every Article will apply to every organization.

Chapter 1 General Provisions

Chapter 1 provides some high level statements such as people having fundamental rights concerning the protection of their personal data, the regulation's scope in regard to the processing of that personal data, the geographic reach or scope of the GDPR, and lastly a set of definitions of pertinent terms.

- Article 1 Subject-matter and objectives
- Article 2 Material Scope
- Article 3 Territorial Scope

- Article 4 Definitions

Chapter 2 Principles

Chapter 2 covers some key data protection principles such as: lawfulness of processing, collecting and processing data only for stated purposes, keeping the data accurate, and providing and withdrawing consent for processing data. These Articles contain some of the GDPR's most important rules about when it is permissible to process people's personal data.

- Article 5 Principles relating to processing of personal data
- Article 6 Lawfulness of processing
- Article 7 Conditions for consent
- Article 8 Conditions applicable to child's consent in relation to information society services
- Article 9 Processing of special categories of personal data
- Article 10 Processing of personal data relating to criminal convictions and offenses
- Article 11 Processing which does not require identification

Chapter 3 Rights of the data subject

Chapter 3 is a key chapter in that it defines the rights of data subjects regarding their personal data such as: whether their data is being processed and for what reasons, the ability to correct incorrect data, the ability to have their data be erased, the right to move their data to another controller, and the right to not have their data be used for direct marketing purposes and profiling.

Section 1 Transparency and modalities

- Article 12 Transparent information communication and modalities for the exercise of the rights of the data subject.

Section 2 Information and access to personal data

- Article 13 Information to be provided where personal data are collected from the data subject
- Article 14 Information to be provided where personal data have not been obtained from the data subject
- Article 15 Right of access by the data subject.

Section 3 Rectification and erasure

- Article 16 Right to rectification
- Article 17 Right to erasure (‘right to be forgotten’)
- Article 18 Right to restriction of processing
- Article 19 Notification obligation regarding rectification or erasure of personal data or restriction of processing
- Article 20 Right to data portability.

Section 4 Right to object and automated individual decision-making

- Article 21 Right to object
- Article 22 Automated individual decision-making, including profiling

Section 5 Restrictions

- Article 23 Restrictions

Chapter 4 Controller and processor

Chapter 4 defines the responsibilities of the controller and processor entities. It limits the use of personal data to only that required for each specific purpose. It requires keeping records of processing activities.

Additionally, Chapter 4 addresses data protection impact assessments (DPIAs), the security of processing data (the pseudonymisation and encryption of personal data), the hiring of a Data Protection Officer (DPO), and certifications of compliance.

Section 1 General obligations

- Article 24 Responsibility of the controller
- Article 25 Data protection by design and by default
- Article 26 Joint controllers
- Article 27 Representatives of controllers or processors not established in the Union
- Article 28 Processor
- Article 29 Processing under the authority of the controller or processor
- Article 30 Records of processing activities
- Article 31 Cooperation with the supervisory authority

Section 2 Security of personal data

- Article 32 Security of processing
- Article 33 Notification of a personal data breach to the supervisory authority
- Article 34 Communication of a personal data breach to the data subject

Section 3 Data protection impact assessment and prior consultation

- Article 35 Data protection impact assessment
- Article 36 Prior consultation

Section 4 Data protection officer

- Article 37 Designation of the data protection officer
- Article 38 Position of the data protection officer
- Article 39 Tasks of the data protection officer

Section 5 Codes of conduct and certification

- Article 40 Codes of conduct
- Article 41 Monitoring of approved codes of conduct
- Article 42 Certification
- Article 43 Certification bodies

Chapter 5 Transfers of personal data to third countries or international organisations

Chapter 5 establishes principles and rules for handling personal data being transferred to third countries or to an international organization.

- Article 44 General principle for transfers
- Article 45 Transfers on the basis of an adequacy decision
- Article 46 Transfers subject to appropriate safeguards
- Article 47 Binding corporate rules
- Article 48 Transfers or disclosures not authorised by Union law
- Article 49 Derogations for specific situations
- Article 50 International cooperation for the protection of personal data

Chapter 6 Independent supervisory authorities

Chapter 6 defines what the supervisory authority is and how it is established in a member state (country). It then describes the supervisory authority's responsibilities that include monitoring and enforcing compliance to the GDPR in the member state. The supervisory authority's powers are in three main areas:

1. Investigative: this powers enables gathering evidence.
2. Corrective: it can issue warnings and fines to controllers and processors.
3. Authorization and advisory: it can establish standards and ways to obtain certifications of compliance.

Section 1 Independent status

- Article 51 Supervisory authority
- Article 52 Independence
- Article 53 General conditions for the members of the supervisory authority
- Article 54 Rules on the establishment of the supervisory authority

Section 2 Competence, tasks and powers

- Article 55 Competence
- Article 56 Competence of the lead supervisory authority
- Article 57 Tasks
- Article 58 Powers
- Article 59 Activity reports

Note

You will often find the supervisory authority referred to by privacy professionals and consultants as the **data protection authority**.

Chapter 7 Cooperation and consistency

When it comes to GDPR enforcement, it might be helpful to think of it occurring through a hierarchy of authorities: the supervisory authority for each member state (country) described earlier, followed by the European Data Protection Board (EDPB) that is comprised of representatives from each member state's supervisory authority. The EDPB is described in this section. Since the EDPB has representation from all member states, it can strive to apply the GDPR consistently across all the member states. The highest authority in the hierarchy is the EU Commission.

Section 1 Cooperation

- Article 60 Cooperation between the lead supervisory authority and the other supervisory authorities concerned
- Article 61 Mutual assistance
- Article 62 Joint operations of supervisory authorities

Section 2 Consistency

- Article 63 Consistency mechanism
- Article 64 Opinion of the Board
- Article 65 Dispute resolution by the Board
- Article 66 Urgency procedure
- Article 67 Exchange of information

Section 3 European data protection board

- Article 68 European Data Protection Board
- Article 69 Independence
- Article 70 Tasks of the Board
- Article 71 Reports
- Article 72 Procedure
- Article 73 Chair
- Article 74 Tasks of the Chair
- Article 75 Secretariat
- Article 76 Confidentiality

Chapter 8 Remedies, liability and penalties

Chapter 8 defines the rights of data subjects to lodge a complaint with a supervisory authority if they think the processing of their personal data by a controller or processor has been counter to the GDPR and to seek judicial remedies (e.g., potential compensation, or punishments against violators). Article 83 usually gets the most press as it defines the fines and penalties that can be imposed for non-compliance.

- Article 77 Right to lodge a complaint with a supervisory authority
- Article 78 Right to an effective judicial remedy against a supervisory authority
- Article 79 Right to an effective judicial remedy against a controller or processor
- Article 80 Representation of data subjects
- Article 81 Suspension of proceedings
- Article 82 Right to compensation and liability
- Article 83 General conditions for imposing administrative fines
- Article 84 Penalties

> **Note**
>
> The details of Article 83 are covered in the next section of the textbook.

Chapter 9 Provisions relating to specific processing situations

We will not be addressing the special cases called out in Chapter 9.

- Article 85 Processing and freedom of expression and information
- Article 86 Processing and public access to official documents
- Article 87 Processing of the national identification number
- Article 88 Processing in the context of employment
- Article 89 Safeguards and derogations relating to processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes
- Article 90 Obligations of secrecy
- Article 91 Existing data protection rules of churches and religious associations

Chapter 10 Delegated acts and implementing acts

We will not be addressing the administrative items called out in Chapter 10.

- Article 92 Exercise of the delegation
- Article 93 Committee procedure

Chapter 11 Final provisions

Article 94 repeals GDPR's predecessor, the DPD, as of May 25, 2018, and Article 99 states that the GDPR goes into force as of May 25, 2018.

- Article 94 Repeal of Directive 95/46/EC
- Article 95 Relationship with Directive 2002/58/EC
- Article 96 Relationship with previously concluded Agreements
- Article 97 Commission reports
- Article 98 Review of other Union legal acts on data protection
- Article 99 Entry into force and application

What follows is the text of several of the GDPR's Articles that are especially relevant to implementing a compliance program consistent with the GDPR's legal requirements.

Note

This chapter should not be treated as legal advice. For that, consult a practicing attorney.

References

1. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 1.

The Commission's reuse policy is implemented by the Commission Decision of 12 December 2011 on the reuse of Commission documents.

Unless otherwise indicated (e.g. in individual copyright notices), content owned by the EU on this website is licensed under the Creative Commons Attribution 4.0 International (CC BY 4.0) licence. This means that reuse is allowed, provided appropriate credit is given and changes are indicated.

This page titled 3.2: The Regulation is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

3.2.1: Article 1 Subject-matter and objectives

1. This Regulation lays down rules relating to the protection of natural persons with regard to the processing of personal data and rules relating to the free movement of personal data.
2. This Regulation protects fundamental rights and freedoms of natural persons and in particular their right to the protection of personal data.
3. The free movement of personal data within the Union shall be neither restricted nor prohibited for reasons connected with the protection of natural persons with regard to the processing of personal data.

Note

The language of "free movement of personal data" is reflective of the EU's origins as an economic union.

The GDPR is establishing a standard data protection law across all of the EU member states.

Note

In law, there is often a technical distinction between "natural persons" (that is, human beings like you and me) and "artificial persons" (such as corporations). The GDPR is interested in protecting the data of human individuals.

References

1. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 1.

This page titled 3.2.1: Article 1 Subject-matter and objectives is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

3.2.2: Article 3 Territorial Scope

1. This Regulation applies to the processing of personal data in the context of the activities of an establishment of a controller or a processor in the Union, regardless of whether the processing takes place in the Union or not.
2. This Regulation applies to the processing of personal data of data subjects who are in the Union by a controller or processor not established in the Union, where the processing activities are related to:
   a. the offering of goods or services, irrespective of whether a payment of the data subject is required, to such data subjects in the Union; or
   b. the monitoring of their behaviour as far as their behaviour takes place within the Union.
3. This Regulation applies to the processing of personal data by a controller not established in the Union, but in a place where Member State law applies by virtue of public international law.

Note

This territorial scope article defines how companies in the United States might be affected by the GDPR by having controller or processing activities inside or outside of the EU acting on the data of people in the EU.

References

1. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 1.

This page titled 3.2.2: Article 3 Territorial Scope is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

3.2.3: Article 4 Definitions

A subset of Article 4's definitions are extracted to aid in understanding the implementation of GDPR. For example, those that are purely administrative are not included.

Definition: Personal Data

‘personal data’ means any information relating to an identified or identifiable natural person (‘data subject’); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person;

Definition: Processing

‘processing’ means any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction;

Definition: Profiling

'profiling’ means any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements;

Definition: Pseudonymisation

‘pseudonymisation’ means the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person;

Definition: Conroller

‘controller’ means the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data; where the purposes and means of such processing are determined by Union or Member State law,

Definition: Processor

‘processor’ means a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller;

Definition: Third party

'third party' means a natural or legal person, public authority, agency or body other than the data subject, controller, processor and persons who, under the direct authority of the controller or processor, are authorised to process personal data;

Definition: Consent

'consent' of the data subject means any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her;

Definition: Personal data breach

'personal data breach' means a breach of security leading to the accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to, personal data transmitted, stored or otherwise processed;

Definition: Supervisory authority

'supervisory authority' means an independent public authority which is established by a Member State pursuant to Article 51;

Definition: International organisation

'international organisation' means an organisation and its subordinate bodies governed by public international law, or any other body which is set up by, or on the basis of, an agreement between two or more countries.

References

1. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 1.

This page titled 3.2.3: Article 4 Definitions is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

3.2.4: Article 5 Principles related to processing of data

Note

Note: The numbering and lettering are as they appear in the GDPR.

1. Personal data shall be:
   a. processed lawfully, fairly and in a transparent manner in relation to the data subject ('lawfulness, fairness and transparency');
   b. collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes; further processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes shall, in accordance with Article 89(1), not be considered to be incompatible with the initial purposes ('purpose limitation');
   c. adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed ('data minimisation');
   d. accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay ('accuracy');
   e. kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed; personal data may be stored for longer periods insofar as the personal data will be processed solely for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) subject to implementation of the appropriate technical and organisational measures required by this Regulation in order to safeguard the rights and freedoms of the data subject ('storage limitation');
   f. processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures ('integrity and confidentiality').

Exercise 3.2.4.1

What are the six major data processing principles described here?

Answer available at the end of this section.

References

1. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 1.

Answers

Exercise 3.2.4.1

1. Lawfulness, fairness and transparency
2. Purpose limitation
3. Data minimisation (minimization)
4. Accuracy
5. Storage limitation
6. Integrity and confidentiality

This page titled 3.2.4: Article 5 Principles related to processing of data is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

3.2.5: Article 6 Lawfulness of processing

1. Processing shall be lawful only if and to the extent that at least one of the following applies:

   (a) the data subject has given consent to the processing of his or her personal data for one or more specific purposes;

   (b) processing is necessary for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract;

   (c) processing is necessary for compliance with a legal obligation to which the controller is subject;

   (d) processing is necessary in order to protect the vital interests of the data subject or of another natural person;

   (e) processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller;

   (f) processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child.

Point (f) of the first subparagraph shall not apply to processing carried out by public authorities in the performance of their tasks.

Note

Lawfulness of processing can be achieved either via

- Consent of the data subject (point (a) above, and see Article 7)

or

- Necessity of the processing (the 5 purposes in points (b) through (f) above)

References

1. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 1.

This page titled 3.2.5: Article 6 Lawfulness of processing is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

3.2.6: Article 7 Conditions for consent

1. Where processing is based on consent, the controller shall be able to demonstrate that the data subject has consented to processing of his or her personal data.
2. If the data subject's consent is given in the context of a written declaration which also concerns other matters, the request for consent shall be presented in a manner which is clearly distinguishable from the other matters, in an intelligible and easily accessible form, using clear and plain language. Any part of such a declaration which constitutes an infringement of this Regulation shall not be binding.
3. The data subject shall have the right to withdraw his or her consent at any time. The withdrawal of consent shall not affect the lawfulness of processing based on consent before its withdrawal. Prior to giving consent, the data subject shall be informed thereof. It shall be as easy to withdraw as to give consent.
4. When assessing whether consent is freely given, utmost account shall be taken of whether, inter alia, the performance of a contract, including the provision of a service, is conditional on consent to the processing of personal data that is not necessary for the performance of that contract.

Definition: inter alia

Inter alia is from the Latin, meaning "among other things."

References

1. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 1.

This page titled 3.2.6: Article 7 Conditions for consent is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

3.2.7: Article 9 Processing of special categories of personal data

1. Processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation shall be prohibited.

Note

There are some exceptions that are itemized in item 2 (not included here).

Exceptions allow for processing (Art. 9(2)(a-g)).

- Some of these exceptions follow the consent/necessity pattern noted in Article 6.
- Other exceptions include non-profits and if the data subject manifestly made it public.

References

1. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 1.

This page titled 3.2.7: Article 9 Processing of special categories of personal data is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

3.2.8: Article 13 Information to be provided where personal data are collected from the data subject

Note

Data Subjects' Rights are covered in these Articles:

- Article 12: controllers furnishing information in clear, transparent ways
- Article 13: controller duties of disclosure to data subjects when collecting or intending to process data about them (akin to implementing the Fair Information Practice Principles, or FIPPs)
- Article 15-17: right of access to, and rectification or deletion of, data
- Article 20: portability of data between controllers
- Article 21: right to object to processing based on public interest or legitimate controller or 3rd party interest (Art. 6(1)(e-f))

1. Where personal data relating to a data subject are collected from the data subject, the controller shall, at the time when personal data are obtained, provide the data subject with all of the following information:
   a. the identity and the contact details of the controller and, where applicable, of the controller's representative;
   b. the contact details of the data protection officer, where applicable;
   c. the purposes of the processing for which the personal data are intended as well as the legal basis for the processing;
   d. where the processing is based on point (f) of Article 6(1), the legitimate interests pursued by the controller or by a third party;
   e. the recipients or categories of recipients of the personal data, if any;
   f. where applicable, the fact that the controller intends to transfer personal data to a third country or international organisation and the existence or absence of an adequacy decision by the Commission, or in the case of transfers referred to in Article 46 or 47, or the second subparagraph of Article 49(1), reference to the appropriate or suitable safeguards and the means by which to obtain a copy of them or where they have been made available.
2. In addition to the information referred to in paragraph 1, the controller shall, at the time when personal data are obtained, provide the data subject with the following further information necessary to ensure fair and transparent processing:
   a. the period for which the personal data will be stored, or if that is not possible, the criteria used to determine that period;
   b. the existence of the right to request from the controller access to and rectification or erasure of personal data or restriction of processing concerning the data subject or to object to processing as well as the right to data portability;
   c. where the processing is based on point (a) of Article 6(1) or point (a) of Article 9(2), the existence of the right to withdraw consent at any time, without affecting the lawfulness of processing based on consent before its withdrawal;
   d. the right to lodge a complaint with a supervisory authority;
   e. whether the provision of personal data is a statutory or contractual requirement, or a requirement necessary to enter into a contract, as well as whether the data subject is obliged to provide the personal data and of the possible consequences of failure to provide such data;
   f. the existence of automated decision-making, including profiling, referred to in Article 22(1) and (4) and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject.
3. Where the controller intends to further process the personal data for a purpose other than that for which the personal data were collected, the controller shall provide the data subject prior to that further processing with information on that other purpose and with any relevant further information as referred to in paragraph 2.
4. Paragraphs 1, 2 and 3 shall not apply where and insofar as the data subject already has the information.

Note

Note item 2 (f) that discusses automated decision making. This will be discussed further later.

References

1. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 1.

This page titled 3.2.8: Article 13 Information to be provided where personal data are collected from the data subject is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

3.2.9: Article 15 Right of access by the data subject

1. The data subject shall have the right to obtain from the controller confirmation as to whether or not personal data concerning him or her are being processed, and, where that is the case, access to the personal data and the following information:
   a. the purposes of the processing;
   b. the categories of personal data concerned;
   c. the recipients or categories of recipient to whom the personal data have been or will be disclosed, in particular recipients in third countries or international organisations;
   d. where possible, the envisaged period for which the personal data will be stored, or, if not possible, the criteria used to determine that period;
   e. the existence of the right to request from the controller rectification or erasure of personal data or restriction of processing of personal data concerning the data subject or to object to such processing;
   f. the right to lodge a complaint with a supervisory authority;
   g. where the personal data are not collected from the data subject, any available information as to their source;
   h. the existence of automated decision-making, including profiling, referred to in Article 22(1) and (4) and, at least in those cases, meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject.
2. Where personal data are transferred to a third country or to an international organisation, the data subject shall have the right to be informed of the appropriate safeguards pursuant to Article 46 relating to the transfer.
3. The controller shall provide a copy of the personal data undergoing processing. For any further copies requested by the data subject, the controller may charge a reasonable fee based on administrative costs. Where the data subject makes the request by electronic means, and unless otherwise requested by the data subject, the information shall be provided in a commonly used electronic form.
4. The right to obtain a copy referred to in paragraph 3 shall not adversely affect the rights and freedoms of others.

References

1. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 1.

This page titled 3.2.9: Article 15 Right of access by the data subject is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

3.2.10: Article 16 Right to rectification

The data subject shall have the right to obtain from the controller without undue delay the rectification of inaccurate personal data concerning him or her. Taking into account the purposes of the processing, the data subject shall have the right to have incomplete personal data completed, including by means of providing a supplementary statement.

Note

In the GDPR, this is a single unnumbered paragraph.

References

1. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 1.

This page titled 3.2.10: Article 16 Right to rectification is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

3.2.11: Article 17 Right to erasure (‘right to be forgotten’)

1. The data subject shall have the right to obtain from the controller the erasure of personal data concerning him or her without undue delay and the controller shall have the obligation to erase personal data without undue delay where one of the following grounds applies:
   a. the personal data are no longer necessary in relation to the purposes for which they were collected or otherwise processed;
   b. the data subject withdraws consent on which the processing is based according to point (a) of Article 6(1), or point (a) of Article 9(2), and where there is no other legal ground for the processing;
   c. the data subject objects to the processing pursuant to Article 21(1) and there are no overriding legitimate grounds for the processing, or the data subject objects to the processing pursuant to Article 21(2);
   d. the personal data have been unlawfully processed;
   e. the personal data have to be erased for compliance with a legal obligation in Union or Member State law to which the controller is subject;
   f. the personal data have been collected in relation to the offer of information society services referred to in Article 8(1).
2. Where the controller has made the personal data public and is obliged pursuant to paragraph 1 to erase the personal data, the controller, taking account of available technology and the cost of implementation, shall take reasonable steps, including technical measures, to inform controllers which are processing the personal data that the data subject has requested the erasure by such controllers of any links to, or copy or replication of, those personal data.
3. Paragraphs 1 and 2 shall not apply to the extent that processing is necessary:
   a. for exercising the right of freedom of expression and information;
   b. for compliance with a legal obligation which requires processing by Union or Member State law to which the controller is subject or for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller;
   c. for reasons of public interest in the area of public health in accordance with points (h) and (i) of Article 9(2) as well as Article 9(3);
   d. for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) in so far as the right referred to in paragraph 1 is likely to render impossible or seriously impair the achievement of the objectives of that processing; or
   e. for the establishment, exercise or defence of legal claims.

Note

The right to be forgotten is new compared with the Data Protection Directive (DPD). Organizations that were compliant with the DPD need to address this to attain GDPR compliance.

References

1. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 1.

This page titled 3.2.11: Article 17 Right to erasure (‘right to be forgotten’) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

3.2.12: Article 18 Right to restriction of processing

1. The data subject shall have the right to obtain from the controller restriction of processing where one of the following applies:
   a. the accuracy of the personal data is contested by the data subject, for a period enabling the controller to verify the accuracy of the personal data;
   b. the processing is unlawful and the data subject opposes the erasure of the personal data and requests the restriction of their use instead;
   c. the controller no longer needs the personal data for the purposes of the processing, but they are required by the data subject for the establishment, exercise or defence of legal claims;
   d. the data subject has objected to processing pursuant to Article 21(1) pending the verification whether the legitimate grounds of the controller override those of the data subject.
2. Where processing has been restricted under paragraph 1, such personal data shall, with the exception of storage, only be processed with the data subject's consent or for the establishment, exercise or defence of legal claims or for the protection of the rights of another natural or legal person or for reasons of important public interest of the Union or of a Member State.
3. A data subject who has obtained restriction of processing pursuant to paragraph 1 shall be informed by the controller before the restriction of processing is lifted.

References

1. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 1.

This page titled 3.2.12: Article 18 Right to restriction of processing is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

3.2.13: Article 20 Right to data portability

1. The data subject shall have the right to receive the personal data concerning him or her, which he or she has provided to a controller, in a structured, commonly used and machine-readable format and have the right to transmit those data to another controller without hindrance from the controller to which the personal data have been provided, where:
   a. the processing is based on consent pursuant to point (a) of Article 6(1) or point (a) of Article 9(2) or on a contract pursuant to point (b) of Article 6(1); and
   b. the processing is carried out by automated means.
2. In exercising his or her right to data portability pursuant to paragraph 1, the data subject shall have the right to have the personal data transmitted directly from one controller to another, where technically feasible.
3. The exercise of the right referred to in paragraph 1 of this Article shall be without prejudice to Article 17. That right shall not apply to processing necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller.
4. The right referred to in paragraph 1 shall not adversely affect the rights and freedoms of others.

Note

The right to data portability is new compared with the Data Protection Directive (DPD). Organizations that were compliant with the DPD need to address this to attain GDPR compliance.

References

1. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 1.

This page titled 3.2.13: Article 20 Right to data portability is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

3.2.14: Article 21 Right to object

1. The data subject shall have the right to object, on grounds relating to his or her particular situation, at any time to processing of personal data concerning him or her which is based on point (e) or (f) of Article 6(1), including profiling based on those provisions. The controller shall no longer process the personal data unless the controller demonstrates compelling legitimate grounds for the processing which override the interests, rights and freedoms of the data subject or for the establishment, exercise or defence of legal claims.
2. Where personal data are processed for direct marketing purposes, the data subject shall have the right to object at any time to processing of personal data concerning him or her for such marketing, which includes profiling to the extent that it is related to such direct marketing.
3. Where the data subject objects to processing for direct marketing purposes, the personal data shall no longer be processed for such purposes.
4. At the latest at the time of the first communication with the data subject, the right referred to in paragraphs 1 and 2 shall be explicitly brought to the attention of the data subject and shall be presented clearly and separately from any other information.
5. In the context of the use of information society services, and notwithstanding Directive 2002/58/EC, the data subject may exercise his or her right to object by automated means using technical specifications.
6. Where personal data are processed for scientific or historical research purposes or statistical purposes pursuant to Article 89(1), the data subject, on grounds relating to his or her particular situation, shall have the right to object to processing of personal data concerning him or her, unless the processing is necessary for the performance of a task carried out for reasons of public interest.

Note

Note especially items 3 and 4 regarding personal data used for direct marketing purposes.

References

1. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 1.

This page titled 3.2.14: Article 21 Right to object is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

3.2.15: Article 22 Automated individual decision-making, including profiling

1. The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.
2. Paragraph 1 shall not apply if the decision:
   a. is necessary for entering into, or performance of, a contract between the data subject and a data controller;
   b. is authorised by Union or Member State law to which the controller is subject and which also lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests; or
   c. is based on the data subject's explicit consent.
3. In the cases referred to in points (a) and (c) of paragraph 2, the data controller shall implement suitable measures to safeguard the data subject's rights and freedoms and legitimate interests, at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision.
4. Decisions referred to in paragraph 2 shall not be based on special categories of personal data referred to in Article 9(1), unless point (a) or (g) of Article 9(2) applies and suitable measures to safeguard the data subject's rights and freedoms and legitimate interests are in place.

Note

The data subject has the right NOT to be subject to a decision based solely on automated processing and **has the right to obtain human intervention**.

References

1. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 1.

This page titled 3.2.15: Article 22 Automated individual decision-making, including profiling is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

3.2.16: Article 24 Responsibility of the controller

1. Taking into account the nature, scope, context and purposes of processing as well as the risks of varying likelihood and severity for the rights and freedoms of natural persons, the controller shall implement appropriate technical and organisational measures to ensure and to be able to demonstrate that processing is performed in accordance with this Regulation. Those measures shall be reviewed and updated where necessary.
2. Where proportionate in relation to processing activities, the measures referred to in paragraph 1 shall include the implementation of appropriate data protection policies by the controller.
3. Adherence to approved codes of conduct as referred to in Article 40 or approved certification mechanisms as referred to in Article 42 may be used as an element by which to demonstrate compliance with the obligations of the controller.

Note

In a subsequent section, we will look further into implementation specifics and the creation of organizational procedures.

References

1. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 1.

This page titled 3.2.16: Article 24 Responsibility of the controller is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

3.2.17: Article 25 Data protection by design and by default

1. Taking into account the state of the art, the cost of implementation and the nature, scope, context and purposes of processing as well as the risks of varying likelihood and severity for rights and freedoms of natural persons posed by the processing, the controller shall, both at the time of the determination of the means for processing and at the time of the processing itself, implement appropriate technical and organisational measures, such as pseudonymisation, which are designed to implement data-protection principles, such as data minimisation, in an effective manner and to integrate the necessary safeguards into the processing in order to meet the requirements of this Regulation and protect the rights of data subjects.
2. The controller shall implement appropriate technical and organisational measures for ensuring that, by default, only personal data which are necessary for each specific purpose of the processing are processed. That obligation applies to the amount of personal data collected, the extent of their processing, the period of their storage and their accessibility. In particular, such measures shall ensure that by default personal data are not made accessible without the individual's intervention to an indefinite number of natural persons.
3. An approved certification mechanism pursuant to Article 42 may be used as an element to demonstrate compliance with the requirements set out in paragraphs 1 and 2 of this Article.

Note

This Article as well as several of the Recitals, actively push pseudonymisation/pseudonymization as an implementation option.

References

1. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 1.

This page titled 3.2.17: Article 25 Data protection by design and by default is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

3.2.18: Article 28 Processor

1. Where processing is to be carried out on behalf of a controller, the controller shall use only processors providing sufficient guarantees to implement appropriate technical and organisational measures in such a manner that processing will meet the requirements of this Regulation and ensure the protection of the rights of the data subject.
2. The processor shall not engage another processor without prior specific or general written authorisation of the controller. In the case of general written authorisation, the processor shall inform the controller of any intended changes concerning the addition or replacement of other processors, thereby giving the controller the opportunity to object to such changes.

Note

For item (2), having the controller be aware of which processors are processing which data will be useful in implementing some Articles, such as the Right to be Forgotten.

3. Processing by a processor shall be governed by a contract or other legal act under Union or Member State law, that is binding on the processor with regard to the controller and that sets out the subject-matter and duration of the processing, the nature and purpose of the processing, the type of personal data and categories of data subjects and the obligations and rights of the controller. That contract or other legal act shall stipulate, in particular, that the processor:

   With regard to point (h) of the first subparagraph, the processor shall immediately inform the controller if, in its opinion, an instruction infringes this Regulation or other Union or Member State data protection provisions.

   a. processes the personal data only on documented instructions from the controller, including with regard to transfers of personal data to a third country or an international organisation, unless required to do so by Union or Member State law to which the processor is subject; in such a case, the processor shall inform the controller of that legal requirement before processing, unless that law prohibits such information on important grounds of public interest;
   b. ensures that persons authorised to process the personal data have committed themselves to confidentiality or are under an appropriate statutory obligation of confidentiality;
   c. takes all measures required pursuant to Article 32;
   d. respects the conditions referred to in paragraphs 2 and 4 for engaging another processor;
   e. taking into account the nature of the processing, assists the controller by appropriate technical and organisational measures, insofar as this is possible, for the fulfilment of the controller's obligation to respond to requests for exercising the data subject's rights laid down in Chapter III;
   f. assists the controller in ensuring compliance with the obligations pursuant to Articles 32 to 36 taking into account the nature of processing and the information available to the processor;
   g. at the choice of the controller, deletes or returns all the personal data to the controller after the end of the provision of services relating to processing, and deletes existing copies unless Union or Member State law requires storage of the personal data;
   h. makes available to the controller all information necessary to demonstrate compliance with the obligations laid down in this Article and allow for and contribute to audits, including inspections, conducted by the controller or another auditor mandated by the controller.
4. Where a processor engages another processor for carrying out specific processing activities on behalf of the controller, the same data protection obligations as set out in the contract or other legal act between the controller and the processor as referred to in paragraph 3 shall be imposed on that other processor by way of a contract or other legal act under Union or Member State law, in particular providing sufficient guarantees to implement appropriate technical and organisational measures in such a manner that the processing will meet the requirements of this Regulation. Where that other processor fails to fulfil its data protection obligations, the initial processor shall remain fully liable to the controller for the performance of that other processor's obligations.
5. Adherence of a processor to an approved code of conduct as referred to in Article 40 or an approved certification mechanism as referred to in Article 42 may be used as an element by which to demonstrate sufficient guarantees as referred to in paragraphs 1 and 4 of this Article.
6. Without prejudice to an individual contract between the controller and the processor, the contract or the other legal act referred to in paragraphs 3 and 4 of this Article may be based, in whole or in part, on standard contractual clauses referred to in paragraphs 7 and 8 of this Article, including when they are part of a certification granted to the controller or processor

pursuant to Articles 42 and 43.
7. The Commission may lay down standard contractual clauses for the matters referred to in paragraph 3 and 4 of this Article and in accordance with the examination procedure referred to in Article 93(2).
8. A supervisory authority may adopt standard contractual clauses for the matters referred to in paragraph 3 and 4 of this Article and in accordance with the consistency mechanism referred to in Article 63.
9. The contract or the other legal act referred to in paragraphs 3 and 4 shall be in writing, including in electronic form.
10. Without prejudice to Articles 82, 83 and 84, if a processor infringes this Regulation by determining the purposes and means of processing, the processor shall be considered to be a controller in respect of that processing.

Note

The spelling of "fulfil" is the British English spelling of "fulfill." It is not a typo since this is taken from an EU document.

References

1. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 1.

This page titled 3.2.18: Article 28 Processor is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

3.2.19: Article 32 Security of processing

Note

This Article covers pseudonymisation, data encryption, availability, and backups.

1. Taking into account the state of the art, the costs of implementation and the nature, scope, context and purposes of processing as well as the risk of varying likelihood and severity for the rights and freedoms of natural persons, the controller and the processor shall implement appropriate technical and organisational measures to ensure a level of security appropriate to the risk, including inter alia as appropriate:
   a. the pseudonymisation and encryption of personal data;
   b. the ability to ensure the ongoing confidentiality, integrity, availability and resilience of processing systems and services;
   c. the ability to restore the availability and access to personal data in a timely manner in the event of a physical or technical incident;
   d. a process for regularly testing, assessing and evaluating the effectiveness of technical and organisational measures for ensuring the security of the processing.
2. In assessing the appropriate level of security account shall be taken in particular of the risks that are presented by processing, in particular from accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to personal data transmitted, stored or otherwise processed.
3. Adherence to an approved code of conduct as referred to in Article 40 or an approved certification mechanism as referred to in Article 42 may be used as an element by which to demonstrate compliance with the requirements set out in paragraph 1 of this Article.
4. The controller and processor shall take steps to ensure that any natural person acting under the authority of the controller or the processor who has access to personal data does not process them except on instructions from the controller, unless he or she is required to do so by Union or Member State law.

References

1. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 1.

This page titled 3.2.19: Article 32 Security of processing is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

3.2.20: Article 33 Notification of a personal data breach to the supervisory authority

1. In the case of a personal data breach, the controller shall without undue delay and, where feasible, not later than 72 hours after having become aware of it, notify the personal data breach to the supervisory authority competent in accordance with Article 55, unless the personal data breach is unlikely to result in a risk to the rights and freedoms of natural persons. Where the notification to the supervisory authority is not made within 72 hours, it shall be accompanied by reasons for the delay.
2. The processor shall notify the controller without undue delay after becoming aware of a personal data breach.
3. The notification referred to in paragraph 1 shall at least: (a) describe the nature of the personal data breach including where possible, the categories and approximate number of data subjects concerned and the categories and approximate number of personal data records concerned; (b) communicate the name and contact details of the data protection officer or other contact point where more information can be obtained; (c) describe the likely consequences of the personal data breach; (d) describe the measures taken or proposed to be taken by the controller to address the personal data breach, including, where appropriate, measures to mitigate its possible adverse effects.
4. Where, and in so far as, it is not possible to provide the information at the same time, the information may be provided in phases without undue further delay.
5. The controller shall document any personal data breaches, comprising the facts relating to the personal data breach, its effects and the remedial action taken. That documentation shall enable the supervisory authority to verify compliance with this Article.

Note

In addition to notifying the supervisory authority of a data breach, the next Article, Article 34, covers notifying the data subject.

References

1. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 1.

This page titled 3.2.20: Article 33 Notification of a personal data breach to the supervisory authority is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

3.2.21: Article 34 Communication of a personal data breach to the data subject

1. When the personal data breach is likely to result in a high risk to the rights and freedoms of natural persons, the controller shall communicate the personal data breach to the data subject without undue delay.
2. The communication to the data subject referred to in paragraph 1 of this Article shall describe in clear and plain language the nature of the personal data breach and contain at least the information and measures referred to in points (b), (c) and (d) of Article 33(3).
3. The communication to the data subject referred to in paragraph 1 shall not be required if any of the following conditions are met:
   a. the controller has implemented appropriate technical and organisational protection measures, and those measures were applied to the personal data affected by the personal data breach, in particular those that render the personal data unintelligible to any person who is not authorised to access it, such as encryption;
   b. the controller has taken subsequent measures which ensure that the high risk to the rights and freedoms of data subjects referred to in paragraph 1 is no longer likely to materialise;
   c. it would involve disproportionate effort. In such a case, there shall instead be a public communication or similar measure whereby the data subjects are informed in an equally effective manner.
4. If the controller has not already communicated the personal data breach to the data subject, the supervisory authority, having considered the likelihood of the personal data breach resulting in a high risk, may require it to do so or may decide that any of the conditions referred to in paragraph 3 are met.

Note

Whereas Article 33 required notifying the supervisory authority within 72 hours, notifying the data subject needs to be done "without undue delay."

References

1. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 1.

This page titled 3.2.21: Article 34 Communication of a personal data breach to the data subject is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

3.2.22: Articles 35-36 Data Protection Impact Assessments

Note

Articles 35 and 36 cover Data Protection Impact Assessments for new technology, to include discussion of necessity and proportionality, risk assessment, and risk mitigation. Where a controller assesses high risk, the controller shall consult the supervisory [data protection] authority.

Data Protection Impact Assessments are among the compliance documentation mandated and specified by the GDPR.

This page titled 3.2.22: Articles 35-36 Data Protection Impact Assessments is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

3.2.23: Article 37 Designation of the data protection officer

1. The controller and the processor shall designate a data protection officer in any case where:
   a. the processing is carried out by a public authority or body, except for courts acting in their judicial capacity;
   b. the core activities of the controller or the processor consist of processing operations which, by virtue of their nature, their scope and/or their purposes, require regular and systematic monitoring of data subjects on a large scale; or
   c. the core activities of the controller or the processor consist of processing on a large scale of special categories of data pursuant to Article 9 and personal data relating to criminal convictions and offences referred to in Article 10.
2. A group of undertakings may appoint a single data protection officer provided that a data protection officer is easily accessible from each establishment.
3. Where the controller or the processor is a public authority or body, a single data protection officer may be designated for several such authorities or bodies, taking account of their organisational structure and size.
4. In cases other than those referred to in paragraph 1, the controller or processor or associations and other bodies representing categories of controllers or processors may or, where required by Union or Member State law shall, designate a data protection officer. The data protection officer may act for such associations and other bodies representing controllers or processors.
5. The data protection officer shall be designated on the basis of professional qualities and, in particular, expert knowledge of data protection law and practices and the ability to fulfil the tasks referred to in Article 39.
6. The data protection officer may be a staff member of the controller or processor, or fulfil the tasks on the basis of a service contract.
7. The controller or the processor shall publish the contact details of the data protection officer and communicate them to the supervisory authority.

Note

In item 5, the GDPR uses "fulfil" (British English spelling) versus "fulfill" (American English spelling). It is not a typo.

References

1. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 1.

This page titled 3.2.23: Article 37 Designation of the data protection officer is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

3.2.24: Article 38 Position of the data protection officer

1. The controller and the processor shall ensure that the data protection officer is involved, properly and in a timely manner, in all issues which relate to the protection of personal data.
2. The controller and processor shall support the data protection officer in performing the tasks referred to in Article 39 by providing resources necessary to carry out those tasks and access to personal data and processing operations, and to maintain his or her expert knowledge.
3. The controller and processor shall ensure that the data protection officer does not receive any instructions regarding the exercise of those tasks. He or she shall not be dismissed or penalised by the controller or the processor for performing his tasks. The data protection officer shall directly report to the highest management level of the controller or the processor.
4. Data subjects may contact the data protection officer with regard to all issues related to processing of their personal data and to the exercise of their rights under this Regulation.
5. The data protection officer shall be bound by secrecy or confidentiality concerning the performance of his or her tasks, in accordance with Union or Member State law.
6. The data protection officer may fulfil other tasks and duties. The controller or processor shall ensure that any such tasks and duties do not result in a conflict of interests.

Note

This Article is positioning the DPO to have a degree of independence and a position in the organization chart to avoid conflicts of interest.

References

1. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 1.

This page titled 3.2.24: Article 38 Position of the data protection officer is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

3.2.25: Article 39 Tasks of the data protection officer

1. The data protection officer shall have at least the following tasks:
   a. to inform and advise the controller or the processor and the employees who carry out processing of their obligations pursuant to this Regulation and to other Union or Member State data protection provisions;
   b. to monitor compliance with this Regulation, with other Union or Member State data protection provisions and with the policies of the controller or processor in relation to the protection of personal data, including the assignment of responsibilities, awareness-raising and training of staff involved in processing operations, and the related audits;
   c. to provide advice where requested as regards the data protection impact assessment and monitor its performance pursuant to Article 35;
   d. to cooperate with the supervisory authority;
   e. to act as the contact point for the supervisory authority on issues relating to processing, including the prior consultation referred to in Article 36, and to consult, where appropriate, with regard to any other matter.
2. The data protection officer shall in the performance of his or her tasks have due regard to the risk associated with processing operations, taking into account the nature, scope, context and purposes of processing.

References

1. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 1.

This page titled 3.2.25: Article 39 Tasks of the data protection officer is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

3.2.26: Article 42 Certification

1. The Member States, the supervisory authorities, the Board and the Commission shall encourage, in particular at Union level, the establishment of data protection certification mechanisms and of data protection seals and marks, for the purpose of demonstrating compliance with this Regulation of processing operations by controllers and processors. The specific needs of micro, small and medium-sized enterprises shall be taken into account.
2. In addition to adherence by controllers or processors subject to this Regulation, data protection certification mechanisms, seals or marks approved pursuant to paragraph 5 of this Article may be established for the purpose of demonstrating the existence of appropriate safeguards provided by controllers or processors that are not subject to this Regulation pursuant to Article 3 within the framework of personal data transfers to third countries or international organisations under the terms referred to in point (f) of Article 46(2). Such controllers or processors shall make binding and enforceable commitments, via contractual or other legally binding instruments, to apply those appropriate safeguards, including with regard to the rights of data subjects.
3. The certification shall be voluntary and available via a process that is transparent.
4. A certification pursuant to this Article does not reduce the responsibility of the controller or the processor for compliance with this Regulation and is without prejudice to the tasks and powers of the supervisory authorities which are competent pursuant to Article 55 or 56.
5. A certification pursuant to this Article shall be issued by the certification bodies referred to in Article 43 or by the competent supervisory authority, on the basis of criteria approved by that competent supervisory authority pursuant to Article 58(3) or by the Board pursuant to Article 63. Where the criteria are approved by the Board, this may result in a common certification, the European Data Protection Seal.
6. The controller or processor which submits its processing to the certification mechanism shall provide the certification body referred to in Article 43, or where applicable, the competent supervisory authority, with all information and access to its processing activities which are necessary to conduct the certification procedure.
7. Certification shall be issued to a controller or processor for a maximum period of three years and may be renewed, under the same conditions, provided that the relevant requirements continue to be met. Certification shall be withdrawn, as applicable, by the certification bodies referred to in Article 43 or by the competent supervisory authority where the requirements for the certification are not or are no longer met.
8. The Board shall collate all certification mechanisms and data protection seals and marks in a register and shall make them publicly available by any appropriate means.

Note

Item 3 makes it clear that certification of compliance is voluntary, not mandatory.

References

1. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 1.

This page titled 3.2.26: Article 42 Certification is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

3.2.27: Article 83 General conditions for imposing administrative fines

Note

Note: See especially 4, 5, and 6 regarding the potential severity of the fines.

1. Each supervisory authority shall ensure that the imposition of administrative fines pursuant to this Article in respect of infringements of this Regulation referred to in paragraphs 4, 5 and 6 shall in each individual case be effective, proportionate and dissuasive.
2. Administrative fines shall, depending on the circumstances of each individual case, be imposed in addition to, or instead of, measures referred to in points (a) to (h) and (j) of Article 58(2). When deciding whether to impose an administrative fine and deciding on the amount of the administrative fine in each individual case due regard shall be given to the following:
   a. the nature, gravity and duration of the infringement taking into account the nature scope or purpose of the processing concerned as well as the number of data subjects affected and the level of damage suffered by them;
   b. the intentional or negligent character of the infringement;
   c. any action taken by the controller or processor to mitigate the damage suffered by data subjects;
   d. the degree of responsibility of the controller or processor taking into account technical and organisational measures implemented by them pursuant to Articles 25 and 32;
   e. any relevant previous infringements by the controller or processor;
   f. the degree of cooperation with the supervisory authority, in order to remedy the infringement and mitigate the possible adverse effects of the infringement;
   g. the categories of personal data affected by the infringement;
   h. the manner in which the infringement became known to the supervisory authority, in particular whether, and if so to what extent, the controller or processor notified the infringement;
   i. where measures referred to in Article 58(2) have previously been ordered against the controller or processor concerned with regard to the same subject-matter, compliance with those measures;
   j. adherence to approved codes of conduct pursuant to Article 40 or approved certification mechanisms pursuant to Article 42; and
   k. any other aggravating or mitigating factor applicable to the circumstances of the case, such as financial benefits gained, or losses avoided, directly or indirectly, from the infringement.
3. If a controller or processor intentionally or negligently, for the same or linked processing operations, infringes several provisions of this Regulation, the total amount of the administrative fine shall not exceed the amount specified for the gravest infringement.
4. Infringements of the following provisions shall, in accordance with paragraph 2, be subject to administrative fines up to 10 000 000 EUR, or in the case of an undertaking, up to 2 % of the total worldwide annual turnover of the preceding financial year, whichever is higher:
   a. the obligations of the controller and the processor pursuant to Articles 8, 11, 25 to 39 and 42 and 43;
   b. the obligations of the certification body pursuant to Articles 42 and 43;
   c. the obligations of the monitoring body pursuant to Article 41(4).
5. 5.Infringements of the following provisions shall, in accordance with paragraph 2, be subject to administrative fines up to 20 000 000 EUR, or in the case of an undertaking, up to 4 % of the total worldwide annual turnover of the preceding financial year, whichever is higher:
   a. the basic principles for processing, including conditions for consent, pursuant to Articles 5, 6, 7 and 9;
   b. the data subjects' rights pursuant to Articles 12 to 22;
   c. the transfers of personal data to a recipient in a third country or an international organisation pursuant to Articles 44 to 49;
   d. any obligations pursuant to Member State law adopted under Chapter IX;
   e. non-compliance with an order or a temporary or definitive limitation on processing or the suspension of data flows by the supervisory authority pursuant to Article 58(2) or failure to provide access in violation of Article 58(1).
6. Non-compliance with an order by the supervisory authority as referred to in Article 58(2) shall, in accordance with paragraph 2 of this Article, be subject to administrative fines up to 20 000 000 EUR, or in the case of an undertaking, up to 4 % of the total worldwide annual turnover of the preceding financial year, whichever is higher.

7. Without prejudice to the corrective powers of supervisory authorities pursuant to Article 58(2), each Member State may lay down the rules on whether and to what extent administrative fines may be imposed on public authorities and bodies established in that Member State.
8. The exercise by the supervisory authority of its powers under this Article shall be subject to appropriate procedural safeguards in accordance with Union and Member State law, including effective judicial remedy and due process.
9. Where the legal system of the Member State does not provide for administrative fines, this Article may be applied in such a manner that the fine is initiated by the competent supervisory authority and imposed by competent national courts, while ensuring that those legal remedies are effective and have an equivalent effect to the administrative fines imposed by supervisory authorities. In any event, the fines imposed shall be effective, proportionate and dissuasive. Those Member States shall notify to the Commission the provisions of their laws which they adopt pursuant to this paragraph by 25 May 2018 and, without delay, any subsequent amendment law or amendment affecting them.

References

1. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 1.

This page titled 3.2.27: Article 83 General conditions for imposing administrative fines is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

3.3: Implementing GDPR Compliance

This is an important section as it will be useful in completing the assignment of preparing a GDPR implementation schedule.

As noted earlier, the GDPR applies to the processing of personal data in the context of the activities of an establishment of a controller or a processor in the EU, regardless of whether the processing takes place in the Union or not. The GDPR applies to data controllers and processors outside the Union if they are engaged in offering goods and services to data subjects in the Union or are “monitoring the behavior of data subjects within the Union”.

Based on what the GDPR requires, what an organization needs to do is going to be based on an assessment of what the current state of compliance is and what needs to be done to achieve the desired level of compliance. This can be a self-assessment or one aided by external consultants who specialize in GDPR compliance.

Now that you have been able to read the actual text of relevant GDPR articles, let's start translating this into implementation.

> **Note**
>
> The EU Commission began considering proposed amendments and substantive changes to the GDPR in late 2025 as part of an omnibus reform to its data regulations, sparking considerable controversy. It is important for your implementation planning to ensure that you are working with the latest text of the Regulation, including any amendments the EU may have made. Additionally, it is worthwhile to include in your compliance plan any relevant contingencies your organization should take in the event the EU amends the GDPR in the future.

This page titled 3.3: Implementing GDPR Compliance is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

3.3.1: Implementation Considerations

Summary of the relevant GDPR requirements:

- Consent (Article 6)
- Data subject rights (GDPR Chapter 3, Articles 12-23)
- Right to erasure/Right to be forgotten (Article 17)
- Marketing opt-out (Article 21)
- Data encryption (Article 32)
- Pseudonymization (Article 25 and 32)
- Handling data breach notifications. Prepare relevant processes. (Articles 33-34)
- Hire DPO (Articles 37 and 38)
- Certifications (Article 42)
- Cross-border data transfers (Articles 44-49)

Overarching Considerations

Companies were aware of the regulation two years before it went into effect. Some worked to attain compliance by the date the GDPR went into effect, while others did not and have had issues. Achieving compliance requires investing resources to comply with the consent, data mapping, and cross-border data transfer requirements, among others.

Organizations that want to attain GDPR compliance have a few options:

- Change their existing business systems to handle the EU as a special case. The user's profile can be inspected to determine their address. The user's IP address can be used to ascertain location.
- Change business systems to extend GDPR protections to all users globally.

Taking stock

Before launching into implementation, it would be wise to do a detailed assessment of the status quo, the current state of the organization's level of compliance with the GDPR articles. Organizations that had already implemented compliance with the predecessor Data Protection Directive (DPD) had a head start.

This would involve assessing what user data is currently collected, what operations are performed on it, how and where is the data stored, and where is it transmitted. This assessment should include levels of pseudonymization and encryption. This assessment would include controllers and processors.

An output of this assessment would be the identification existing gaps in GDPR compliance.

Consent (Article 6)

The data subject provides “informed consent” for each specific purpose that the data is going to be used for. The implementation cannot bundle multiple data processing types together when seeking consent. The data subject can revoke consent/permission any time.

There is some good guidance provided by Recital 32.

> "Consent should be given by a clear affirmative act establishing a freely given, specific, informed and unambiguous indication of the data subject's agreement to the processing of personal data relating to him or her, such as by a written statement, including by electronic means, or an oral statement. This could include ticking a box when visiting an internet website, choosing technical settings for information society services or another statement or conduct which clearly indicates in this context the data subject's acceptance of the proposed processing of his or her personal data. Silence, pre-ticked boxes or inactivity should not therefore constitute consent. Consent should cover all processing activities carried out for the same purpose or purposes. When the processing has multiple purposes, consent should be given for all of them. If the

data subject's consent is to be given following a request by electronic means, the request must be clear, concise and not unnecessarily disruptive to the use of the service for which it is provided."

Note

Note the specific reference to "pre-ticked boxes or inactivity."

Note also that when the processing has multiple purposes, it is necessary to obtain consent of all of them (i.e., each of them).

See also Marketing Opt-out (Article 21).

Data Subject Rights (Articles 12-23)

- Right to be informed about what is being collected, how is it used, how long is it retained.
- Right to objection and “profiling”(e.g., for direct marketing): can request end to such use.
- Right to request copy of data collected.
- Right to have data erased.

In all of these cases, a user interface needs to make this option available to interested data subjects and to obtain their input and take the actions required to fulfill the intent of the request. The user should be informed that the request has been accepted and to provide a timeframe within which the request likely will be completed and notify the user when the request has been completed.

Right to erasure/Right to be forgotten (Article 17)

This will require a user interface to make this option available to interested data subjects. Presumably the controller will gather this request from the data subject. The controller will also need to pass the request on to all relevant downstream processors.

In addition to erasing all primary copies of the data of the data subject, all other copies of the data need to be erased.

This raises an interesting question about how to handle data backups that include a broad range of data across multiple data subjects.

Direct Marketing Opt-out (Article 21)

GDPR Article 21 allows a data subject to object to using personal information for marketing, sales, or non-service related purposes (including “profiling”).

Users “opting out” could impact a major source of advertising revenue for some companies. Companies that do most of their business on the internet can derive a significant amount of revenue from advertising. The ads might be targeted to certain users or user classes based on their data. If a company derives 95% of its revenues from direct marketing, asking a user if they want to opt out from having their data be used for direct marketing purposes to be in compliance with the GDPR would likely generate some interesting boardroom discussions.

For example, in designing the user interface for this, one might consider the following possible wireframes to be presented to the user.

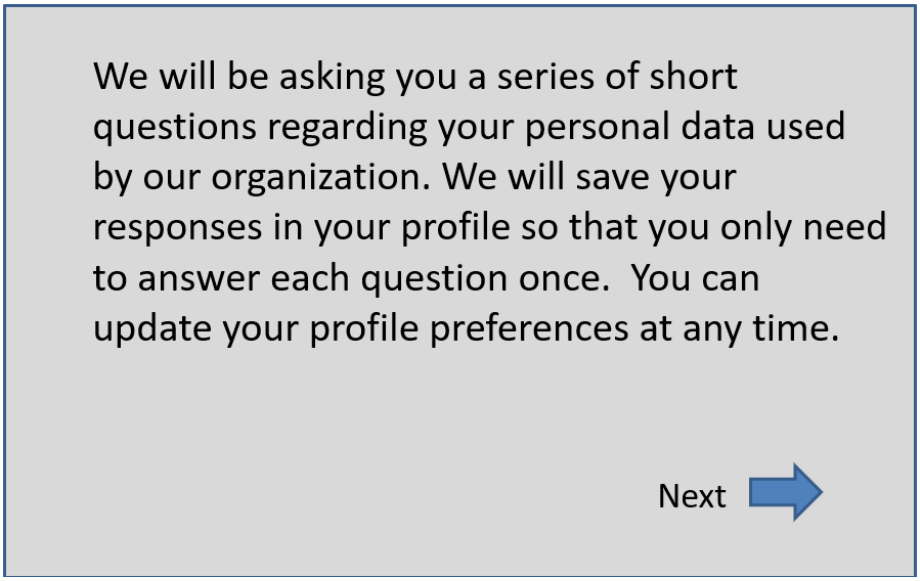


Figure 3.3.1.1: Wireframe for GDPR Starting Point. (Robert Slate, CIS 79 course materials)

Accessibility Note

Figure 3.3.1.1: Wireframe for GDPR Starting Point shows a mock-up creen stating "We will be asking you a series of short questions regarding your personal data used by our organization. We will save your responses in your profile so that you only need to answer each question once. You can update your profile preferences at any time." Next blue-colored arrow button.

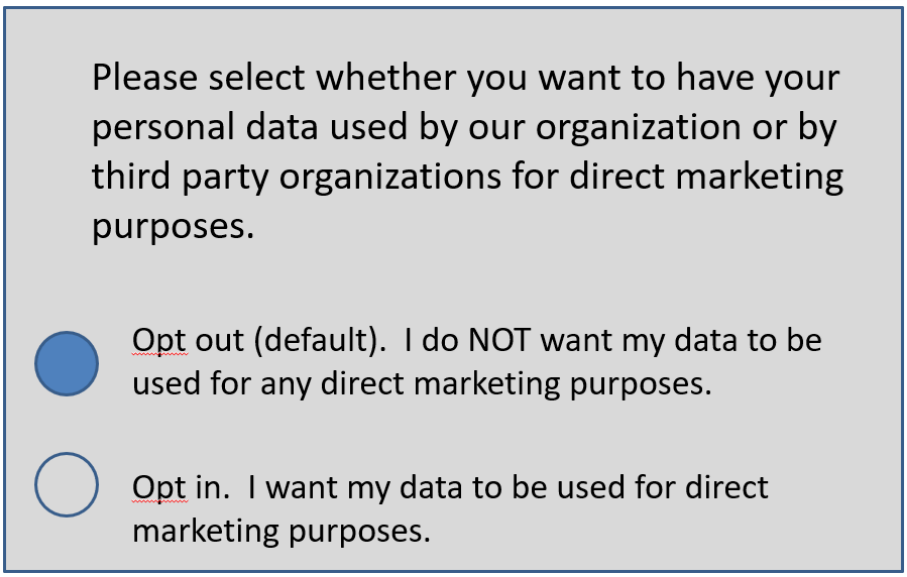


Figure 3.3.1.2: Wireframe for GDPR Opt-Out. (Robert Slate, CIS 79 course materials)

Accessibility Note

Figure 3.3.1.2: Wireframe for GDPR Opt-Out shows a mock-up screen with text: "Please select whether you want to have your personal data used by our organization or by third party organizations for direct marketing purposes." With Opt-out and Opt-in options described in the text.

Note that this screen specifically pre-selects the "Opt-out" option as the default option as required. The user would have to specifically select the "Opt-in" option to allow his or her data to be used for direct marketing purposes.

Pseudonymisation/Pseudonymization

- Article 25 requires that data protection mechanisms be designed into business systems.
- With Pseudonymisation/Pseudonymization: data cannot be tied to a specific “data subject.” Note that data encryption can be employed as a data protection mechanism.

Pseudonymization is a procedure by which the most identifying fields within a data record are replaced by one or more artificial identifiers, or pseudonyms. [Ref 1]

Definition: Pseudonymisation/Pseudonymization

‘pseudonymisation’ means the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person.

GDPR Article 4 (5)

Example: John Smith maps to "Person 1", Sally Jones maps to "Person 2".

Note

However, 87% of the United States actually has a unique combination of birth date, gender, and zip code, so someone who has those 3 data items might be able to re-identify people. [Ref. 1]

Full/complete anonymization could be employed as well.

Default: highest privacy settings, data not publicly available.

Relevant GDPR Recitals concerning Pseudonymisation/Pseudonymization

GDPR Recitals 26, 28, and 29 provide additional context on Pseudonymisation/Pseudonymization.

Recital 26

> The principles of data protection should apply to any information concerning an identified or identifiable natural person. Personal data which have undergone pseudonymisation, which could be attributed to a natural person by the use of additional information should be considered to be information on an identifiable natural person. To determine whether a natural person is identifiable, account should be taken of all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly. To ascertain whether means are reasonably likely to be used to identify the natural person, account should be taken of all objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments. The principles of data protection should therefore not apply to anonymous information, namely information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable. This Regulation does not therefore concern the processing of such anonymous information, including for statistical or research purposes.

Recital 28

> The application of pseudonymisation to personal data can reduce the risks to the data subjects concerned and help

controllers and processors to meet their data-protection obligations. The explicit introduction of 'pseudonymisation' in this Regulation is not intended to preclude any other measures of data protection.

Recital 29

In order to create incentives to apply pseudonymisation when processing personal data, measures of pseudonymisation should, whilst allowing general analysis, be possible within the same controller when that controller has taken technical and organisational measures necessary to ensure, for the processing concerned, that this Regulation is implemented, and that additional information for attributing the personal data to a specific data subject is kept separately. The controller processing the personal data should indicate the authorised persons within the same controller.

Hire DPO (Articles 37 and 38)

The DPO can be either an internal employee or someone who works outside of the organization. Where the DPO reports into the organization is an important consideration in order to maintain the required independence and not have conflicts of interest. The hiring of the DPO may or may not be part of the compliance project as it could be handled as an executive staff action item.

Certifications (Article 42)

The company or organization can decide whether it wishes to pursue a formal certification per Article 42 to help mitigate risk and to inform its customers on its level of compliance. Remember, formal certification is voluntary, not mandatory.

Implement new data protection and network protection procedures

The data protection requirements have been previously discussed. Since data breaches need to be reported both to supervisory authorities as well as to affected data subjects, there likely would need to be an assessment of existing network security and closing identified gaps as required. It would be useful to document the relevant organizational processes and procedures.

Handling data breach notifications (Articles 33-34)

Implementing this item is more about preparing the relevant processes. Ideally the network and corporate data processing infrastructure should be proactively strengthened to minimize the potential for data breaches. However, in the event a data breach occurs occur, the processes need to be in place to allow for smooth execution to meet the GDPR requirements and notification timelines.

This could be done in many ways. Conceptually, a list of tasks needs to be created along with a RACI matrix to show who is Responsible, Accountable, Consulted, and Informed.

References

1. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 1.
2. Bowman, Gesher, Grant, and Slate, The Architecture of Privacy, 2015, page 166.

This page titled 3.3.1: Implementation Considerations is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

3.4: Project Management Considerations

Your boss has dropped by and informed you that you have been selected to be the project manager of your organization's GDPR compliance efforts. "I hope you know something about GDPR" you boss says before walking off.

Well, you actually have never heard about GDPR and have no idea what it is. If you are going to run the project, you will need to come up to speed quickly so that you can exhibit some confidence when you assemble your team and have your first project meeting. You likely will use your search engine to do some reading about the topic. The first sections of this chapter have summarized some of the salient points. It has provided you with some of the necessary terminology and pulled some of the GDPR Articles that could be part of most compliance efforts.

Assessing the current state

Based on what the GDPR regulation requires, what an organization needs to do is going to be based on an assessment of what the current state of compliance is and what needs to be done to achieve the desired level of compliance.

The assessment likely would be conducted by a cross-functional team with representatives from, but not limited to, the following organizations:

- Legal staff
- Information technology or product development staff familiar with the relevant applications, data bases, user interfaces, etc.
- Network staff who can address network security issues
- Business analysts familiar with the company's processes and flows
- Corporate risk analysis staff
- Project Management Office/Project manager

Identifying gaps

After completing the assessment, you are now aware of gaps. What gaps? The gaps could be, but are not limited to, any of the following:

- Implementation gaps. Based on your current organizational processes, implementation of key applications, and networking infrastructure, you have itemized things that need to be done to close the gaps to get the organization to what your team perceives to be the desired state of compliance.
- Capability gaps. You have determined that your organization does not currently have the staff on hand in the organization to proceed with certain elements of the implementation. Your organization may not currently have a Data Protection Officer (DPO).

Preparing a strawman plan of action

A Work Breakdown Structure (WBS) can be prepared based on a preliminary scope for the project. A preliminary straw-man schedule can be prepared along with a preliminary project budget and resource plan.

Getting project steering committee approval (Concept)

At this point, a recommendation about scope, budget, resources, and schedule can be presented as part of a concept review with an executive project steering committee. The project steering committee can review this preliminary work and determine what changes are needed to be consistent with corporate goals and strategy and notify the project team to proceed accordingly.

Preparing a formal project plan

The project team can then incorporate these changes and begin work on formal project planning. The project scope and detailed

requirements can be documented. A formal project plan can be prepared including a schedule, resource plan, and budget.

This is where it gets real, in the sense that the resources that are applied to the project activities must actually be available to be assigned to project work. Some hiring for in-house resources may need to be identified and hired by the responsible development groups. Some external consulting resources may need to be procured as external contractors.

For all of the required development work, the technical teams can prepare formal specifications.

Getting project steering committee approval (Planning/Specification)

At this point, a formal Phase 1 exit review team is scheduled with the project steering committee to gain official approval to proceed.

The list of potential work was covered in the "Implementing GDPR Compliance" section. The work to be done will vary from company to company based on each company's starting point and corporate strategy and objectives. The sense of urgency in completing the work will also span a broad spectrum. In addition, the risk appetite of the executive team may be such that they will take the risk to not implement certain work.

The project scope could also include whether the company is planning to seek a formal certification and whether outside consultants might be required for that certification effort.

The potential hiring of a Data Protection Officer needs to be addressed. That may or may not be within the scope of the project. It may be an executive action item.

After gaining that approval, the project can proceed to implement the agreed-upon scope, using the approved resources, and go so within the approved budget.

On-going project status monitoring

It is likely that a compliance project such as GDPR compliance is going to attract executive attention. As a result, plans should be made to have on-going project status reviews with the project steering committee and any other stakeholders. The project manager can reference the stakeholder register to determine what other executive check-ins are needed along with their frequency.

This page titled 3.4: Project Management Considerations is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

3.5: Legal Considerations

Learning Objectives

- Understand why US companies are affected by this EU regulation
- Understand the size of potential penalties
- Understand why companies that violate the GDPR are being fined substantial amounts of money
- Understand what arguments are being put forward to justify the positions of non-compliant companies

Why US Companies can be affected

The GDPR imposes rules on organizations that offer goods and services to people in the EU, or that collect and analyze data tied to EU residents. The GDPR applies to any entity processing personal information of "data subjects" inside the EU/EEA (regardless of company's location or data subject's citizenship/residence). The GDPR covers international companies with data processing centers in the EU.

Penalties

The following sets out the GDPR fines:

- "The less severe infringements could result in a fine of **up to €10 million, or 2% of the firm's worldwide annual revenue** from the preceding financial year, **whichever amount is higher."**
- "The more serious infringements go against the very principles of the right to privacy and the right to be forgotten that are at the heart of the GDPR. These types of infringements could result in a fine of up to **€20 million, or 4% of the firm's worldwide annual revenue** from the preceding financial year, **whichever amount is higher**." [Ref 1]

Business decision analysis may result in accepting legal penalties in lieu of compliance. This could occur when a large percentage of a company's revenue is derived from direct marketing activities based on user data. A risk assessment might conclude that if a high percentage of users opt out of having their data be used for direct marketing purposes, that loss of revenue will be large in comparison to paying the penalty resulting from non-compliance.

GDPR Fines in Practice

The following are examples of fines that have been imposed.

- Meta Platforms 1.2 billion euros for EU to US data transfers.
- Uber Technologies, Inc 825 million euros for using automated systems to suspend driver accounts.[2] Note: this is an Article 22 violation.
- Amazon 746 million euros (sent back to the regulator)
- TikTok 530 million euros
- Instagram 405 million euros
- Meta 390 million euros for EU to US data transfers.
- TikTok 345 million euros
- LinkedIn 310 million euros
- Uber 290 million euros
- Meta 265 million euros
- Whatsapp 225 million euros

Data Subjects' Rights in court

- Article 77: private right to lodge complaints with supervisory [data protection] authority of the Member State of one's

"habitual residence"
- Article 79: data subjects can bring action in courts of the Member State where the controller or processor has an establishment, or where the data subject has "habitual residence"

Argumentation

Here, we will briefly explore two interpretations issued by the second-highest authority in applying the GDPR, the European Data Protection Board (EDPB).

EDPB Binding Decision 3/2022 on Meta IE

- Ireland's GDPR supervisory authority circulated a draft decision to the other EU data protection authorities (see Art. 60 GDPR). They disputed the finding that contract was a lawful basis for Meta's processing. The Irish DPA ("IE SA") asked the EDPB for a binding decision to resolve the dispute.
- The EDPB's decision concerns how to analyze contract as a lawful basis for necessary processing (Art. 6(1)(b) GDPR). Relevant excerpts from the decision are included and annotated below.

Contract as Lawful Basis

- "It is important to determine the exact rationale of the contract, i.e., its substance and fundamental objective, as it is against this that it will be tested whether the data processing is necessary for its performance" (para. 112)
- "[R]egard should be given to the particular aim, purpose, or objective of the service, and…it is required that the processing is objectively necessary for a purpose and integral to the delivery of that contractual service to the data subject" (para. 112)
- "…the controller should be able to justify the necessity of its processing by reference to the fundamental and mutually understood contractual purpose."
- "…as a general rule, processing of personal data for behavioral advertising is not necessary for the performance of a contract for online services" (para. 114)
- "…the EDPB recalls that the concept of necessity has its own independent meaning under EU law. It must be interpreted in a manner that fully reflects the objective pursued by an EU instrument, in this case, the GDPR. Accordingly, the concept of necessity cannot be interpreted in a way that undermines this provision and the GDPR's general objective of protecting the right to the protection of personal data or contradicts Article 8 of the [EU] Charter" (para. 116)

> **Note**
>
> Note the recurrence of the concept of necessity.

EDPB Binding Decision

- The EDPB found that users sign up for Facebook mainly for the purpose of communicating with others, not to get personalized, behavioral ads (para. 124)
- A reasonable user could not expect their signing up to have their data processed for behavioral advertising; the terms of service and related documents are too sparse in their discussion about this type of processing (para. 123)

Artificial Intelligence and the GDPR

- In late 2024, the EDPB issued a major opinion on AI Labs' compliance (or lack thereof) with the GDPR in the training of their models
- Recall: under Article 6, data processing must be lawful (e.g., via the consent of the data subject (Art. 6(1)(a), contracting with a data subject (Art. 6(1)(b)), or "legitimate interests" (Art. 6(1)(f))
- At the time, most AI labs appeared to be relying on 6(1)(f)

The EDPB issued Guidelines 1/2024 (October) on how to fulfill Art. 6(1)(f). Three conditions included:

1. Pursuit of a legitimate interest by controller or 3rd party)

2. The need to process personal data for the purpose of (1)
3. The interests or fundamental freedoms and rights of data subjects affected by (2) do not take precedence over (1).

References

1. What are the GDPR Fines? - GDPR.eu
2. Charlotte Hughes-Morgan. Bloomberg. "Dutch Regulators penalize Uber under EU data policy." The Mercury News, Section C, page 8, Sunday August 23, 2026.

This page titled 3.5: Legal Considerations is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

Terms

Note

Note that many of these terms are defined in GDPR Article 4, covered earlier in this chapter.

- **Consent:** "Consent'"of the data subject means any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her.
- **Controller:** The natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data.
- **Cross-border processing:** 'cross-border processing' means either: (a) processing of personal data which takes place in the context of the activities of establishments in more than one Member State of a controller or processor in the Union where the controller or processor is established in more than one Member State; or (b) processing of personal data which takes place in the context of the activities of a single establishment of a controller or processor in the Union but which substantially affects or is likely to substantially affect data subjects in more than one Member State.
- **Data Protection Officer** (DPO)
- **Data Subject:** an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person
- **EEA** (European Economic Area)
- **EU** (European Union)
- **GDPR:** the European Union's General Data Protection Regulation.
- **Personal data:** Personal data' means any information relating to an identified or identifiable natural person ('data subject').
- **Processor:** A natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller.
- **Processing:** 'processing' means any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction;
- **Profiling:** Any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements."
- **Pseudonymization/Pseudonymisation:** Pseudonymization is a procedure by which the most identifying fields within a data record are replaced by one or more artificial identifiers, or pseudonyms. [Ref 1]

References

1. Bowman, Gesher, Grant, and Slate, The Architecture of Privacy, 2015, page 166.
2. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 1.

This page titled 3.6: Glossary is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

4: NETWORK SWITCH DEVELOPMENT

Thumbnail figure of RJ-45 Ethernet Ports on a Cisco Catalyst 4006 switch. Ben Franske. File:RJ45 Switch Ports.jpg - Wikimedia Commons. August 29, 2009.This file is licensed under the Creative Commons Attribution-Share Alike 4.0 International, 3.0 Unported, 2.5 Generic, 2.0 Generic and 1.0 Generic license.

4.1: Introduction

4.2: Types of Network Switches

4.3: System Scope

4.4: Hardware System Design

4.4.1: Network ASICs

4.4.2: Chassis Design

4.4.3: Board Design

4.4.4: Signal integrity

4.4.5: Electro-magnetic Regulatory Compliance

4.4.6: Testing (HALT)

4.5: Software System Design

4.6: Project Management Considerations

4.6.1: Procurement

4.6.1.1: Bid Documents (RFI, RFP, etc.)

4.6.1.2: PR to PO procedure

4.6.2: Team Management

4.6.2.1: Content Theories of Motivation

4.6.2.2: Leadership

4.6.2.2.1: Introduction

4.6.2.2.2: The Nature of Leadership

4.6.2.2.3: The Leadership Process

4.6.2.2.4: Types of Leaders and Leader Emergence

4.6.2.2.5: The Trait Approach to Leadership

4.6.2.2.6: Behavioral Approaches to Leadership

4.6.2.2.7: Situational (Contingency) Approaches to Leadership

4.6.2.2.8: Substitutes for and Neutralizers of Leadership

4.6.2.2.9: Transformational, Visionary, and Charismatic Leadership

4.6.2.2.10: Leadership Needs in the 21st Century

4.6.2.2.11: Summary

4.7: Legal Considerations

4.7.1: Patents

4.7.1.1: The Patenting Process

4.7.1.2: Patent Disclosures

4.7.1.3: Patent cross-licensing

4.7.2: Trade Secrets

4.7.2.1: Trade Secret Protections

4.7.2.2: The Foundations of Trade Secrets Law

4.7.2.3: Elements of a Trade Secret

4.7.2.4: The Secrecy Requirement

4.7.2.5: Misappropriation of Trade Secrets

4.7.2.6: Remedies Available for the Misappropriation of Trade Secrets

4.7.2.7: Assessment Questions

4.7.3: Trademark Basics

4.7.3.1: Core Concepts

4.7.3.2: Early Trademark Systems

4.7.3.3: U.S. Trademark Law

4.7.3.4: Four Types of Trademarks

4.7.3.5: Subject Matter of Trademarks

4.7.3.6: The Spectrum of Distinctiveness

4.7.3.7: Bars to Trademark

4.7.3.8: Establishing Trademark Protection

4.7.3.9: Trademark Infringement

4.7.3.10: Trademark Remedies

4.7.3.11: Fair Use of Trademarks

4.7.3.12: Assessment Questions

4.8: Glossary

This page titled 4: Network Switch Development is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

4.1: Introduction

A network switch or Ethernet switch is a multi-port networking hardware system that connects devices on a computer network using packet switching to receive, process, and forward data to a destination device whose address is in the packet.[1]

A simple network hub broadcasts the same data out of each of its ports. However, a network switch forwards data only to one or multiple devices that need to receive the packet.[2]

Packet Switching

Packet switching is a method of grouping data to be transmitted over a telecommunications network into short messages in a fixed format, called a packet.

Packets consist of a "header" section and a "payload" section.

- The header provides the means for directing a packet to its destination (e.g., Destination MAC address)
- The payload consists of the actual data that an application wants to transmit.[3]

In the Open Systems Interconnect (OSI) model developed by the International Organization for Standardization (ISO), there are several layers:

1. Physical layer
2. Data Link layer
3. Network layer
4. Transport layer
5. Session layer
6. Presentation layer
7. Application layer.

Ethernet frame

An Ethernet frame is a data link layer (Layer 2) protocol data unit that uses the underlying Ethernet physical layer (Layer 1) transport mechanisms.

- Layer 2 is the layer at which one transmits data frames between two network nodes that are connected by a physical layer.
- Layer 1 handles the transmission and reception of bit streams over a physical medium (e.g., copper cable, fiber optic cable).

An Ethernet frame (described below) is preceded by a "preamble" and a "start frame delimiter (SFD):, which are both part of the Ethernet packet at the physical layer (Layer 1).

An Ethernet frame consists of:

- An Ethernet header, which contains destination and source Media Access Control (MAC) Addresses as its first two fields. A MAC address is a unique identifier assigned by network device manufacturers and are typically store in the device's read-only memory.
- The middle section of the frame is payload data, including any headers for other protocols (for example, Internet Protocol) carried in the frame.
- The frame ends with a Frame Check Sequence (FCS), which is a 32-bit Cyclic Redundancy Check (CRC) that is used to detect any in-transit corruption of data.[4]

An Ethernet packet is shown in Figure 4.1.1.

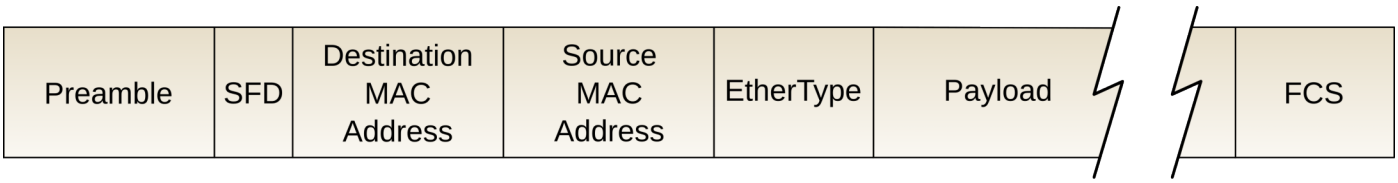


Figure 4.1.1: Ethernet packet. The SFD (start frame delimiter) marks the end of the packet preamble. It is immediately followed by the Ethernet frame, which starts with the destination MAC address. (Ethernet frame - Wikipedia, in the public domain)

Accessibility Note

An Ethernet packet is structured as follows:

- Preamble
- Start Frame Delimiter (SFD)
- Destination MAC Address
- Source MAC Address
- Ethertype
- Payload
- Frame Check Sequence (FCS)

A data packet on the wire and the frame as its payload consist of binary data. Ethernet transmits data with the most-significant octet of eight bits (byte) first; within each octet, however, the least-significant bit is transmitted first.[4]

The internal structure of an Ethernet frame is specified in the IEEE 802.3 specification. Figure 4.1.2 shows the complete Ethernet packet and the frame inside, as transmitted, for the payload size up to the Maximum Transmission Unit (MTU) of 1500 octets. Some implementations of Gigabit Ethernet and other higher-speed variants of Ethernet support larger frames that are called jumbo frames.[4]

802.3 Ethernet packet and frame structure

Layer	Preamble	Start frame delimiter (SFD)	MAC destination	MAC source	802.1Q tag (optional)	Ethertype (Ethernet II) or length (IEEE 802.3)	Payload	Frame check sequence (32-bit CRC)	Interpacket gap (IPG)
Length (octets)	7	1	6	6	(4)	2	42–1500[c]	4	12
Layer 2 Ethernet frame	(not part of the frame)		← 64–1522 octets →						(not part of the frame)
Layer 1 Ethernet packet & IPG	← 72–1530 octets →								← 12 octets →

Figure 4.1.2: 802.3 Ethernet packet and frame structure. (Ethernet frame - Wikipedia)

Accessibility Note

802.3 Ethernet packet and frame structure showing the number of octets in each component:

- Preamble: 7 octets

- Start Frame Delimiter (SFD): 1 octet
- Frame: 64 to 1522 octets
  - Destination MAC Address: 6 octets
  - Source MAC Address: 6 octets
  - Ethertype or length: 2 octets
  - Payload: 42-1500 octets
  - Frame Check Sequence (FCS): 4 octets

Ethernet II framing

Ethernet II framing defines the two-octet Ethertype field in an Ethernet frame, that is located right after the destination and source MAC addresses. It identifies an upper layer protocol encapsulated by the frame data.

For example:

- an EtherType value of 0x0800 indicates that the frame contains an IPv4 datagram (IPv4 supports 4 billion (2^{32}) addresses with a 32-bit address).
- 0x0806 indicates an Address Resolution Protocol (ARP) datagram,
- 0x86DD indicates an IPv6 datagram.[4] IPv6 uses 128-bit addresses, supporting an address space of 2^{128}, or 10^{38}

As this industry-developed standard went through a formal Institute of Electrical and Electronics Engineers (IEEE) standardization process, the EtherType field was changed to a (data) length field in the new 802.3 standard.

Since the packet receiverl needs to know how to interpret the frame, the standard required an IEEE 802.2 header to follow the length and specify the type. IEEE 802.2 defines Logical Link Control (LLC) as the upper portion of the data link layer of the OSI model.[5]

In the 802.3x-1997 standard, and later versions of the 802.3 standard, formally approved of both types of framing. Ethernet II framing is the most common in Ethernet local area networks, due to its simplicity and lower overhead.[4]

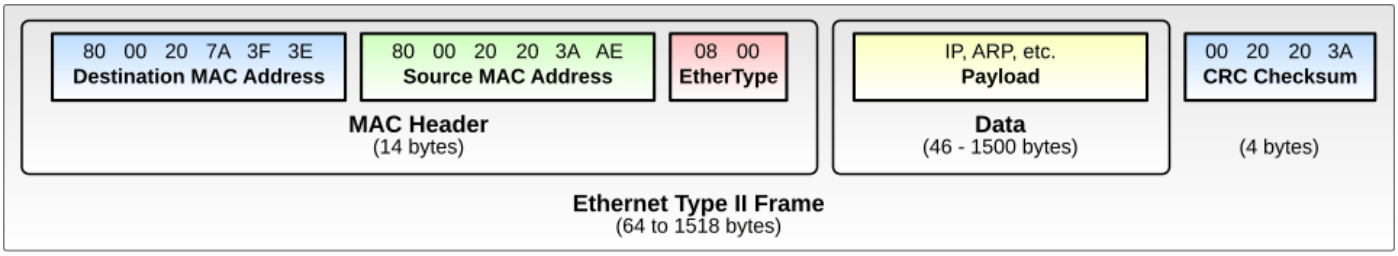


Figure 4.1.3: Ethernet Type II frame. (Ethernet frame - Wikipedia[4] in the public domain)

Accessibility Note

An Ethernet Type II frame is structured as follows:

- MAC Header (14 bytes)
  - Destination MAC Address
  - Source MAC Address
  - Ethertype
- Data/Payload (46-1500 bytes)
- CRC Checksum (4 bytes)

Location in a network

Access Switches

Access switches reside at the edge of a network, connecting WiFi access points, IP telephones sitting on the desks of employees, Ethernet-cabled equipment such as servers, security cameras, and so on. Some of the connections may require support for Power over Ethernet (PoE) in which the Ethernet cable provides power from the switch to the connected device, supporting 15, 30, or 60 watts.

Access switches often are placed in wiring closets.

Aggregation Switches

Aggregation switches serve to connect access switches, aggregating traffic from access switches for transport to the network core switch or switches. The up-link speeds of aggregation switches tend to be higher. Additionally, they have more Layer 3 routing functionality.

Core Switch

Core switches are connected to aggregation switches and would be connected to a router for connectivity to a Wide Area Network (WAN) and the internet.

Switch functions

A network switch needs to learn MAC addresses. As part of the learning process, it might store the addresses in a pice of specialized hardware called a Content Addressable Memory (CAM).

The switch has to forward packets to the specified destination, i.e, the Destination MAC address at Layer 2 or the IP address in a Layer 3 network. It may forward the packet to the next hop on its way to its ultimate destination.

More advanced switches might incorporate security features such as Virtual Private Networks and encryption.

References

1. Network switch - Wikipedia. Accessed February 4, 2026.
2. Networking hardware - Wikipedia. Accessed February 4, 2026.
3. Packet switching - Wikipedia. Accessed February 4, 2026.
4. Ethernet frame - Wikipedia. Accessed February 4, 2026.
5. IEEE 802.2 - Wikipedia. Accessed February 15, 2026.
6. Ethernet packet image.

This page titled 4.1: Introduction is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

4.2: Types of Network Switches

Networking equipment is typically mounted in standard dimension frames or enclosures that are called racks. An example is the 19-inch rack that is 19 inches (482.6mm) wide. The equipment that is placed in and secured in racks is called "rack-mountable." The equipment has side edges that have holes that allow for fastening the equipment to the rack using bolts or screws.

The height of rack-mountable equipment is standardized in rack units (RU) of 1.75 inches. So one might say, "we are designing a 1 RU switch." A standard rack is 42 RU tall.

Rack-mounted switches consist of the following basic types:

- stand-alone units,
- stackable switches, or
- large chassis systems with swap-able line cards types.[3]

Standalone Switch

A standalone switch operates independently as a single system by itself, hence the name standalone. In this mode, it does not operate in a stack as a stackable switch that operates as a grouped with other similar units.

It could be a stackable switch that is not operating in a stack, so in this unstacked mode, it is operating standalone. The standalone switch also could be a switch that has not been designed to operate in a stack.

Stackable Switch[1]

A **s**tackable switch is a switch that can operate in standalone mode (described above), but it can also be set up to operate as a stack together with one or more other switches. This group of switches, when operating in stack mode, appears a single switch but whose port count is the sum of the port counts of the stacked switches. There typically is hardware that makes the stacking connection among the members of the stack called a stacking backplane.

Definition: Stacking backplane

Stacking backplane refers to the connections between stacked units, and the bandwidth of that connection. Switches that have primarily 10Mbps ports would have at minimum 1 Gbps connections for its stacking backplane. Switches that primarily have 1 Gbps ports would have at minimum 10 Gbps connections.

The term "stack" refers to the group of switches that have been set up in this manner per the hardware supplier's instructions.

The common characteristic of a stack acting as a single switch is that there is a single IP Address for remote administration or management (by a network administrator) of the stack as a whole, not an IP address for the administration or management of each individual switch in the stack.

Stackable switches are generally one to two rack unit (RU) in height, with a fixed set of data ports on the front panel. Depending on the data port, it accepts either a fiber optic or copper connection. Some models have slots for optional slide-in modules to add ports or features to the base stackable unit. The most common stackable switches contain 24-ports or 48-ports.

Some stackable switches when operating as a stack can offer resiliency and redundancy features. For example, multiple switches can bypass a “down” switch in a stack, allowing the remaining units to continue to function as a stack even with a failed or removed unit.

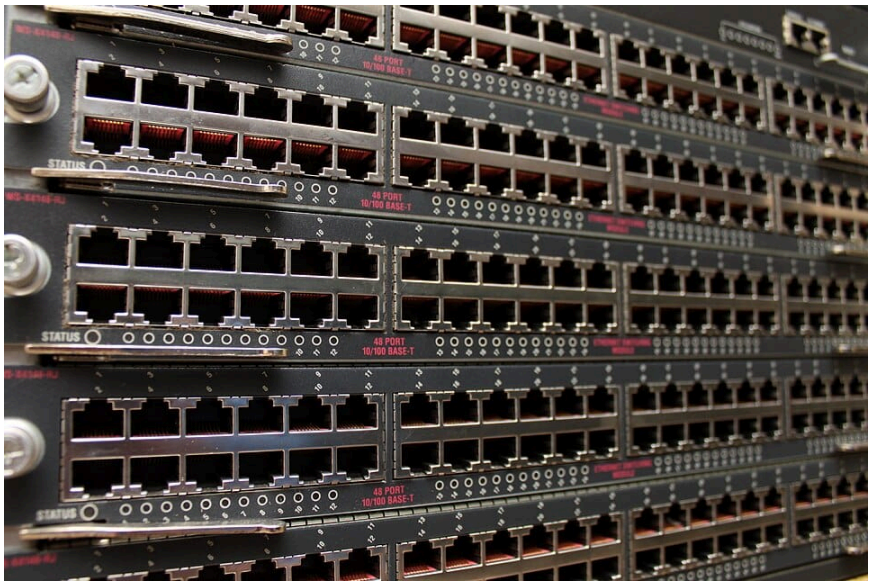


Figure 4.2.1: RJ-45 Ethernet Ports on Cisco Catalyst 4006 Switches. ([4])

Modular Switch[2]

A modular switch or chassis switch can be configured using field-replaceable units (FRUs). These units are called I/O blades (or just blades) or line cards. These line cards can add more ports, bandwidth, I/O connection types, and other capabilities to a switch.

These line cards can be heterogenous (of different types), and this allows for a network based on multiple different protocols and cable types, thereby providing flexibility to network designers and administrators.

Line cards can typically be configured in a parallel or failover configuration, which can allow for higher bandwidth, or redundancy in the event of a failure.

Modular switches also typically support hot-swapping of switch modules. This can be very important in managing system availability. If a line card fails, it can be removed from the system enclosure and a replacement line card can be inserted in its place without powering down the entire system. In a standalone switch, the configuration is fixed so one cannot hot-swap the main board from the enclosure.

Other system elements that can be hot-swapped include:

- Power supplies
- Fan trays (used for cooling the system)
- Management or supervisor modules.

Management or supervisor modules are typically configures with two in a system to offer resilience. These modules run the system's network operating system and provide an interface for managing the system by a network administrator.

Modular switches also support additional line cards which can provide new functions to the switch that would previously have been unavailable, such security functionality.

Compared with a modular chassis switch, stackable switches have the following disadvantages:

1. For a customer site needing numerous ports, a modular chassis may cost less. With stackable switching, each unit in a stack has its own mechanical enclosure and at minimum a single power supply. With modular switches, there is one mechanical enclosure and one set of power supplies.
2. High-end modular switches have high-resiliency and high-redundancy features not available in all stackable architectures.
3. There may bel overhead when sending stacking data between switches. For example, some stacking protocols add additional

headers to frames.

The design of a modular switch is much more complex than designing a standalone 1 or 2 RU standalone switch. In a modular switch, as mentioned before, there are multiple types of line cards that typically need to be designed. The two management modules and line cards need to be connected. This is done via a separate board not visible to the customer called the backplane. It has specialized high speed connectors with multiple connector pins for the management modules and for the line cards. In order to safely insert management modules and line cards into the system to connect with the backplane, the boards typically have guide pins to help ensure proper card alignment with the backplane. In addition, there may be a fair amount of insertion force required to make the board connection with the backplane, so specialized "ejector" mechanisms need to be designed to provide the necessary leverage.

To distribute power from the Power Supply Units (PSUs) the system design team needs to design a power backplane or bus bar system.

References

1. Stackable switch - Wikipedia. Accessed February 4, 2026.
2. Modular switch - Wikipedia. Accessed February 4, 2026.
3. Network switch - Wikipedia. Accessed February 4, 2026.
4. Ben Franske. Wikimedia. Figure of RJ-45 Ethernet Ports on a Cisco Catalyst 4006 switch. August 29, 2009. This file is licensed under the Creative Commons(opens in new window) Attribution-Share Alike 4.0 International(opens in new window), 3.0 Unported(opens in new window), 2.5 Generic(opens in new window), 2.0 Generic(opens in new window) and 1.0 Generic(opens in new window) license.
5. HPE Aruba Networking CX 6300L Switch Series, Accessed February 12, 2026.

This page titled 4.2: Types of Network Switches is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

4.3: System Scope

Now that we have defined what a network switch is and have described the types of network switches, we are going to look at the scope that this case study is going to address.

For simplicity, we are not going to address the development of a large modular switch. As mentioned in the prior section, its scope is complex, requiring the development of:

- Multiple line card types
- A management or supervisor board that typically would be used in a redundant configuration (two management boards)
- A backplane
- A power backplane or other power distribution system.
- A large mechanical enclosure to house the management modules, line cards, backplane, power distribution system, redundant hot-swappable fan trays and power supply units. Typically a complex finite element cooling analysis needs to be performed to ensure adequate air flow, especially to hot spots (components that rise to a high temperature when operating). This may require designing special baffles to intelligently direct airflow.

Instead, we are addressing the smaller pizza-box sized 1-2 RU standalone switches with fixed I/O port configurations.

Since this CIS 79 Managing Technology Projects course is part of the De Anza College project management certificate and degree curriculum, we are going to look at this development effort through the lens of the project manager assigned to the project rather than through the lens of a system design team chartered with the engineering effort to develop the system.

Network switches can be complex systems, comprised of hardware and software components. A project manager's scope might be to handle the entire system, just the hardware, or just the software. There is also a middle ground in which the project manager handles the hardware as well as a piece of the platform software, the software that provides support for the underlying hardware, especially the networking Application Specific Integrated Circuits (ASICs). The hardware project manager would be working closely with the software project manager, often called a software release project manager.

Since a lot of the De Anza curriculum is software centric, for balance, we will concentrate on addressing the hardware elements. This will provide better preparation to be "industry-ready."

Design choices

In De Anza College's CIS 43 Systems Design course (which one of the authors teaches), there are design alternatives and various "make versus buy" decisions that need to be addressed. How one approaches these decisions will be guided by the requirements filed by the product manager regarding the required system capabilities.

Network Application Specific Integrated Circuits (ASICs) do the heavy-lifting of the needed switching functionality such as packet switching, forwarding, routing, quality of Service (QoS) and some security protocols. Some networking companies maintain their own ASICs development teams to retain their edge in the competitive marketplace. There are companies that develop network ASICs that they sell to network switch companies for incorporation into their product offerings. These third party or merchant silicon ASICs come with an SDK (Software Development Kit) that provides software support for the ASICs. The SDK needs to be incorporated into the network switch vendor's network operating system. The SDK can come with a Command Line Interface (CLI) that provides an almost ready to use system, albeit usually at the low end.

A classic "make versus buy" decision is who does the major work on the hardware development effort. Should it be done using the network switch company's hardware design team, or could it be done by an external ODM (Original Design Manufacturing or Manufacturer)? An ODM can not only perform the design work, but also do the manufacturing. There may be a catalog from the ODM of already developed products or very similar ones from which to choose.

Another design choice (albeit a software one) is what CLI (Command Line Interface) to offer the user for interacting with the switch. One can use a proprietary interface (your own commands and syntax) or consider offering a CISCO-compatible CLI that many networking professionals are used to using.

There will be more on this in subsequent sections.

This page titled 4.3: System Scope is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

4.4: Hardware System Design

This section provides a brief overview of the hardware design teams, typically a networking ASIC development team and a system design team.

ASIC Design Team

As mentioned earlier Network ASICs can be developed in-house (custom ASICs) or be procured from several merchant silicon vendors. The ASIC chipsets may have specific target markets: Enterprise, hyperscalers, or Service Providers.

Systems Design Team

An in-house system design team is usually of modest size, much smaller in general than the software team that is working on the project. The areas represented typically include:

- **Chassis design.** The chassis design is often done in concert with a senior system designer by mechanical engineers who prepare 3D models of the desired enclosure which will be fabricated using sheet metal, with necessary cutouts for front-panel I/O interfaces, removable power supplies, and removable fan trays.
- **Printed Circuit Board (PCB) design**. A PCB designer does the logical design of the board, connecting the pins of component to other components. If an outside vendor is used to perform the physical routing, the board designer would supervise that.
- **Signal integrity engineering**. This group ensures that signals remain usable as they traverse the full length of traces from one component to another.
- **Power supply design**. PSU designers are involved with working with outside vendors on new PSUs, validating off the shelf PSUs, or working on custom changes to existing PSUs.
- **Compliance engineering.** These professionals ensure compliance to EMC standards by going requisite testing in-house or working with outside vendors.
- **Testing.** This includes board bring-up and HALT testing. It may include people who write diagnostics, although they may reside in the software group.

If an ODM is being used to do an original design or port a previously existing design, a much smaller hardware team will likely be in place to oversee the work of the ODM. The team will likely be a group of experienced former hardware designers with various areas of expertise.

This page titled 4.4: Hardware System Design is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

4.4.1: Network ASICs

As mentioned earlier Network ASICs can be developed in-house (custom ASICs) or be procured from several merchant silicon vendors. The ASIC chipsets may have specific target markets: Enterprise, hyperscalers, or Service Providers.

Handling ASIC revisions

Whether the ASICs are developed in-house or are from merchant silicon vendors, they need to be tested on boards subjected to network traffic loads.

Despite the best efforts of the ASIC designers doing their testing in simulation before the chip tapes out (is sent to the ASIC foundry for fabrication), it is possible that errors are detected during these on-board tests that will require a new revision of the ASIC, called a new "spin." Depending on the nature of the error, it sometimes can be corrected with a metal layer spin, or in the worst case, an all-layer spin which would come at considerable expense to redo the chip masks.

Integration into schedules

When preparing the project schedule, the project manager must make risk assessments of the number of prototype board design revisions (to be discussed later) as well as the number of prototype ASIC spins. These assessments can include such criteria as the past track record of the design team, an assessment of the technology risk (are many new features being added and are the new features described in newly published specifications).

Testing

In addition to specific function testing, there typically is a part of the network switch developer's System Quality Assurance lab that tests maximum configurations, such as all ports passing full traffic at maximum rated speed.

Process Lot/Corner Lot Testing

The semiconductor manufacturing process that produces the network ASICs can have variability. In order to assess whether these process variations could potentially impact one's design, it is useful to consider ordering corner lots of the ASICs from the ASIC manufacturer. Corner lots also go by the following names: process lots and skewed lots.

The semiconductor foundry will, upon request, make chip wafers that they purposely skew to different process corners by altering the doping of NMOS and PMOS devices for the purpose of this production process variation immunity testing. The five "corners" that are skewed have a two-letter designation, where the first letter refers to the NMOS device, and the second refers to the PMOS device. There are 5 corners:[1]

- FF (fast fast), meaning both the P and N devices are purposely skewed to the fast corner.
- SF (slow fast)
- SS (slow slow)
- FS (fast slow)
- TT (typical typical), i.e., normal process is used.

Note

As the project manager, you may be involved in ordering the corner lot ASICs from the ASIC manufacturer and scheduling the placement of the corner lot parts on boards in conjunction with corner lot parts for other critical components and scheduling the testing activity. The timing would be not long after ASIC tape-out or before the ASIC is released to production.

> In order to build these boards with the skewed components, one must exercise special care in tracking which parts come from which corner and have a willing, capable prototype board manufacturer to handle these special cases.
>
> There should be money in the project for this activity. The testing resource is likely a signal integrity engineer.

References

1. Understanding semiconductor Process Lots (Corner Lots). Accessed February 16, 2026. This article provides an excellent overview of corner lots.

This page titled 4.4.1: Network ASICs is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

4.4.2: Chassis Design

The chassis design is often done in concert with a lead system designer by mechanical engineers.

The mechanical engineering team prepares 3D models of the desired enclosure which will be fabricated using sheet metal usually by an outside vendor. The design model contains the necessary metal openings for front-panel I/O interfaces, removable power supplies, and removable fan trays.

The desired direction of air flow is usually specified by the product manager, based on where the product will be placed in racks. Air-flow direction can include front to back (pull in air from the front panel with the I/O interfaces) or side to side.

Typically a plastic overlay is placed on the front panel around the I/O module cages, labeling the types of interfaces and numbering them.

The mechanical design team needs to be keenly aware of EMC considerations and to proactively take countermeasures. During the EMC testing of the first revision of the chassis and board mounted within it, traffic flow testing can reveal electronic immunity issues or emissions issues. This may result in a second revision of the chassis design.

The chassis design team must also arrange to paint the chassis according to the current prevailing look and feel and industrial design guidelines required by Marketing as reflected in the product manager's requirements.

Also on the chassis is an emblem indicating the compliance status of the system (e.g., CE mark, FCC rating).

The team works with the technical publications team to ensure that correct documentation for the system is prepared.

This page titled 4.4.2: Chassis Design is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

4.4.3: Board Design

A printed circuit board (PCB) is a laminated sandwich structure of conductive and insulating layers, each with a pattern of traces, planes and other features (similar to wires on a flat surface) etched from one or more sheet layers of copper laminated onto or between sheet layers of a non-conductive substrate.[3]

PCBs are used to connect components to one another. Electrical components may be fixed to conductive pads on the outer layers of the board (top side, bottom side), generally by a manufacturing soldering process.

The soldering process does two things:

- it electrically connects the components to the printed circuit board and
- it mechanically fastens the components to the printed circuit board.

Another PCB manufacturing process adds vias (discusses later in this section). Vias are metal-lined drilled holes that enable electrical interconnections between conductive layers of the PCB.[3]

PCBs can be single-sided (one copper layer), double-sided (two copper layers on both sides of one substrate layer), or multi-layer (stacked layers of substrate with copper plating sandwiched between each and on the outside layers). Multi-layer PCBs provide much higher component density, because circuit traces on the inner layers would otherwise take up surface space between components. The rise in popularity of multilayer PCBs with more than two, and especially with more than four, copper planes was concurrent with the adoption of surface mount technology. However, multilayer PCBs make repair, analysis, and field modification of circuits much more difficult and usually impractical.[3]

Definition: Board stack-up

The term "stack-up" refers to the organized arrangement of the various laminated layers of the PCB as determined by the systems design team.

Logical Design

The printed circuit board designer's primary task is to properly connect components to one another.

Component selection

Before the board designer can connect any components, the board designer needs to select the components that will meet the functional requirements filed by the product manager. This will include the networking ASICs, CPUs, memory devices, transceivers, DC-DC converters, resistors, capacitors, inductors, and other components.

Once the component has been selected, the system designer needs to ensure that it is added to the design team's database of components. A part number is assigned to it with an Approved Vendor List (AVL) of suppliers of that component.

The board designer needs to create "symbols" for new components, including definitions and names for all the component's pins.

Connecting components

Once all the component symbols have been created, the board designer can connect component pins logically, not worrying about the physical routing connection in the logical design phase.

Component power requirements

The data sheets for components specify the voltage levels required and provide estimates of power consumption when

operational.

The board designer, sometimes with help from an interested project manager, can use a spreadsheet to keep track of all the necessary voltage levels and the expected power consumption at each voltage level.

Some systems will see 48V DC supplied to the system. In other cases, the system may have PSUs that are operating off 110/220V that can supply different voltages to the board. The board designer can use on-board DC-DC converters to get the needed voltages and include enough DC-DC converters to meet the power budget.

Mounting the board in the chassis

The board designer designates certain physical spaces as "keep outs" so no routing takes place there. For example, this is done to accommodate the placement of stand-offs that are secured to the chassis sheet metal and which go through the entire width of the board for secure tightening.

Physical Design/Layout

The main tasks here is to take the logical design that logically connected components discussed previously and transform it into a physical design that routes the connections called traces through the various layers of the board. There are design automation tools to ease the pain of dealing with this daunting task. The work is often done by outside services that specialize in physical layout.

Vias

A **via** (Latin, 'path' or 'way') is an electrical connection between two or more metal layers of a PCB. A via is a small drilled hole that goes through two or more adjacent layers; the hole is plated with metal (often copper) that forms an electrical connection through the insulating layers.[1] The holes are drilled as part of the PCM manufacturing process.

Vias are an important concern in PCB manufacturing. As vertical structures crossing multiple layers, they are specified differently from most of the design, which increases the chance for errors. They place the strictest demands on registration (how closely aligned different layers are). They are manufactured with different tooling from other features -- tooling that typically has looser tolerances. If either the hole or any layer is slightly out of place, the wrong electrical connections could be made. This may not be visible from the surface of the PCB. After the hole is drilled, it must also be lined with conductive material, as opposed to simply leaving conductive material in place on copper layers. Even an initially good board may develop problems later because the via reacts to heat differently from the substrate around it. Vias also represent a discontinuity in the electrical impedance, which can cause problems for signal integrity[1] that will be discussed in a subsequent section of this textbook's chapter.

In PCB design, a via consists of two pads in corresponding positions on different copper layers of the board, that are electrically connected by a hole through the board. The hole is made conductive by electroplating, or is lined with a tube or a rivet. High-density multilayer PCBs may have microvias:

- **blind vias** are exposed only on one side of the board,
- **buried vias** connect internal layers without being exposed on either surface
- A **through-hole via** is a drilled and copper-plated hole that connects all layers of a multilayer PCB, from the top layer to the bottom layer, providing both electrical connectivity and mechanical strength.[1] It allows signals and power to travel vertically through the PCB stackup.[1]

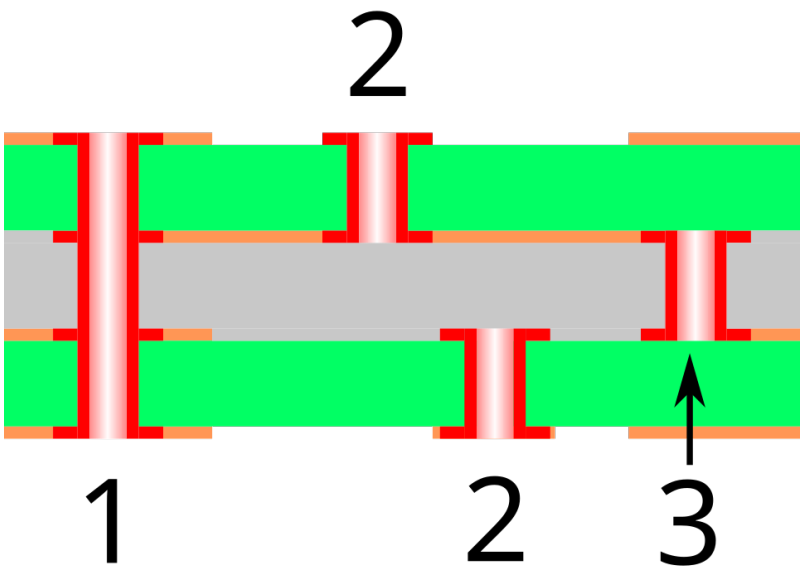


Figure 4.4.3.1: PCB Via Types. (M Adler, CC-BY 3.0)[2]

Accesssibility Note

Figure 4.4.3.1 illustrates several different types of printed circuit board vias:

1. Through hole.
2. Blind via.
3. Buried via.

The gray and green layers are nonconducting, while the thin orange layers and red vias are conductive.

PCB fabrication

Once the physical design has completed, the work is encapsulated in a BRD file.

One typically will get quotes for manufacturing the board from multiple board fabricator vendors. The quotations typically provide a table of costs for the specified quantity, gated by how fast you want the boards fabricated. The faster you want the boards, the more expensive they will be. Doing a board in 14 days would be called a "14-day turn."

This BRD file is transmitted to a PCB fabricator to manufacture the bare multilayer printed circuit board. The bare board is often referred to as the "PCB fab".

PCB assembly (PCBA)

Then board designer working with a project manager structures the Bill of Materials (BOM) for the board. It includes all the components that are needed along with their quantities. It also specifies the Approved Vendor List or suppliers for each component.

In order to place components on a board, the PCB needs to be procured as discussed, as well as all the components to be placed on the board.

During the engineering prototyping phase, the components may be procured by a contract manufacturer that specializes in prototypes. The system design team may do the component sourcing, but this is rare. When new components are being used such as merchant silicon ASICs, they may be delivered to the system design team or project manager for transfer to the prototype board assembler.

The prototype board assembler puts the components to be placed on the board in machines for soldering on the board. The assembly process requires placing components on the top side of the board as well as on the bottom side of the board.

After the components have been placed on both sides of the board, various tests called for by the system design team (e.g., flying probe) are conducted before boards can be delivered to the system design team for testing.

References

1. Via (electronics) - Wikipedia. Accessed February 7, 2026.
2. Via (electronics) - Wikipedia. Accessed February 7, 2026. Figure 4.4.3.1 showing different types of PCB vias.

This file is licensed under the Creative Commons Attribution 3.0 Unported license.

You are free:
- **to share** – to copy, distribute and transmit the work
- **to remix** – to adapt the work

Under the following conditions:
- **attribution** – You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

3. Printed circuit board - Wikipedia. Accessed February 7, 2027.

This page titled 4.4.3: Board Design is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

4.4.4: Signal integrity

Signal Integrity (SI) refers to the quality of electrical signals as they travel through traces, vias, and components on a printed circuit board. Ensuring good signal integrity is critical for high-speed and high-frequency designs. Poor signal quality can lead to data errors, signal distortion, and system malfunction.[1]

All signals are subject to effects such as noise, distortion, and loss. Over short distances and at low bit rates, a simple conductor can transmit this with sufficient fidelity. However, at high bit rates and over longer distances or through various mediums, various effects can degrade the electrical signal to the point where errors occur and the system or device fails. Signal integrity engineering is the task of analyzing and mitigating these effects.[1] The system design team member who has responsibility for this analysis usually has a job title of Signal Integrity Engineer.

On printed circuit boards, signal integrity became a serious concern when the transition (rise and fall) times of signals started to become comparable to the propagation time across the board. This typically happens when system speeds exceed a few tens of MHz. At first, only a few of the most important, or highest speed, signals needed detailed analysis or design. As speeds increased, a larger and larger fraction of signals needed SI analysis and design practices. In modern (> 100 MHz) circuit designs, essentially all signals must be designed with SI in mind.[1]

Factors Affecting Signal Integrity

The following is a list of factors that affect signal integrity:[1]

- Reflection
- Crosstalk
- Transmission Line Effects
- Impedance Mismatch
- Electromagnetic Interference (EMI)
- Power Integrity (PI)
- Rise and Fall Time

A PCB designer removes signal integrity problems through the use of the following mitigation techniques:[1]

- Placing a solid reference plane adjacent to the signal traces to control crosstalk
- Controlling the trace width spacing to the reference plane to create consistent trace impedance
- Using terminations to control ringing
- Route traces perpendicular on adjacent board layers to reduce crosstalk
- Increasing spacing between traces to reduce crosstalk
- Providing sufficient ground (and power) connections to limit "ground bounce" (this sub-discipline of signal integrity is sometimes called out separately as power integrity)
- Distributing power with solid plane layers to limit power supply noise
- Adding a pre-emphasis filter to the transmitter driving cell

Cross-talk

Definition: Crosstalk

Crosstalk (XT) is a phenomenon by which a signal transmitted on one circuit or channel of a transmission system creates an undesired effect in another circuit or channel. Crosstalk is usually caused by undesired capacitive, inductive, or conductive coupling from one circuit or channel to another.[3]

Eye-patterns

If you were to go to a hardware bring-up lab and observe a signal integrity engineer at work, you likely would encounter the engineer analyzing an eye pattern on an oscilloscope.

An eye pattern, or eye diagram, is a display on an oscilloscope screen in which a digital signal from a receiver is repetitively sampled and applied to the vertical input (*y-axis*), while the data rate is used to trigger the horizontal sweep (*x-axis*). It is called an eye pattern because, for several types of coding, the pattern looks like a series of eyes between a pair of rails. The eye pattern a tool for the evaluation of the combined effects of channel noise, dispersion and inter-symbol interference on the performance of a baseband pulse-transmission system.[5]

Definition: Symbol

A symbol is a waveform, a state or a significant condition of the communication channel that *persists*, for a fixed period of time.[6]

Definition: Inter-symbol interference

inter-symbol interference (**ISI**) is a form of distortion of a signal in which one symbol (waveform) interferes with subsequent symbols (waveforms). This is an unwanted phenomenon as the previous symbols (waveforms) have a similar effect as noise, thus making the communication less reliable.[7]

Several system performance measurements can be derived by analyzing the display. If the signals are too long, too short, poorly synchronized with the system clock, too high, too low, too noisy, or too slow to change, or have too much undershoot or overshoot, this can be observed from the eye diagram. An open eye pattern corresponds to minimal signal distortion. Distortion of the signal waveform due to inter-symbol interference and noise appears as closure of the eye pattern.[5]

Definition: Noise

Noise is an error or undesired random disturbance of a useful information signal in a communication channel.[8]

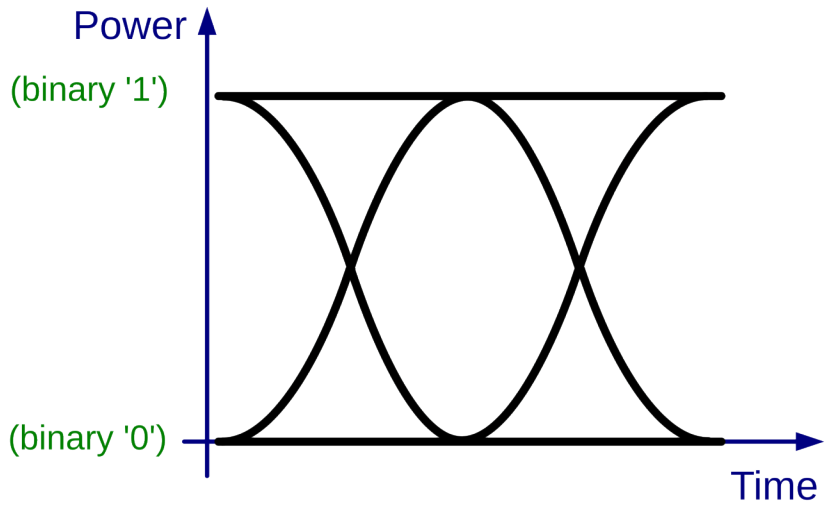


Figure 4.4.4.1: Example of an Eye Diagram. (By Gmoose1 - Own work, CC BY-SA 3.0[4])

Accessibility Note

Graphical eye pattern showing an example of two power levels in an OOK modulation scheme. Constant binary 1 and 0 levels are shown, as well as transitions from 0 to 1, 1 to 0, 0 to 1 to 0, and 1 to 0 to 1 in a sinusoidal pattern.

Manhattan lengths of signal traces

Manhattan length (or distance), refers to the total distance traveled by a trace on a printed circuit board when following a grid pattern (like streets on Manhattan Island, New York), calculated by summing the horizontal and vertical distances between two points.

Manhattan length helps printed circuit board designers to optimize trace routing and maintain good signal integrity.

Pre-emphasis settings

Pre-emphasis is a technique to protect against anticipated noise and loss. The idea is to boost (and hence distort) the frequency range that is most susceptible to noise and loss beforehand, so that after a noisy and lossy process more information can be recovered from that frequency range. Removal of the distortion caused by pre-emphasis is called de-emphasis, making the output accurately reproduce the original input.[2]

In serial data transmission, emphasis is used to improve signal quality at the output of a communication channel. In transmitting signals at high data rates, the transmission medium may introduce distortions, so emphasis is used to distort the transmitted signal to correct for this distortion. This produces a received signal that more closely resembles the original or desired signal, allowing the use of higher data rates or producing fewer bit errors. Most real world channels have loss that increases with frequency (effectively a low pass filter), so emphasis needs to invert this effect (functioning as a high pass filter).[2]

Emphasis can be implemented either by boosting high frequencies (pre-emphasis, increasing the amplitude of transition bits) or attenuating low frequencies (de-emphasis, reducing the amplitude of non-transition bits). Both have the same net effect of producing a flatter system frequency response. De-emphasis is typically more convenient to do in real circuits since it only requires attenuation rather than amplification.[2]

In practice, after a signal integrity engineer conducts the requisite analysis, a set of needed pre-emphasis settings is calculated and saved in a file. This file is then referenced when the system is operational to ensure that the intended pre-emphasis actions are performed.

References

1. Signal integrity - Wikipedia. Accessed February 16, 2026.
2. Emphasis (telecommunications) - Wikipedia. Accessed February 16, 2026. Text is available under the Creative Commons Attribution-ShareAlike 4.0 License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.
3. Crosstalk - Wikipedia. Accessed February 16, 2026.
4. By Gmoose1 - Own work, CC BY-SA 3.0, Wikimedia Eye pattern diagram. [Eye pattern diagram in Wikipedia-Eye pattern.] Accessed February 16, 2026.
5. Eye pattern - Wikipedia. Accessed February 17, 2026.
6. Symbol rate - Wikipedia. Accessed February 17, 2026.
7. Intersymbol interference - Wikipedia. Accessed February 17, 2026.
8. Noise (electronics) - Wikipedia. Accessed February 17, 2026.

This page titled 4.4.4: Signal integrity is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

4.4.5: Electro-magnetic Regulatory Compliance

Electromagnetic compatibility

Electromagnetic compatibility (EMC) is the ability of electrical equipment and systems to function acceptably in their electromagnetic environment, by limiting the unintentional generation, propagation and reception of electromagnetic energy which may cause unwanted effects such as electromagnetic interference (EMI) or even physical damage to operational equipment. The goal of EMC is the correct operation of different equipment in a common electromagnetic environment.[1]

Electromagnetic compatibility covers the following three classes of issues:[1]

1. ***Emission*** is the generation of electromagnetic energy, whether deliberate or accidental, by some source and its release into the environment. EMC studies the unwanted emissions and the countermeasures which may be taken in order to reduce unwanted emissions.
2. The second class, ***susceptibility***, is the tendency of electrical equipment, referred to as the victim, to malfunction or break down in the presence of unwanted emissions, which are known as radio-frequency interference (RFI). ***Immunity*** is the opposite of susceptibility, being the ability of equipment to function correctly in the presence of RFI.
3. A third class studied is ***coupling***, which is the mechanism by which emitted interference reaches the victim.

Interference mitigation and hence electromagnetic compatibility may be achieved by addressing any or all of these issues, i.e., quieting the sources of interference, inhibiting coupling paths and/or hardening the potential victims. In practice, many of the engineering techniques used, such as grounding and shielding, apply to all three issues.

Definition: Electromagnetic Interference

"Electromagnetic interference" (EMI) is defined as the "*degradation in the performance of equipment or transmission channel or a system caused by an electromagnetic disturbance*" (IEV 161-01-06) while "electromagnetic disturbance" is defined as "*an electromagnetic phenomenon that can degrade the performance of a device, equipment or system, or adversely affect living or inert matter* (IEV 161-01-05). The terms "electromagnetic disturbance" and "electromagnetic interference" designate respectively the cause and the effect[1]

EMC ensures the correct operation, in the same electromagnetic environment, of different equipment items which use or respond to electromagnetic phenomena, and the avoidance of any interference. Another way of saying this is that EMC is the control of EMI so that unwanted effects are prevented.

Besides understanding these phenomena, EMC also addresses the countermeasures, such as control regimes, design and measurement, which should be taken by the hardware systems design team in order to prevent emissions from causing any adverse effect.

Coupling

When a source emits interference, it follows a route to the victim known as the coupling path.

There are four basic coupling mechanisms:

1. conductive,
2. capacitive,
3. magnetic or inductive, and
4. radiative.

Any coupling path can be broken down into one or more of these coupling mechanisms working together.

- **Conductive coupling** occurs when the coupling path between the source and victim is formed by direct electrical contact

with a conducting body.

- **Capacitive coupling** occurs when a varying electrical field exists between two adjacent conductors, inducing a change in voltage on the receiving conductor.
- **Inductive coupling** or **magnetic coupling** occurs when a varying magnetic field exists between two parallel conductors, inducing a change in voltage along the receiving conductor.
- **Radiative coupling** or **electromagnetic coupling** occurs when source and victim are separated by a large distance. Source and victim act as radio antennas: the source emits or radiates an electromagnetic wave which propagates across the space in between and is picked up or received by the victim.

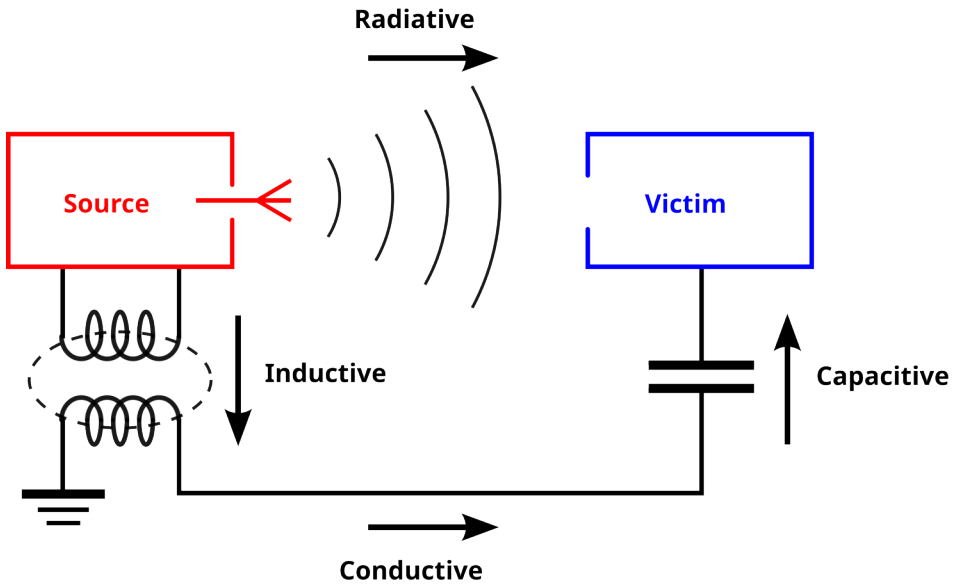


Figure 4.4.5.1: The four EMI Coupling modes. (By Steelpillow - Own work, CC BY-SA 3.0,)[5]

Accessibility Note

Figure 4.4.5.1: There are four coupling mechanisms:

1. conductive (shown as a wire),
2. capacitive (shown as a capacitor),
3. magnetic or inductive (shown as coupled indictive coils), and
4. radiative (shown as a radiating antenna).

Controlling EMI

The control of electromagnetic interference (EMI) and assurance of EMC comprises a series of related disciplines:

- Characterizing the threat.
- Setting standards for emission and susceptibility levels.
- Design for standards compliance.
- Testing for standards compliance.

Grounding and shielding aim to reduce emissions or divert EMI away from the victim by providing an alternative, low-impedance path. Techniques include:

- Grounding or earthing schemes such as *ground planes* for RF. The scheme must also satisfy safety regulations.
- Shielded cables, where the signal wires are surrounded by an outer conductive layer that is grounded at one or both ends.
- Shielded housings. A conductive metal housing will act as an interference shield. In order to access the interior, such a housing is typically made in sections (such as a box and lid); an RF gasket may be used at the joints to reduce the amount of interference that leaks through. RF gaskets come in various types. A plain metal gasket may be either braided wire or a

flat strip slotted to create many springy "fingers". Where a waterproof seal is required, a flexible elastomeric base may be impregnated with chopped metal fibers dispersed into the interior or long metal fibers covering the surface or both.

EMC Testing

Testing is required to confirm that a particular device meets the required standards. It is divided broadly into:

- emissions testing and
- susceptibility testing.

Open-area test sites, or OATS, are the reference sites in most standards. They are especially useful for emissions testing of large equipment systems.

However, RF testing of a physical prototype (such as our network switch) is most often carried out indoors, in a specialized EMC test chamber. Types of test chambers include:

- anechoic,
- reverberation, and
- gigahertz transverse electromagnetic (GTEM) cell

It is important that the test equipment, including the test chamber or site and any software used, be properly calibrated and maintained. Typically, a given run of tests for a particular piece of equipment will require an *EMC test plan* and a follow-up *test report*. The full test program may require the production of several such documents.

> **Note**
>
> Keep these deliverables in mind when preparing the project WBS and schedule.

Emissions are typically measured for radiated field strength and, where appropriate, for conducted emissions along cables and wiring.

Inductive (magnetic) and capacitive (electric) field strengths are near-field effects and are only important if the device under test (DUT) is designed for a location close to other electrical equipment.

For conducted emissions, typical transducers include the LISN (line impedance stabilization network) or AMN (artificial mains network) and the RF current clamp.

For radiated emission measurement, antennas are used as transducers. Typical antennas specified include dipole, biconical, log-periodic, double ridged guide and conical log-spiral designs. Radiated emissions must be measured in all directions around the DUT. Specialized EMI test receivers or EMI analyzers are used for EMC compliance testing. These incorporate bandwidths and detectors as specified by international EMC standards. An EMI receiver may be based on a spectrum analyzer to measure the emission levels of the DUT across a wide band of frequencies (frequency domain), or on a tunable narrower-band device which is swept through the desired frequency range. EMI receivers along with specified transducers can often be used for both conducted and radiated emissions. Pre-selector filters may also be used to reduce the effect of strong out-of-band signals on the front-end of the receiver. Some pulse emissions are more usefully characterized using an oscilloscope to capture the pulse waveform in the time domain.

Radiated field susceptibility testing typically involves a high-powered source of RF or EM energy and a radiating antenna to direct the energy at the potential victim or device under test (DUT). Conducted voltage and current susceptibility testing typically involves a high-powered signal generator, and a current clamp or other type of transformer to inject the test signal. Transient or EMP signals are used to test the immunity of the DUT against powerline disturbances including surges, lightning strikes and switching noise. The transient pulse may be generated digitally and passed through a broadband pulse amplifier, or applied directly to the transducer from a specialized pulse generator. Electrostatic discharge testing is typically performed with a piezo spark generator.

EMC test standards

EMC testing standards attempt to standardize product EMC performance, with respect to conducted or radiated radio interference from electrical or electronic equipment, imposition of other types of disturbance on the mains supply by such equipment, and the sensitivity of such equipment to received interference.[2]

For the United States, this includes:

- FCC Part 15 regulates unlicensed radio-frequency transmissions, both intentional and unintentional.
  - FCC Part 15 Subpart A contains a general provision that "devices may not cause interference and must accept interference from other sources."
  - FCC Part 15 Subpart B US limits and methods of measurement of radio disturbance, measuring radio waves accidentally emitted from devices not specifically designed to emit radio waves ("unintentional"), both directly ("radiated") and indirectly ("conducted")
  - The rest of FCC Part 15 (subparts C through H) deal with unlicensed devices specifically designed to emit radio waves ("intentional"), such as wireless LAN, cordless telephones, low-power broadcasting, walkie-talkies, etc.
  - Conducted emissions are regulated from 150 kHz to 30 MHz, and radiated emissions are regulated from 30 MHz and up.

The standards that likely will be referenced in network switch data sheets might include the following:

- Europe:
  - EN 55032:2015 +A11:2020, Class A
  - EN 55035:2017 +A11:2020 EN 61000-3-2:2014, Class A
  - EN 61000-3-3:2013

Note

- IEC EN 61000-3-3, Electromagnetic compatibility (EMC) - Part 3-3 - Limits - Limitation of voltage changes, voltage fluctuations and flicker in public low-voltage supply systems, for equipment with rated current ≤ 16 A per phase and not subject to conditional connection.[2]

- U.S./Canada:
  - FCC CFR47 Part 15:2014, Class A (Code of Federal Regulations,Title 47). FCC Part 15 regulates unlicensed radio-frequency transmissions, both intentional and unintentional. FCC Part 15 Subpart A contains a general provision that "devices may not cause interference and must accept interference from other sources."[2]
  - ICES-003 Class A

Definition: ICES-003

ICES-003 stands for Interference-Causing Equipment Standard 003 covering Information Technology Equipment (including Digital Apparatus) for Canada.[4]

- Worldwide:
  - VCCI Class A
  - CISPR 32 Class A
  - CISPR 35:2016

Definition: VCCI

The Voluntary Control Council for Interference by Information Technology Equipment or VCCI is the Japanese organization governing radio frequency (RF) emissions (i.e. electromagnetic interference) standards.[3]

Definition: CISPR

CISPR is the acronym of **C**omité **I**nternational **S**pécial des **P**erturbations **R**adio, or the International Special Committee for Radio Protection of IEC. CISPR Standards aim to the protection of radio reception in the range 9 kHz to 400 GHz from interference caused by operation of electrical or electronic appliances and systems in the electromagnetic environment. CISPR standards cover product emission and immunity requirements as well as defining test methods and equipment.[2]

- CISPR 32, Electromagnetic Compatibility of multimedia equipment – Emission requirements.
- CISPR 35, Electromagnetic Compatibility of multimedia equipment – Immunity requirements.

NEBS

Definition: Network Equipment Building System (NEBS)

Network Equipment-Building System (NEBS) is a telecommunications term that describes the environment of a typical United States central office covering the most common set of safety, spatial and environmental design guidelines applied to telecommunications equipment in the United States. [5]

NEBS Levels

"NEBS Level 1" is a very low threshold of equipment hazards and network degradation. NEBS Level 1 addresses the personnel and equipment safety requirements of GR-63-CORE and GR-1089-CORE. Operability requirements are not enforced for NEBS Level 1 certification. It is primarily used for getting prototypes into lab trials.[5]

"NEBS Level 2" addresses equipment operability in a controlled environment (usually datacenters) that will not be subjected to environmental stress. Because of ambiguity, this level of certification is rarely (if ever) used.[5]

"NEBS Level 3" is a term from Bellcore special report, SR-3580, and means the equipment meets all of the requirements of GR-63-CORE and GR-1089-CORE. NEBS Level 3 has strict specifications for fire suppression, thermal margin testing, vibration resistance (earthquakes), airflow patterns, acoustic limits, failover and partial operational requirements (such as chassis fan failures), failure severity levels, RF emissions and tolerances, and testing/certification requirements.[5]

NEBS Project Management Considerations

The product manager will specify whether the network switch being developed needs to be NEBS-compliant and at what level. Sometimes the product manager will be content to state the the product has been designed for NEBS, but has not been tested for NEBS-compliance.

It the highest level of NEBS (Level 3) is required, the hardware design team needs to familiarize themselves with the requirements it places on the design.

When the hardware design has been finalized, a production (or near production level) set of units will be taken to an outside testing company for the NEBS testing. It comes at considerable expense. Since NEBS Level 3 has requirements for fire suppression, one unit will be damaged or destroyed as part of that testing that is usually done as one of the last testing items.

EMC Project Management Considerations

The project manager is responsible for determining how many prototype systems will be build at each prototype version or revision level and to whom they are allocated for what purposes. The initial prototypes are called "Proto 1" units.

When a "Proto 1" chassis with its system printed circuit board, power supplies, and fan trays is available as a complete system, one or more units usually are allocated to the EMC Compliance group. The EMC Compliance group conducts its testing as summarized earlier in this section, usually by running network traffic through the system. If the EMC Compliance group finds issues, they will work with the system design hardware team to find countermeasures that will be incorporated into the next revision (Proto 2 or Rev 2) of the chassis or PCB.

When the Proto 2 system units are available, the The EMC Compliance group conducts its testing again, and specifically checks that the countermeasures that were implemented are indeed effective.

When the EMC Compliance group believes that the system can pass the required testing standards, they can send systems to outside testing labs for testing and to receive the testing report that demonstrates compliance.

References

1. Electromagnetic compatibility - Wikipedia. Accessed February 17, 2026.
2. List of common EMC test standards - Wikipedia. Accessed February 17, 2026.
3. Voluntary Control Council for Interference by Information Technology Equipment - Wikipedia. Accessed February 17, 2026.
4. CES-003 — Information Technology Equipment (including Digital Apparatus). Accessed February 17, 2026.
5. Network Equipment-Building System - Wikipedia. Accessed February 19, 2026.
6. Figure 4.4.5.1: The four EMI Coupling modes. (By Steelpillow - Own work, CC BY-SA 3.0. Wikimedia - The four EMI Coupling modes.

This page titled 4.4.5: Electro-magnetic Regulatory Compliance is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

4.4.6: Testing (HALT)

Aside from the EMC Regulatory compliance testing that has been described, other testing is done by the hardware team.

Initial Bring-up

When the system printed circuit board assembly arrives from the prototype assembly facility, the PCB designer will take it to the hardware lab and start the "bring up" process by following steps like the following:

- **Visual Inspection**: For this step, no power is applied to the board. Check for any visible issues such as damage to the edges of the PCB, possible board delamination, scratches across visible PCB traces, bridges between copper traces, solder bridging (adjacent solder joints have solder that melted together), board bowing, problems with drilled through-holes, cold joints (solder did not completely melt), tomb-stoned components (an end of a component has lifted off its solder pad), incorrect components, incorrectly mounted components, or open pins.
- **Physical checks:** For this step, no power is applied to the board. Check continuity of power traces from power rails to integrated circuits.
- **Power Supply Bring-Up**: Ensure the board's power supply is safe and within limits. In this step, power is applied to the board. Check the power supply voltage polarity. Do a temperature check to assess whether there might be a short circuit. Check power rail voltages.
- **Functional Testing**: Verifying the board's ability to process data and communicate with other components.
- **Debugging.**

HALT Testing

Definition: HALT

A **highly accelerated life test (HALT)** is a stress testing methodology for enhancing product reliability in which system prototypes are stressed to a much higher degree than expected from actual use in order to identify weaknesses in the design or manufacturing of the product.[1]

HALT can be effectively used multiple times over a product's life time.

- During product development, it can find design weakness earlier in the product life cycle when changes are much less costly to make. By finding weaknesses and making changes early, HALT can lower product development costs and reduce time to market.
- When HALT is used at the time a product is being introduced into the market, it can expose problems caused by new manufacturing processes.
- When used after a product has been introduced into the market, HALT can be used to audit product reliability caused by changes in components, manufacturing processes, suppliers, etc.[1]

What is HALT Testing

Highly accelerated life testing (HALT) techniques are important in uncovering many of the weak links of a new product. These discovery tests rapidly find weaknesses using accelerated stress conditions. The goal of HALT is to proactively find weaknesses and fix them, thereby increasing product reliability. Because of its accelerated nature, HALT is typically faster and less expensive than traditional testing techniques.

HALT is a test technique called test-to-fail, where a product is tested until failure. HALT does not help to determine or demonstrate the reliability value or failure probability in field. Many accelerated life tests are test-to-pass, meaning they are used to demonstrate the product life or reliability.

It is highly recommended to perform HALT in the initial phases of product development to uncover weak links in a product, so

that there is better chance and more time to modify and improve the product.

HALT uses several stress factors (decided by a Reliability Test Engineer) and/or the combination of various factors. Commonly used stress factors for electronics and mechanical products are:

- temperature,
- vibration, and
- humidity.

Other factors can include voltage, current, power cycling and combinations of them.[1]

HALT procedures

Environmental stresses are applied in a HALT procedure, eventually reaching a level significantly beyond that expected during use.

The stresses used in HALT are typically:

- hot and cold temperatures,
- temperature cycles,
- random vibration,
- power margining, and
- power cycling.[1]

The system under test is in operation during HALT and is continuously monitored for failures. As stress-induced failures occur, the cause should be determined, and if possible, the problem should be repaired so that the test can continue to find other weaknesses.

HALT Test Chamber

A specialized environmental chamber is required for HALT.

A suitable chamber also has to be capable of applying pseudo-random vibration with a suitable profile in relation to frequency. The HALT chamber should be capable of applying random vibration energy from 2 to 10,000 Hz in 6 degrees of freedom and temperatures from -100 to +200°C. Sometimes HALT chambers are called repetitive shock chambers because pneumatic air hammers are used to produce vibration. Test fixtures must transmit vibration to the system under test.

The chamber should also be capable of rapid changes in temperature, 50°C per minute should be considered a minimum rate of change. Usually high power resistive heating elements are used for heating and liquid nitrogen (LN_2) is used for cooling.

The network system under test must be monitored so that if the equipment fails under test, the failure is detected.

Monitoring is typically performed with thermocouple sensors, vibration accelerometers, multi-meters, and data loggers.

Common causes of failures during HALT are:

- poor product design,
- workmanship, and
- poor manufacturing.

Failures to individual components such as resistors, capacitors, diodes, and printed circuit boards occur because of these type of issues. Failure types found during HALT testing are associated with the infant mortality region of the reliability bathtub curve as shown in Figure 4.4.6.1.

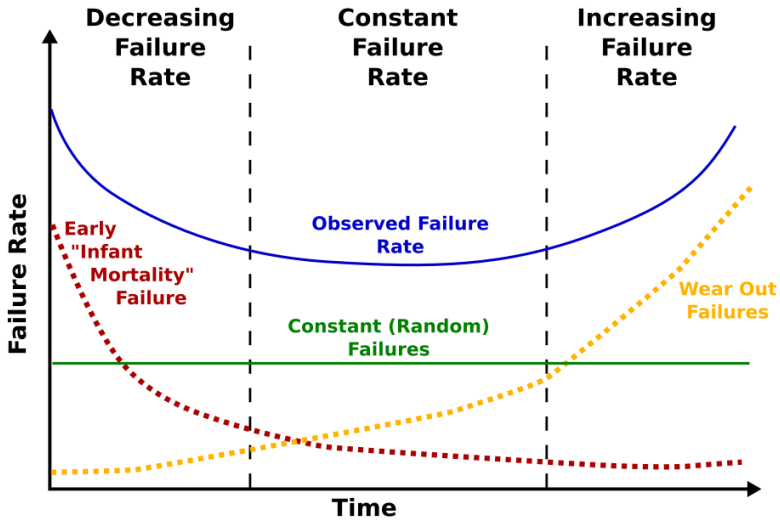


Figure 4.4.6.1: Bathtub curve. (Wyatts derivative work: McSush (talk) - Bathtub_curve.jpg, Public Domain[2])

Accessibility Note

The 'bathtub curve' hazard function (blue, upper solid line) is a combination of a decreasing hazard of early failure (red dotted line) and an increasing hazard of wear-out failure (yellow dotted line), plus some constant hazard of random failure (green, lower solid line).

The bathtub curve has 3 regions:[2]

1. The first region has a decreasing failure rate due to early failures (a.k.a. the "Infant Mortality Phase").
2. The middle region is a constant failure rate due to random failures (a.k.a. the "Useful Life Phase").
3. The last region is an increasing failure rate due to wear-out failures (a.k.a. the "Wear-Out Phase").

References

1. Highly accelerated life test - Wikipedia. Accessed February 19, 2026. Text is available under the Creative Commons Attribution-ShareAlike 4.0 License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.
2. Bathtub curve - Wikipedia. Accessed February 19, 2026.

This page titled 4.4.6: Testing (HALT) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

4.5: Software System Design

As mentioned earlier, this case study is intended to be hardware centric. Perhaps in the future, it will be enhanced with more software material.

For now, this section will provide a very high level overview of the software that might be running on the network switch.

Network Operating system

Definition: Network Operating System

A network operating system (NOS) is a specialized operating system for a network device such as a router, switch, or firewall.[2]

The network operating system provides support for the switch device's underlying hardware, provides support for a myriad of network protocols, helps manage network traffic, provides a user interface, provides support for network operations management by network administrators through a network management system (NMS) using SNMP for example, provides monitoring and alerts, and tracks network performance.

Examples of network operating systems are:

- Cisco IOS,
- Juniper Junos, and
- Arista EOS.

Platform Software

Platform software manages the switch's hardware. This includes the networking ASICs, I/O port interfaces, power supplies, and fan trays.

Software can monitor the speed of fans in the fan trays and fans in power supplies to detect a degradation of speed, which might be indicative of a need to replace the unit before it fails.

Since some vendors support products using their own custom networking ASICs as well as merchant silicon networking ASICs, the networking operating system might be structured to provide a hardware abstraction layer (HAL) to more easily support both types of platforms.

Broadcom SDK(from Google search summary, rephrase)

As discussed earlier, merchant silicon is available, providing the network ASICs that do the heavy lifting of providing switching functionality.

One of the critical deliverables in a Software Development Kit (SDK) or Switch Development Kit (SDK) provided by the ASIC supplier are the device drivers for the networking ASICs. The SDK needs to be integrated into the network operating system software that runs on the switch.

The switch vendor's network operating system likely has a Hardware Abstraction Layer (HAL) as mentioned previously. It will use Application Programming Interfaces (APIs) to communicate with the ASIC supplier's SDK. The SDK in turn interfaces with the ASICs to initialize them, configure them, and monitor them.[3]

Broadcom is one of the leading suppliers of merchant silicon ASICs. The Broadcom Unified SDK (or OpenBCM SDK) provides APIs, drivers, and tools for programming the various Broadcom network switch silicon platforms.

The SDK provides low-level access to Broadcom ASIC hardware, enabling the network operating system to configure packet forwarding, VLANs, QoS, multicast, and other advanced networking features.

It supports various ASIC families as follows:

- XGS devices (e.g., Trident TD2/TD3/TD4, Tomahawk TH/TH2/TH3) and
- DNX devices (e.g., Ramon, Qumran Q2A/Q2U, Jericho JR2/J2C).

Key SDK pieces include:

- Application Programming Interfaces (APIs) for direct hardware programming.
- Abstraction of low level hardware constructs (registers) to simplify programming.
- Device drivers for various Broadcom ASICs. The device drivers work with a layer called the SOC layer (presumably for System on a Chip) to communicate with the ASICs.
- Sample applications demonstrating feature usage.
- A build system for compiling and linking SDK modules into the switch platform.

Support for Network Protocols

A network switch developer publishes a data sheet when the system product is launched into the market. The data sheet usually has a section entitled something like "Standards and protocols" that provides a long list of the networking standards and protocols that the network switch supports. This includes (by general categories):

- ANSI standards (for example, ANSI/TIA-1057 LLDP Media Endpoint Discovery (LLDP-MED)
- IEEE networking standards, for example:
  - IEEE 802.3ab 1000 BASE-T
  - IEEE 802.3ae 10-Gigabit Ethernet
  - IEEE 802.3ae 10-Gigabit Ethernet
  - IEEE 802.1Q VLANs
- RFCs. The The Internet Engineering Task Force (IETF) publishes its technical documentation as RFCs (Requests for Comments.) They describe the Internet's technical foundations, such as addressing, routing, and transport technologies. For example:
  - RFC 826 ARP

Features

Some features that have been incorporated into a specific software release are documented in the "Release Notes" for that release.

Command Line Interface (CLI)

The network operating system provides a command line interface as a means for issuing commands to the switch to configure and manage it.

Network Management

A Network Management System (NMS) is software that helps with monitoring, controlling, and automating the operation of a network switch. It also provides a means for monitoring the health of the switch and in maintaining it.

References

1. IETF | Introduction. Accessed February 19, 2026.

2. Network operating system - Wikipedia. Accessed February 19, 2026. Text is available under the Creative Commons Attribution-ShareAlike 4.0 License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.
3. Broadcom switch SDK software enables rapid development and deployment. Accessed February 23, 2026.

This page titled 4.5: Software System Design is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

4.6: Project Management Considerations

In this section we will look at special considerations in a few areas:

- **Project schedule.**
- **Project budget.**
- **Procurement.** There is a separate subsection dealing with the RFP/RFQ process as well as the general process for filing Purchase Requisitions (PRs) that will be transformed into Purchase Orders (POs) that are sent to outside vendors.
- **Team Management**. Since the project involves a fair number of people over a lengthy period of time, this separate subsection provides some background on motivation theory and leadership.

Project Schedule considerations

Number of prototype revisions

When preparing the project schedule, the project manager needs to make some assumptions about the following:

- number of prototype printed circuit board revisions,
- number of prototype ASIC revisions, and
- number of prototype chassis revisions.

Based on a number of factors, including technology risk and past track record of the hardware team, one might conclude that there will be two PCB revisions. Testing of revision 1 (Rev 1) prototypes might reveal issues that need to be corrected in a second revision.

As mentioned earlier, the networking ASICs can be merchant silicon or produced as custom ASICs by the network switch company's ASIC team. In either case, one needs to assess how many revisions of each ASIC will likely be needed. The ASIC development schedules are important as they will gate when the PCBs that use them can be assembled. When using a merchant silicon provider, new ASICs of any revision are often on allocation. This means that they are initially scarce, and limited quantities are provided to network switch companies. If insufficient numbers of ASICs are available, the PCB assemblies may need to be split, with the second build starting when the balance of the needed ASICs are delivered.

Number of prototype switches needed

As the project manager, you will need to determine how many prototype switches of each revision level need to be built. Typically this can be done in the project core team meetings.

When the switches are assembled, they need to be delivered to their intended recipients.

Project Budget management

Capex and Opex

There are two major categories of expenses:

- Operating expenses
- Capital expenses.

Expense Categories

Term	Abbreviation	Definition
Operating expense	OpEx	An expense that a business incurs through its normal business operations. Operating expenses include categories such as: rent, equipment, inventory costs, marketing, payroll, insurance, and funds allocated for research and development
Capital expense	CapEx	Capital expenses are financial outlays by a company to acquire, upgrade, and maintain physical assets such as property, plants, buildings, technology, or equipment.

As the project manager, you will assemble the project budget which is a tally of all project Opex expenses over the duration of the project. The budget is time-based, showing in which month or quarter the project expects to incur the various project expenses, shown by financial accounting categories.

This project budget is approved as the project baseline budget when the project is approved.

The following is a list of typical financial budget accounting categories that a technology project might use.

Expense Accounting Categories

Accounting Category	Description
Prototype materials	Costs associated with building prototypes
NRE	**N**on-**R**ecurring **E**ngineering Expenses
Outside Services	Costs associated with services provided by parties outside of the organization
Depreciation expense	Expense associated with the depreciation of capital equipment over a period of time (useful life of asset)
Design costs	Expenses for external design activities
Project equipment rental	Rental charges for equipment (e.g., traffic generators)
Software licensing	License charges for software tools
Tooling	Tooling charges for test fixtures
Regulatory testing	Expenses for EMI, EMC, FCC, VDE, NEBS regulatory testing

Accounting Category	Description
Travel	Expenses associated with business travel (airfare, hotel)

Capital Equipment/Depreciation

When a piece of capital equipment is purchased, you will need to work with your finance representative to ensure that it is handled properly.

In the network switch project, examples of capital purchases would be buying network traffic generators. The network traffic generator company would be paid the total amount of the purchase. This is a "**capital expense**."

However, internal to the network switch company, the total purchase price of the capital tangible asset is put on the accounting books for the total amount of the purchase. The asset is "depreciated" over a set period of time depending on the type of asset.
Note

> The term "**amortization**" is used for loans and for intangible assets (e.g., patents).

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	**Depreciation Schedule**												
2													
3	Capital equipment cost	$360,000											
4	Residual value	$0											
5	Depreciation period (years)	3											
6	Depreciation period (quarters)	12											
7	Depreciation expense per quarter	$30,000											
8													
9	**Depreciation schedule (straight line)**												
10	**Period (quarter)**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
11	**Asset value on the books**	$360,000	$330,000	$300,000	$270,000	$240,000	$210,000	$180,000	$150,000	$120,000	$90,000	$60,000	$30,000
12	**Depreciation amount**	$30,000	$30,000	$30,000	$30,000	$30,000	$30,000	$30,000	$30,000	$30,000	$30,000	$30,000	$30,000
13	**Asset value (end of period)**	$330,000	$300,000	$270,000	$240,000	$210,000	$180,000	$150,000	$120,000	$90,000	$60,000	$30,000	$0

Figure 4.6.1: Sample straight-line depreciation schedule. (Copyright; Robert Slate, CIS 79 course materials)

Assuming the asset is depreciated using "straight-line depreciation" over three years (12 quarters), one twelfth (1/12) of the asset's initial value is treated as a **depreciation expense** in the current quarter, and that same amount is deducted from the value of the asset on the accounting books. This is assuming that "straight line depreciation" is being used rather than other depreciation methods that might tilt the amount of depreciation to the beginning periods. This continues for the remaining eleven (11) quarters.

At the end of twelve (12) quarters, the asset's value on the accounting books would be zero, or zero residual value.

From a budget management perspective, these quarterly depreciation expenses are included in the project's Opex budget.

In the example, the original cost of the equipment is $360,000. Since it is being depreciated over 12 quarters, 1/12 of $360,000 or $30,000 is treated as a depreciation expense each quarter. At the end of the first quarter, the $360,000 asset value originally on the books is reduced by $30,000, so that the asset is now on the books for $330,000. This is repeated for the remaining 11 quarters.

Expense Accruals

Definition: Accrual or Accrued Expense

"An accrual, or accrued expense, is a means of recording an **expense that was incurred** in one accounting period **but not paid until a future accounting period**. Accruals differ from Accounts Payable transactions in that an invoice is usually not yet received and entered into the system."[1]

Here are a few important things to consider with accruals:

- “Any accrued expense a company incurs but hasn’t paid is known as an **accrual liability**. Also called expense accruals, these transactions are **recorded on the balance sheet as accounts payable liabilities and on the company’s income statement as an expense**.”[2]
- “At the end of each accounting period, all accrued liabilities are adjusted on the balance sheet to provide documentation for goods delivered and services performed that have not yet been billed.”[2]

Operationally, accruals typically appear on a project manager's radar toward the end of a financial quarter. The organization's financial representative will typically schedule a meeting looking for any specific accruals that need to be made on behalf of your project, based on goods delivered and services performed that have not yet been billed. It is in the project manager's interest to take care of this accurately. When the bill or invoice is received, the accrual can be "unwound", tapping into the funds that were set aside in the accounting system for this purpose.

Timing of Expenses

Another interaction with your financial partner might include the timing of your project expenses.

Suppose the financial outlook for the company is looking a bit bleak one quarter due to any number of potential factors, such as slow sales if the overall economic outlook is poor and the company's customers are slowing their expenditures. The company's Chief Financial Officer (CFO) may issue a directive to slow spending during the quarter, including project spending.

The reverse can occur where the economy is booming and the sales team is projecting much higher revenue for the current quarter. Project managers might be asked if there are any expenses that could be accelerated into the current quarter.

References

1. Period ending accruals.
2. Accounts Payable Accruals.

This page titled 4.6: Project Management Considerations is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

4.6.1: Procurement

In this section we will be looking at two major topics:

1. the bid process for work to be done outside the company and relevant documents and
2. the process of creating a Purchase Requisition (PR) and having a Purchase Order (PO) sent to an outside supplier.

This page titled 4.6.1: Procurement is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

4.6.1.1: Bid Documents (RFI, RFP, etc.)

In most large companies, there is a “Purchasing” department that has company-wide responsibility for managing procurements. It is often in organizations with sames such as Manufacturing, Supply Chain, or Operations.

Usually, only someone in the Purchasing department can formally issue a “**Purchase Order**” (PO) listing the quantities of the ordered items and terms and conditions (Ts&Cs). You request a Purchase Order via the “**Purchase Requisition**” process, to be discussed in a subsequent section of the textbook.

This Purchasing group also manages Approved Vendor Lists (AVLs).

Bid Documents

As part of the project work, an assessment is made for certain pieces of work to be done with resources internal to the organization, or in other cases, the work may be done by an outside supplier.

If some work is going to be contracted out, the following documents are typically used.

Contracting Document Types

Document	Meaning	Description
RFI	Request for Information	Request for information from potential suppliers about their capabilities to meet your needs
RFP	Request for Proposal	Request for suppliers/vendors to provide a solution
RFQ	Request for Quote	Request for a price quotation from suppliers/vendors, typically when the requested item is generally available in the market and price is the major factor

Bid Process Flow

The process of handling the bid process is as follows:

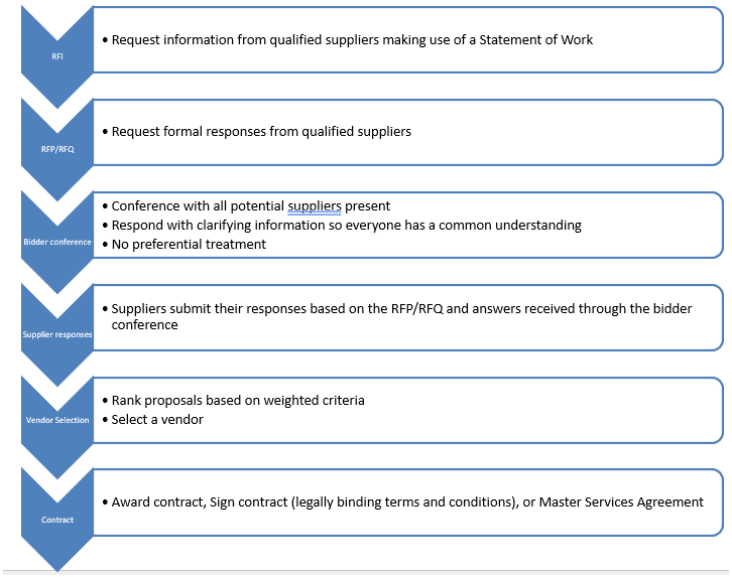


Figure 4.6.1.1.1: Bid Process Flow. (Robert Slate, CIS course materials)

The process steps in Figure 4.6.1.1.1 are further described here.

- RFI (Request for Information): the organization puts out a Request for Information from potential suppliers about their capabilities to meet your organization's needs, usually using a Statement of Work that you provide.
- RFP/RFQ: the organization requests responses from qualified suppliers.
- Bidder conference: In the interest of fairness, the following is usually done:
  - There is a conference held with all of the potential suppliers present.
  - If any clarifications are in order, your organization responds with the clarification send to all parties so that everyone has a common understanding.
  - The goal is to have no preferential treatment of any suppliers.
- Supplier responses: the suppliers submit their responses based on the RFP or RFQ and answers that they received through the bidder conference.
- Vendor selection. There are several ways to do this. One is by ranking vendors based on weighted criteria and scored assigned to each criterion. The supplier with the highest score wins.
- Contract. The contract is awarded to the supplier. The contract is signed by both parties with legally binding terms and conditions (Ts & Cs). Depending on the supplier, your organization may have a Master Services Agreement in place with the supplier.

Contracts

Usually only corporate executives and Purchasing organization personnel are authorized to sign contracts on behalf of an organization.

Contracts go through approval with the legal department.

The legal department usually has responsibility for maintaining the official repository of contracts, including any agreed-to changes to contracts.

This page titled 4.6.1.1: Bid Documents (RFI, RFP, etc.) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

4.6.1.2: PR to PO procedure

There is a business process for creating a Purchase Order (PO) that a company sends to a supplier. That process is described in Figure 4.6.1.2.1.

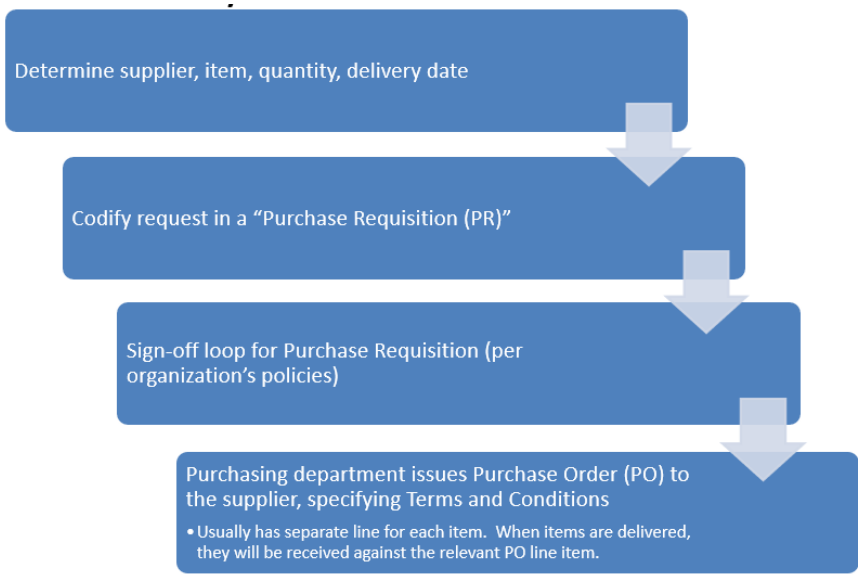


Figure 4.6.1.2.1: Summary of Purchase Order Process. (Robert Slate, CIS 79 course materials)

Accessibility Note

The process steps in Figure 4.6.1.2.1 are described below:

1. Determine supplier, item, quantity, delivery date. This is typically done by the requestor.
2. Codify the request in a "Purchase Requisition (PR)". This is done using the organization's selected procurement software tool. A project code specifies which project will be incurring the expense.
3. Sign-off loop for a Purchase Requisition (per organization's policies). The procurement software tool will route the Purchase Requisition in sequence to everyone in the sign-off chain or signature loop.
4. The Purchasing department issues a Purchase Order (PO) to the supplier, specifying Terms and Conditions. The Purchase Order usually has separate lines for each item, called "line items." When items are delivered to the organization's Receiving department, they will be received against the relevant PO line item.

Sign-off on Purchase Requisitions

In order to have visibility into the flow of project expenses, it is useful for the project manager to be on the approvers list for all Purchase Requisitions (that result in Purchase Orders), or perhaps only those over a certain threshold dollar amount. This can be instrumented by automated triggering by the specified project code. Alternatively, it might be organizational policy that you request that everyone filing a Purchase Requisition for your project add you (the project manager) as an approver of the Purchase Requisition.

Having the project manager on the Purchase Requisition approver list provides visibility into:

- the correct item being ordered
- the correct item quantity
- the cost (including tax!), including any expedite fees, shipping costs, and import duties. The project manager can cross-check this expense against planned cost in the project budget.

- expected delivery timeframe, including clearing Customs (cross check against schedule).

The project manager can bounce requisitions back to the originator of the Purchase Requisition if information is not correct.

Note

The design team and project manager have choices on number of days for board fabs. The quotation from the board fab vendor will quote prices based on 'turn time", the number of days to complete the board fabrication work. Doing the work very quickly will have the highest cost, whereas doing it more slowly will cost less.

Making this decision is an example of the application of Project Management Triangle.

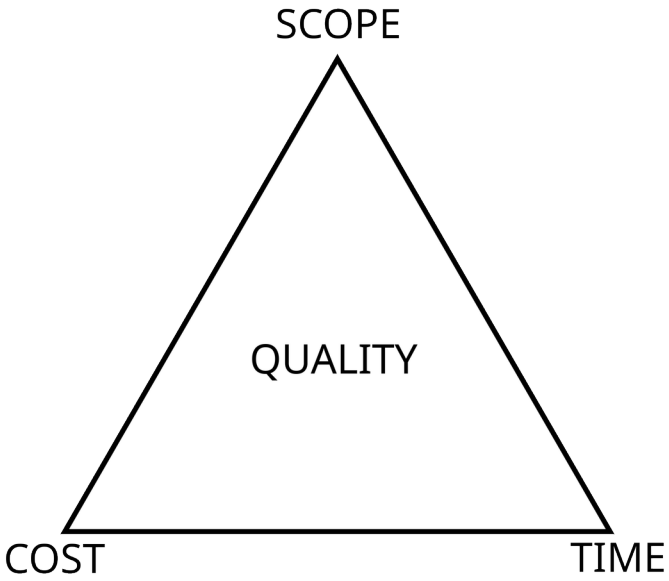


Figure 4.6.1.2.2: Project Management Triangle. (Public domain)

The **project management triangle** is a model of the constraints experienced when conducting project management, as depicted in Figure 4.6.1.2.2. Its premises are:

1. The quality (depicted in the center of the triangle) of work is constrained by the project's cost (budget), time (deadlines), and scope (features) that are shown at the vertices of the triangle.
2. Scope, time, and cost are interdependent and the project manager can trade between the constraints.
3. Changes in one constraint necessitate changes in others to compensate, otherwise quality will suffer.[1]

Anticipating upcoming expenses

As part of monitoring the project, the project manager should know when Purchase Requisitions/Purchase Orders need to be in place by consulting project documents:

- Project schedule
- Project budget (expenses by quarter)

The project manager can ping the expected filers of Purchase Requisitions as needed to ensure that they take action in a timely manner.

References

1. Project management triangle - Wikipedia. Accessed January 23, 2026.

This page titled 4.6.1.2: PR to PO procedure is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

4.6.2: Team Management

Since the project involves a fair number of people over a lengthy period of time, this section provides some background readings on:

- Motivation theory and
- Leadership.

This page titled 4.6.2: Team Management is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

4.6.2.1: Content Theories of Motivation

The theories presented in this section focus on the importance of human needs.

A common thread through all of them is that people have a variety of needs. A **need** is a human condition that becomes "energized" when people feel deficient in some respect. When we are hungry, for example, our need for food has been energized.

Two features of needs are key to understanding motivation.

1. First, when a need has been energized, we are motivated to satisfy it. We strive to make the need disappear. **Hedonism,** one of the first motivation theories, assumes that people are motivated to satisfy mainly their own needs (seek pleasure, avoid pain). Long since displaced by more refined theories, hedonism clarifies the idea that needs provide direction for motivation.
2. Second, once we have satisfied a need, it ceases to motivate us. When we've eaten to satiation, we are no longer motivated to eat. Other needs take over and we endeavor to satisfy them.

A **manifest need** is whatever need is motivating us at a given time. Manifest needs dominate our other needs.

Instincts are our natural, fundamental needs, basic to our survival. Our needs for food and water are instinctive. Many needs are learned. We are not born with a high (or low) need for achievement; we learn to need success (or failure). The distinction between instinctive and learned needs sometimes blurs; for example, is our need to socialize with other people instinctive or learned?

Manifest Needs Theory

One major problem with the need approach to motivation is that we can make up a need for every human behavior. Do we "need" to talk or be silent? The possibilities are endless. In fact, around the 1920s, some 6,000 human needs had been identified by behavioral scientists!

Henry A. Murray recognized this problem and condensed the list into a few instinctive and learned needs.[3]

- Instincts, which Murray called **primary needs**, include physiological needs for food, water, sex (procreation), urination, and so on.
- Learned needs, which Murray called **secondary needs**, are learned throughout one's life and are basically psychological in nature. They include such needs as the need for achievement, for love, and for affiliation (see Table 4.6.2.1.1).[4]

Table 4.6.2.1.1 *Sample Items from Murray's List of Needs*

Sample Items from Murray's List of Needs	
Social Motive	**Brief Definition**
Abasement	To submit passively to external force. To accept injury, blame, criticism, punishment. To surrender.
Achievement	To accomplish something difficult. To master, manipulate, or organize physical objects, human beings, or ideas.
Affiliation	To draw near and enjoyably cooperate or reciprocate with an allied other (an other who resembles the subject or who likes the subject). To please and win affection of a coveted object. To adhere and remain loyal to a friend.
Aggression	To overcome opposition forcefully. To fight. To revenge an injury. To attack, injure, or kill another. To oppose forcefully or punish another.
Autonomy	To get free, shake off restraint, break out of confinement.
Counteraction	To master or make up for a failure by restriving.
Source: Adapted from C. S. Hall and G. Lindzey, *Theories of Personality.* Sample items from Murray's List of Needs. Copyright 1957 by John Wiley & Sons, New York.	

Sample Items from Murray's List of Needs	
Social Motive	**Brief Definition**
Defendance	To defend the self against assault, criticism, and blame. To conceal or justify a misdeed, failure, or humiliation. To vindicate the ego.
Deference	To admire and support a superior. To praise, honor, or eulogize.
Dominance	To control one's human environment. To influence or direct the behavior of others by suggestion, seduction, persuasion, or command.
Exhibition	To make an impression. To be seen and heard. To excite, amaze, fascinate, entertain, shock, intrigue, amuse, or entice others.
Harm avoidance	To avoid pain, physical injury, illness, and death. To escape from a dangerous situation. To take precautionary measures.
Infavoidance	To avoid humiliation. To quit embarrassing situations or to avoid conditions that may lead to belittlement or the scorn or indifference of others.
Nurturance	To give sympathy and gratify the needs of a helpless object: an infant or any object that is weak, disabled, tired, inexperienced, infirm, defeated, humiliated, lonely, dejected, sick, or mentally confused. To assist an object in danger. To feed, help, support, console, protect, comfort, nurse, heal.
Order	To put things in order. To achieve cleanliness, arrangement, organization, balance, neatness, tidiness, and precision.
Play	To act for "fun" without further purpose. To like to laugh and make jokes. To seek enjoyable relaxation from stress.
Rejection	To separate oneself from a negatively valued object. To exclude, abandon, expel, or remain indifferent to an inferior object. To snub or jilt an object.
Sentience	To seek and enjoy sensuous impressions.
Sex	To form and further an erotic relationship. To have sexual intercourse.
Succorance	To have one's needs gratified by the sympathetic aid of an allied object.
Understanding	To ask or answer general questions. To be interested in theory. To speculate, formulate, analyze, and generalize.
Source: Adapted from C. S. Hall and G. Lindzey, *Theories of Personality.* Sample items from Murray's List of Needs. Copyright 1957 by John Wiley & Sons, New York.	

Murray's main premise was that people have a variety of needs, but only a few are expressed at a given time. When a person is behaving in a way that satisfies some need, Murray called the need manifest. **Manifest needs theory** assumes that human behavior is driven by the desire to satisfy needs. Lucretia's chattiness probably indicates her need for affiliation. This is a manifest need. But what if Lucretia also has a need to dominate others? Could we detect that need from her current behavior? If not, Murray calls this a latent need. A **latent need** cannot be inferred from a person's behavior at a given time, yet the person may still possess that need. The person may not have had the opportunity to express the need. Or she may not be in the proper environment to solicit behaviors to satisfy the need. Lucretia's need to dominate may not be motivating her current behavior because she is with friends instead of coworkers.

Manifest needs theory laid the groundwork for later theories, most notably McClelland's learned needs theory, that have greatly influenced the study of organizational behavior. The major implication for management is that some employee needs are latent. Managers often assume that employees do not have certain needs because the employees never try to satisfy them at work. Such needs may exist (latent needs); the work environment is simply not conducive to their manifestation (manifest needs). A reclusive accountant may not have been given the opportunity to demonstrate his need for achievement because he never received challenging assignments.

Learned Needs Theory

David C. McClelland and his associates (especially John W. Atkinson) built on the work of Murray for over 50 years. Murray studied many different needs, but very few in any detail.

McClelland's research differs from Murray's in that McClelland studied three needs in-depth: the need for achievement, the need for affiliation, and the need for power (often abbreviated, in turn, as nAch, nAff, and nPow).[5] McClelland believes that these three needs are learned, primarily in childhood. But he also believes that each need can be taught, especially nAch. McClelland's research is important because much of the current thinking about organizational behavior is based on it.

Need for Achievement

The **need for achievement (nAch)** is how much people are motivated to excel at the tasks they are performing, especially tasks that are difficult. Of the three needs studied by McClelland, nAch has the greatest impact. The need for achievement varies in intensity across individuals. This makes nAch a personality trait as well as a statement about motivation. When nAch is being expressed, making it a manifest need, people try hard to succeed at whatever task they're doing. We say these people have a high achievement motive. A **motive** is a source of motivation; it is the need that a person is attempting to satisfy. Achievement needs become manifest when individuals experience certain types of situations.

To better understand the nAch motive, it's helpful to describe high-nAch people. You probably know a few of them. They're constantly trying to accomplish something.

McClelland describes three major characteristics of high-nAch people:

1. They feel personally responsible for completing whatever tasks they are assigned. They accept credit for success and blame for failure.
2. They like situations where the probability of success is moderate. High-nAch people are not motivated by tasks that are too easy or extremely difficult. Instead, they prefer situations where the outcome is uncertain, but in which they believe they can succeed if they exert enough effort. They avoid both simple and impossible situations.
3. They have very strong desires for feedback about how well they are doing. They actively seek out performance feedback. It doesn't matter whether the information implies success or failure. They want to know whether they have achieved or not. They constantly ask how they are doing, sometimes to the point of being a nuisance.

Why is nAch important to organizational behavior? The answer is, the success of many organizations is dependent on the nAch levels of their employees.[6] This is especially true for jobs that require self-motivation and managing others. Employees who continuously have to be told how to do their jobs require an overly large management team, and too many layers of management spell trouble in the current marketplace. Today's flexible, cost-conscious organizations have no room for top-heavy structures; their high-nAch employees perform their jobs well with minimal supervision.

Many organizations manage the achievement needs of their employees poorly. A common perception about people who perform unskilled jobs is that they are unmotivated and content doing what they are doing. But, if they have achievement needs, the job itself creates little motivation to perform. It is too easy. There are not enough workers who feel personal satisfaction for having the cleanest floors in a building. Designing jobs that are neither too challenging nor too boring is key to managing motivation. Job enrichment is one effective strategy; this frequently entails training and rotating employees through different jobs, or adding new challenges.

Need for Affiliation

This need is the second of McClelland's learned needs. The **need for affiliation (nAff)** reflects a desire to establish and maintain warm and friendly relationships with other people. As with nAch, nAff varies in intensity across individuals. As you would expect, high-nAff people are very sociable. They're more likely to go bowling with friends after work than to go home and watch television. Other people have lower affiliation needs. This doesn't mean that they avoid other people, or that they dislike others. They simply don't exert as much effort in this area as high-nAff people do.

The nAff has important implications for organizational behavior. High-nAff people like to be around other people, including other people at work. As a result, they perform better in jobs that require teamwork. Maintaining good relationships with their coworkers is important to them, so they go to great lengths to make the workgroup succeed because they fear rejection. So, high-nAff employees will be especially motivated to perform well if others depend on them. In contrast, if high-nAff people perform jobs in isolation from other people, they will be less motivated to perform well. Performing well on this job won't satisfy their need to be around other people.

Effective managers carefully assess the degree to which people have high or low nAff. Employees high in nAff should be placed in

jobs that require or allow interactions with other employees. Jobs that are best performed alone are more appropriate for low-nAff employees, who are less likely to be frustrated.

Need for Power

The third of McClelland's learned needs, the **need for power (nPow)**, is the need to control things, especially other people. It reflects a motivation to influence and be responsible for other people. An employee who is often talkative, gives orders, and argues a lot is motivated by the need for power over others.

Employees with high nPow can be beneficial to organizations. High-nPow people do have effective employee behaviors, but at times they're disruptive. A high-nPow person may try to convince others to do things that are detrimental to the organization. So, when is this need good, and when is it bad? Again, there are no easy answers. McClelland calls this the "two faces of power."[7] A *personal power seeker* endeavors to control others mostly for the sake of dominating them. They want others to respond to their wishes whether or not it is good for the organization. They "build empires," and they protect them.

McClelland's other power seeker is the *social power seeker*. A high social power seeker satisfies needs for power by influencing others, like the personal power seeker. They differ in that they feel best when they have influenced a workgroup to achieve the group's goals, and not some personal agenda. High social power-seekers are concerned with goals that a workgroup has set for itself, and they are motivated to influence others to achieve the goal. This need is oriented toward fulfilling responsibilities to the employer, not to the self.

McClelland has argued that the high need for social power is the most important motivator for successful managers. Successful managers tend to be high in this type of nPow. High need for achievement can also be important, but it sometimes results in too much concern for personal success and not enough for the employer's success. The need for affiliation contributes to managerial success only in those situations where the maintenance of warm group relations is as important as getting others to work toward group goals.

The implication of McClelland's research is that organizations should try to place people with high needs for social power in managerial jobs. It is critical, however, that those managerial jobs allow the employee to satisfy the nPow through social power acquisition. Otherwise, a manager high in nPow may satisfy this need through the acquisition of personal power, to the detriment of the organization.

ETHICS IN PRACTICE

Corporate Social Responsibility as a Motivating Force

Whatever their perspective, most people have a cause that they are passionate about. Bitcoin or net neutrality, sea levels or factory farming—social causes bind us to a larger context or assume a higher purpose for living better.

So what motivates employees to give their all, work creatively, and be fully engaged? According to CB Bhattacharya, the Pietro Ferrero Chair in Sustainability at ESMT European School of Management and Technology in Berlin, Germany, employment engagement, or how positive employees feel about their current job, was at an all-time low globally in 2016: 13 percent. But not all companies battle such low engagement rates. Unilever employees more than 170,000 workers globally and has an employ engagement level around 80 percent. How? Bhattacharya credits the success of Unilever, and other companies with similar engagement levels, to an emphasis on a "sustainable business model." He outlines eight steps that companies take to move sustainability and social responsibility from buzzwords to a company mission capable of motivating employees (Knowledge @ Wharton 2016).

According to Bhattacharya, a company needs to first define what it does and its long-term purpose, and then reconcile its sustainability goals with its economic goals. With its purpose and goals defined, it can then educate the workforce on sustainable methods to create knowledge and competence. Champions for the effort must be found throughout the organization, not just at the top. Competition should be encouraged among employees to find and embrace new goals. Sustainability should be visible both within and outside the company. Sustainability should be tied to a higher purpose and foster a sense of unity not simply among employees, but even with competition at a societal level (Knowledge @ Wharton 2016).

Other companies have made social responsibility an everyday part of what they do. Launched in 2013, Bombas is the brainchild of Randy Goldberg and David Heath. Goldberg and Heath discovered that socks are the most-requested clothing at homeless shelters. In response, the two entrepreneurs launched a line of socks that not only "reinvents" the sock (they claim), but also helps those in need. For each pair of socks purchased, the company donates a pair of socks to someone in need (Mulvey 2017). According to the company website, "Bombas exists to help solve this problem, to support the homeless community, and to bring awareness to an under-publicized problem in the United States" (n.p.). Although the New York-based company is still growing, as of October 2017 Bombas had donated more than four million pairs of socks (Bombas 2017).

In 2016, the Royal Bank of Scotland (RBS) launched a pilot program called Jump in which employees participated in challenges on ways to save water and electricity, as well as other sustainability issues. At the end of the pilot, 95 percent of the employees reported that they felt the program had contributed to employee engagement, team building, and environmental stability. Given the success of the program, in 2017 it was expanded to all RBS sites and a smartphone app was added to help employees participate in the challenges (Barton 2017).

Placing a *company* in a larger context and adding a second, higher purpose than the established company goals motivates employees to police the company itself to be a better global citizen. Companies benefit from reduced waste and increased employee engagement. Many companies are successfully motivating their staff, and working toward more sustainable practices, while improving lives directly.

Sources:

1. Barton, Tynan. 2017. "RBS boosts employee motivation and engagement through its CSR approach." Employee Benefits.
2. Bombas. 2017. "Giving Back."
3. Knowledge @ Wharton. 2016. "How Companies Can Tap Sustainability to Motivate Staff." Sustainability to Motivate Staff.
4. Mulvey, Kelsey. 2017. "This company spent two years perfecting gym socks, and it paid off." *Business Insider*. Athletic Socks.

Questions:

1. Do you think social responsibility to promote sustainable practices? Why or why not?
2. Do you think most companies' CSR programs are essentially PR gimmicks? Why or why not? Give examples.

Maslow's Hierarchy of Needs

Any discussion of needs that motivate performance would be incomplete without considering Abraham Maslow.[8] Thousands of managers in the 1960s were exposed to Maslow's theory through the popular writings of Douglas McGregor.[9] Today, many of them still talk about employee motivation in terms of Maslow's theory.

Maslow was a psychologist who, based on his early research with primates (monkeys), observations of patients, and discussions with employees in organizations, theorized that human needs are arranged hierarchically. That is, before one type of need can manifest itself, other needs must be satisfied. For example, our need for water takes precedence over our need for social interaction (this is also called *prepotency*). We will always satisfy our need for water before we satisfy our social needs; water needs have prepotency over social needs. Maslow's theory differs from others that preceded it because of this hierarchical, prepotency concept.

Maslow went on to propose five basic types of human needs. This is in contrast to the thousands of needs that earlier researchers had identified, and also fewer than Murray identified in his theory. Maslow condensed human needs into a manageable set. Those five human needs, in the order of prepotency in which they direct human behavior, are:

1. ***Physiological and survival needs***. These are the most basic of human needs and include the needs for water, food, sex, sleep, activity, stimulation, and oxygen.
2. ***Safety and security needs***. These needs invoke behaviors that assure freedom from danger. This set of needs involves meeting threats to our existence, including extremes in environmental conditions (heat, dust, and so on), assault from other humans, tyranny, and murder. In other words, satisfaction of these needs prevents fear and anxiety while adding stability and predictability to life.

3. ***Social needs***. These needs reflect human desires to be the target of affection and love from others. They are especially satisfied by the presence of spouses, children, parents, friends, relatives, and others to whom we feel close. Feelings of loneliness and rejection are symptoms that this need has not been satisfied.
4. ***Ego and esteem***. Esteem needs to go beyond social needs. They reflect our need to be respected by others, and to have esteem for ourselves. It is one thing to be liked by others. It is another thing to be respected for our talents and abilities. Ego and esteem needs have internal (self) and external (others) focuses. An internal focus includes desires for achievement, strength, competence, confidence, and independence. An external focus includes desires to have prestige, recognition, appreciation, attention, and respect from others. Satisfaction of external esteem needs can lead to satisfaction of internal esteem needs.
5. ***Self-actualization***. Self-actualization needs are the most difficult to describe. Unlike the other needs, the need for self-actualization is never completely satisfied. Self-actualization involves a desire for self-fulfillment, "to become more and more what one is, to become everything that one is capable of becoming."[10] Because people are so different in their strengths and weaknesses, in capacities and limitations, the meaning of self-actualization varies greatly. Satisfying self-actualization needs means developing all of our special abilities to their fullest degree.

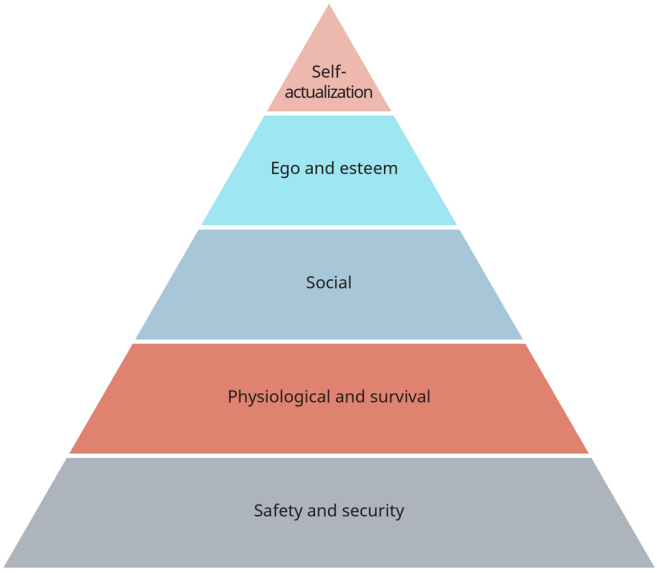


Figure 4.6.2.1.1: Maslow's Hierarchy of Needs (Source: Based on A. H. Maslow. 1943. A theory of human motivation. Psychological Bulletin 50:370–396.)

Accessibiity Note

Figure 4.6.2.1.1 shows Maslow's Hierarchy of Needs as a triangle, going from the top to bottom:

- Self-actualization
- Ego and esteem
- Social
- Physiological and survival
- Safety and security.

Figure 4.6.2.1.1 illustrates Maslow's proposed hierarchy of needs. According to his theory, people first direct their attention to satisfying their lower-order needs. Those are the needs at the bottom of the pyramid (physiological, safety, and security). Once those needs have been satisfied, the next level, social needs, becomes energized. Once satisfied, we focus on our ego and esteem needs. Maslow believed that most people become fixated at this level. That is, most people spend much of their lives developing self-esteem and the esteem of others. But, once those esteem needs are satisfied, Maslow predicted that self-actualization needs would dominate. There are no higher levels in the pyramid, because self-actualization needs can never be fully satisfied. They represent a continuing process of self-development and self-improvement that, once satisfied on one dimension (painting), create motivation to continue on other dimensions (sculpting).

An overriding principle in this theory is that a person's attention (direction) and energy (intensity) will focus on satisfying the lowest-level need that is not currently satisfied. Needs can also be satisfied at some point but become active (dissatisfied) again. Needs must be "maintained" (we must continue to eat occasionally). According to Maslow, when lower-level needs are reactivated, we once again concentrate on that need. That is, we lose interest in the higher-level needs when lower-order needs are energized.

The implications of Maslow's theory for organizational behavior are as much conceptual as they are practical. The theory posits that to maximize employee motivation, employers must try to guide workers to the upper parts of the hierarchy. That means that the employer should help employees satisfy lower-order needs like safety and security and social needs. Once satisfied, employees will be motivated to build esteem and respect through their work achievements. Figure 4.6.2.1.1 shows how Maslow's theory relates to factors that organizations can influence. For example, by providing adequate pay, safe working conditions, and cohesive work groups, employers help employees satisfy their lower-order needs. Once satisfied, challenging jobs, additional responsibilities, and prestigious job titles can help employees satisfy higher-order esteem needs.

Maslow's theory is still popular among practicing managers. Organizational behavior researchers, however, are not as enamored with it because research results don't support Maslow's hierarchical notion. Apparently, people don't go through the five levels in a fixed fashion. On the other hand, there is some evidence that people satisfy the lower-order needs before they attempt to satisfy higher-order needs. Refinements of Maslow's theory in recent years reflect this more limited hierarchy.[11] The self-assessment below will allow you to evaluate the strength of your five needs.

Alderfer's ERG Theory

Clayton Alderfer observed that very few attempts had been made to test Maslow's full theory. Further, the evidence accumulated provided only partial support. During the process of refining and extending Maslow's theory, Alderfer provided another need-based theory and a somewhat more useful perspective on motivation.[12] Alderfer's **ERG theory** compresses Maslow's five need categories into three: existence, relatedness, and growth.[13] In addition, ERG theory details the dynamics of an individual's movement between the need categories in a somewhat more detailed fashion than typically characterizes interpretations of Maslow's work.

As shown in Figure 4.6.2.1.2, the ERG model addresses the same needs as those identified in Maslow's work:



Figure 4.6.2.1.2: Alderfer's ERG Theory (Attribution: Copyright Rice University, OpenStax, under CC-BY 4.0 license)

Accessibility Note

Figure 4.6.2.1.2 shows Alderfer's ERG model that categorizes Maslow's hierarchy of needs into three groups of needs. From bottom upward, the groups are Existence needs, Relatedness needs, and Growth needs.

- ***Existence*** *needs* include physiological and material safety needs. These needs are satisfied by material conditions and not through interpersonal relations or personal involvement in the work setting. These include:
  - Physiological needs
  - Material safety needs
- ***Relatedness*** *needs* include all of Maslow's social needs, plus social safety and social esteem needs. These needs are satisfied through the exchange of thoughts and feelings with other people. These include:
  - Social needs
  - Social esteem needs
  - Interpersonal safety needs
- ***Growth needs*** include self-esteem and self-actualization needs. These needs tend to be satisfied through one's full involvement in work and the work setting. These include:
  - Internal self-esteem needs
  - Self-actualization needs

Figure 4.6.2.1.3 identifies a number of ways in which organizations can help their members satisfy these three needs.



Figure 4.6.2.1.3: Satisfying Existence, Relatedness, and Growth Needs (Attribution: Copyright Rice University, OpenStax, under CC-BY 4.0 license)

Accessibility Note

This is an accessibility note for Figure 4.6.2.1.3 that shows the various ways in which organizations can help their members satisfy three needs. From bottom upward, the needs are Existence needs, Relatedness needs, and Growth needs.

The details of Growth Opportunities, Relatedness Opportunities, and Existence Opportunities are as follows:

- Growth Opportunities
  - Challenging job, Creativity, Organizational advancement, Responsibility, Autonomy, Interesting Work, Achievement, and Participation.
- Relatedness Opportunities
  - Friendship, Interpersonal security, Athletic teams, Social recognition, Quality supervision, Work teams, Social events, and Merit pay.
- Existence Opportunities
  - Heat, Lighting, Base salary, Insurance, Retirement, Air conditioning, Restrooms, Cafeteria, Job security, Health programs, Clean Air, Drinking Water, Safe conditions, No layoffs, and Time off.

Four components—satisfaction progression, frustration, frustration regression, and aspiration—are key to understanding Alderfer's ERG theory.

1. The first of these, *satisfaction progression,* is in basic agreement with Maslow's process of moving through the needs. As we increasingly satisfy our existence needs, we direct energy toward relatedness needs. As these needs are satisfied, our growth needs become more active.
2. The second component, *frustration,* occurs when we attempt but fail to satisfy a particular need. The resulting frustration may make satisfying the unmet need even more important to us—unless we repeatedly fail to satisfy that need.
3. In this case, Alderfer's third component, *frustration regression,* can cause us to shift our attention to a previously satisfied, more concrete, and verifiable need.
4. Lastly, the *aspiration* component of the ERG model notes that, by its very nature, growth is intrinsically satisfying. The more we grow, the more we want to grow. Therefore, the more we satisfy our growth need, the more important it becomes and the more strongly we are motivated to satisfy it.

Alderfer's model is potentially more useful than Maslow's in that it doesn't create false motivational categories. For example, it is difficult for researchers to ascertain when interaction with others satisfies our need for acceptance and when it satisfies our need for recognition. ERG also focuses attention explicitly on movement through the set of needs in both directions. Further, evidence in support of the three need categories and their order tends to be stronger than evidence for Maslow's five need categories and their relative order.

Herzberg's Motivator-Hygiene Theory

Clearly one of the most influential motivation theories throughout the 1950s and 1960s was Frederick Herzberg's motivator-hygiene theory.[14] This theory is a further refinement of Maslow's theory.

Herzberg argued that there are two sets of needs, instead of the five sets theorized by Maslow.

1. He called the first set "motivators" (or growth needs). **Motivators**, which relate to the jobs we perform and our ability to feel a sense of achievement as a result of performing them, are rooted in our need to experience growth and self-actualization.
2. The second set of needs he termed "hygienes." **Hygienes** relate to the work environment and are based in the basic human need to "avoid pain." According to Herzberg, growth needs motivate us to perform well and, when these needs are met, lead to the experience of satisfaction. Hygiene needs, on the other hand, must be met to avoid dissatisfaction (but do not necessarily provide satisfaction or motivation).[15]

Hygiene factors are not directly related to the work itself (job content). Rather, hygienes refer to job context factors (pay, working conditions, supervision, and security). Herzberg also refers to these factors as "dissatisfiers" because they are frequently associated with dissatisfied employees. These factors are so frequently associated with dissatisfaction that Herzberg claims they never really provide satisfaction. When they're present in sufficient quantities, we avoid dissatisfaction, but they do not contribute to satisfaction. Furthermore, since meeting these needs does not provide satisfaction, Herzberg concludes that they do not motivate workers.

Motivator factors involve our long-term need to pursue psychological growth (much like Maslow's esteem and self-actualization needs). Motivators relate to *job content.* Job content is what we actually *do* when we perform our job duties. Herzberg considered job duties that lead to feelings of achievement and recognition to be motivators. He refers to these factors as "satisfiers" to reflect their ability to provide satisfying experiences. When these needs are met, we experience satisfaction. Because meeting these needs provides satisfaction, they motivate workers. More specifically, Herzberg believes these motivators lead to high performance (achievement), and the high performance itself leads to satisfaction.

The unique feature of Herzberg's theory is that job conditions that prevent dissatisfaction do not cause satisfaction. Satisfaction and dissatisfaction are on different "scales" in his view. Hygienes can cause dissatisfaction if they are not present in sufficient levels. Thus, an employee can be dissatisfied with low pay. But paying him more will not cause long-term satisfaction *unless* motivators are present. Good pay *by itself* will only make the employee neutral toward work; to attain satisfaction, employees need challenging job duties that result in a sense of achievement. Employees can be dissatisfied, neutral, or satisfied with their jobs, depending on their levels of hygienes and motivators. Herzberg's theory even allows for the possibility that an employee can be satisfied and dissatisfied at the same time—the "I love my job but I hate the pay" situation!

Herzberg's theory has made lasting contributions to organizational research and managerial practice. Researchers have used it to identify the wide range of factors that influence worker reactions. Previously, most organizations attended primarily to hygiene factors. Because of Herzberg's work, organizations today realize the potential of motivators. Job enrichment programs are among the many direct results of his research.

Herzberg's work suggests a two-stage process for managing employee motivation and satisfaction. First, managers should address the hygiene factors. Intense forms of dissatisfaction distract employees from important work-related activities and tend to be demotivating.[16] Thus, managers should make sure that such basic needs as adequate pay, safe and clean working conditions, and opportunities for social interaction are met. They should then address the much more powerful motivator needs, in which workers experience recognition, responsibility, achievement, and growth. If motivator needs are ignored, neither long-term satisfaction nor high motivation is likely. When motivator needs are met, however, employees feel satisfied and are motivated to perform well.

Self-Determination Theory

One major implication of Herzberg's motivator-hygiene theory is the somewhat counterintuitive idea that managers should focus more on motivators than on hygienes. (After all, doesn't everyone want to be paid well? Organizations have held this out as a chief motivator for decades!) Why might concentrating on motivators give better results? To answer this question, we must examine *types* of motivation. Organizational behavior researchers often classify motivation in terms of what stimulates it. In the case of **extrinsic motivation**, we endeavor to acquire something that satisfies a lower-order need. Jobs that pay well and that are performed in safe, clean working conditions with adequate supervision and resources directly or indirectly satisfy these lower-order needs. These "outside the person" factors are *extrinsic rewards*.

Factors "inside" the person that cause people to perform tasks, **intrinsic motivation**, arise out of performing a task in and of itself because it is interesting or "fun" to do. The task is enjoyable, so we continue to do it *even in the absence* of extrinsic rewards. That is, we are motivated by *intrinsic rewards,* rewards that we more or less give ourselves. Intrinsic rewards satisfy higher-order needs like relatedness and growth in ERG theory. When we sense that we are valuable contributors, are achieving something important, or are getting better at some skill, we like this feeling and strive to maintain it.

Self-determination theory (SDT) seeks to explain not only what causes motivation, but also how extrinsic rewards affect intrinsic motivation.[17] In SDT, extrinsic motivation refers to the performance of an activity in order to attain some valued outcome, while intrinsic motivation refers to performing an activity for the inherent satisfaction of the activity itself. SDT specifies when an activity will be intrinsically motivating and when it will not. Considerable numbers of studies have demonstrated that tasks are intrinsically motivating when they satisfy at least one of three higher-order needs: competence, autonomy, and relatedness. These precepts from SDT are entirely consistent with earlier discussions of theories by McClelland, Maslow, Alderfer, and Herzberg.

SDT takes the concepts of extrinsic rewards and intrinsic motivation further than the other need theories. SDT researchers have consistently found that as the level of extrinsic rewards increases, the amount of intrinsic motivation *decreases*. That is, SDT posits that extrinsic rewards not only do not provide intrinsic motivation, they diminish it. Think of this in terms of hobbies. Some people like to knit, others like to carve wood. They do it because it is intrinsically motivating; the hobby satisfies needs for competence, autonomy, and relatedness. But what happens if these hobbyists start getting paid well for their sweaters and carvings? Over time the hobby becomes less fun and is done in order to receive extrinsic rewards (money). Extrinsic motivation increases as intrinsic motivation decreases! When extrinsic rewards are present, people do not feel like what they do builds competence, is self-determined, or enhances relationships with others.

SDT theory has interesting implications for the management of organizational behavior. Some jobs are by their very nature uninteresting and unlikely to be made interesting. Automation has eliminated many such jobs, but they are still numerous. SDT would suggest that the primary way to motivate high performance for such jobs is to make performance contingent on extrinsic rewards. Relatively high pay is necessary to sustain performance on certain low-skill jobs. On the other hand, SDT would suggest that to enhance intrinsic motivation on jobs that are interesting, don't focus only on increasing extrinsic rewards (like large pay bonuses). Instead, create even more opportunities for employees to satisfy their needs for competence, autonomy, and relatedness. That means giving them opportunities to learn new skills, to perform their jobs without interference, and to develop meaningful relationships with other customers and employees in other departments. Such actions enhance intrinsic rewards.

You may have noticed that content theories are somewhat quiet about what determines the intensity of motivation. For example, some people steal to satisfy their lower-order needs (they have high intensity). But most of us don't steal. Why is this? Process theories of motivation attempt to explain this aspect of motivation by focusing on the intensity of motivation as well as its direction. According to self-determination theory, skilled workers who are given a chance to hone their skills and the freedom to practice their craft will be intrinsically motivated.

Concept check

1. Understand the content theories of motivation.
2. Understand the contributions that Murray, McClelland, Maslow, Alderfer, and Herzberg made toward an understanding of human motivation.

References

1. J.E. Hunter & R.E. Hunter. 1984. Validity and utility of alternative predictors of job performance. *Psychological Bulletin* 96: 72–98.
2. Statistics on the prevalence of this choice are available. "Calling in Well: A Look at leave Time Tracking Trends," *actiTIME* website, June 2016. Leave time tracking trends.
3. H. A. Murray. 1938. *Explorations in personality.* New York: Oxford University Press.
4. Murray also hypothesized that people would differ in the degree to which they felt these needs. His list of secondary needs became a basis for his theory of personality.
5. Representative references include J.W. Atkinson & D.C. McClelland. 1948. The projective expression of needs. II. The effect of different intensities of the hunger drive on thematic apperception. *Journal of Experimental Psychology* 38:643–658; D.C. McClelland, J.W. Atkinson, R.A. Clark, & E.L. Lowell. 1953. *The achievement motive.* New York: Appleton-Century-Crofts; R.C. DeCharms. 1957. Affiliation motivation and productivity in small groups. *Journal of Abnormal Psychology* 55:222– 276; D.C. McClelland. 1961. *The achieving society.* Princeton, NJ: Van Nostrand; and D.C. McClelland. 1975. *Power: The inner experience.* New York: Irvington.
6. In fact, McClelland argued that the success of entire societies is dependent on its achievement needs.
7. D. C. McClelland. 1970. The two faces of power. *Journal of International Affairs* 24:29–47.
8. A.H. Maslow. 1943. A theory of human motivation. *Psychological Bulletin* 50:370–396; A.H. Maslow. 1954. *Motivation and personality.* New York: Harper & Row; A. H. Maslow. 1965. *Eupsychian management.* Homewood, IL: Irwin.
9. D. McGregor. 1960. *The human side of enterprise.* New York: McGraw-Hill; D. McGregor. 1967. *The professional manager.* New York: McGraw-Hill.
10. Maslow, 1943, 382.
11. C.P. Alderfer. 1972. *Existence, relatedness, and growth: Human needs in organizational settings.* New York: Free Press.
12. D.T. Hall & K.E. Nougaim. 1968. An examination of Maslow's need hierarchy in an organizational setting. *Organizational Behavior and Human Performance* 3:12–35; E.E. Lawler, III & J.L. Suttle. 1972. A causal correlational test of the need hierarchy concept. *Organizational Behavior and Human Performance* 7:265–287; M.A. Wahba & L.G. Bridwell. 1973. Maslow reconsidered: A review of research on the need hierarchy theory. *Proceedings of the thirty-third annual meeting of the Academy of Management,* 514–520.
13. C.P. Alderfer. 1972. *Existence, relatedness, and growth: Human needs and organizational settings.* New York: Free Press.
14. Note that Herzberg's theory has often been labeled the "two-factor theory" because it focuses on two continua. This name, however, implies that only two factors are involved, which is not correct. Herzberg prefers not to use the term "two-factor theory" because his two *sets* of needs identify a much larger *number* of needs.
15. F. Herzberg, B. Mausner, & B. Snyderman. 1959. *The motivation to work.* New York: Wiley; F. Herzberg. 1966. *Work and the nature of man.* New York: Crowell; F. Herzberg. 1968. One more time: How do you motivate employees? *Harvard Business Review* 46:54–62.
16. R.B. Dunham, J.L. Pierce, & J.W. Newstrom. 1983. Job context and job content: A conceptual perspective. *Journal of Management* 9:187–202.
17. R.M. Ryan & E.L. Deci. 2000. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist* 55:68–78.
18. B.F. Skinner. 1953. *Science and human behavior.* New York: Free Press; B.F. Skinner. 1969. *Contingencies of reinforcement.* East Norwalk, CT: Appleton Century-Crofts; B.F. Skinner. 1971. *Beyond freedom and dignity.* New York: Bantam Books.
19. Ibid.
20. R. W. Kempen & R. V. Hall. 1977. Reduction of industrial absenteeism: Results of a behavioral approach. *Journal of Organizational Behavior Management* 20:1–21.
21. J.S. Adams. 1965. Inequity in social exchange. In L. Berkowitz (ed.), *Advances in experimental social psychology* (Vol. 2). New York: Academic Press; G.C. Homans. 1961. *Social behavior: Its elementary forms.* New York: Harcourt, Brace, & World.
22. Ibid.
23. J. Kane & E.E. Lawler, III. 1979. Performance appraisal effectiveness. In B. Staw (ed.), *Research in organizational behavior* (Vol. 1). Greenwood, CT: JAI Press.
24. E.E. Lawler, III. 1972. Secrecy and the need to know. In M. Dunnette, R. House, & H. Tosi (eds.), *Readings in managerial*

motivation and compensation. East Lansing: Michigan State University Press.
25. I.R. Andrews. 1967. Wage inequity and job performance: An experimental study. *Journal of Applied Psychology* 51:39–45; J.S. Adams. 1963a. Towards an understanding of inequity. *Journal of Abnormal Social Psychology* 67:422–436; J.S. Adams. 1963b. Wage inequities, productivity and work quality. *Industrial Relations* 3:9–16.
26. R.C. Huseman., J.D. Hatfield, & E.W. Miles. 1987. A new perspective on equity theory: The equity sensitivity construct. *Academy of Management Review* 12:222–234; E.W. Miles, J.D. Hatfield, & R.C. Huseman. 1989. The equity sensitivity construct: Potential implications for worker performance. *Journal of Management* 15:581–588.
27. R.J. Bies. 1987. The predicament of justice: The management of moral outrage. In B.M. Staw & L.L. Cummings (eds.), *Research in organizational behavior* (Vol. 9). Greenwich, CT: JAI Press, 289–319; J. Greenberg. 1987. A taxonomy of organizational justice theories. *Academy of Management Review* 12:9–22.
28. E.L. Locke. 1978. The ubiquity of the technique of goal setting in theories of and approaches to employee motivation. *Academy of Management Review* 3:594–601; F.W. Taylor. 1911. *The principles of scientific management*. New York: Norton; K. Lewin. 1935. *A dynamic theory of personality*. New York: McGraw-Hill; K. Lewin. 1938. *The conceptual representation and the measurement of psychological forces*. Durham, NC: Duke University Press; K. Lewin, T. Dembo, L. Festinger, & P.S. Sears. 1944. Level of aspiration. In J. McVicker Hunt (ed.), *Personality and behavior disorders*. New York: Ronald Press, 333–378; P. Drucker. 1954. *The practice of management*. New York: Wiley; D. McGregor. 1957. An uneasy look at performance appraisal. *Harvard Business Review* 35:89–94; E.A. Locke. 1968. Toward a theory of task motivation and incentives. *Organizational Behavior and Human Performance* 3:157–189; E.A. Locke, K.N. Shaw, L.M. Saari, & G.P. Latham. 1981. Goal setting and task performance: 1969– 1980. *Psychological Bulletin* 90:125–152; G. P. Latham & E.A. Locke. 1984. *Goal setting: A motivational technique that works!* Englewood Cliffs, NJ: Prentice Hall.
29. C.C. Pinder. 1984. *Work motivation: Theory, issues, and applications*. Glenview, IL: Scott, Foresman.
30. Locke, 1979.
31. T.R. Mitchell & W.S. Silver. 1990. Individual and group goals when workers are interdependent: Effects on task strategies and performance. *Journal of Applied Psychology* 75:185–193.
32. A. Bandura. 1977. Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review* 84:191–215; A. Bandura. 1986b. The explanatory and predictive scope of self- efficacy theory. *Journal of Social and Clinical Psychology* 4:359– 373; A. Bandura. 1997. *Self-efficacy: The exercise of control*. New York: Freeman.
33. D.G. Gardner & J.L. Pierce. 1998. Self-esteem and self-efficacy within the organizational context: An empirical comparison. *Group and Organization Management* 23:48–70.
34. Locke, 1978.
35. M.L. Ambrose & C.T. Kulik. 1999. Old friends, new faces: Motivation research in the 1990s. *Journal of Management* 25: 231–292.

Note

References are taken from the section Work Motivation for Performance from:

References - Organizational Behavior | OpenStax.

Note

Original Source Chapter 7.2 Content Theories of Motivation

7.2: Content Theories of Motivation - Business LibreText

Edited for use in CIS 79 Managing Technology Projects.

- Edited for readability.
- Corrected label for Table.
- Restored figures that were damaged by the fork process.
- Removed pictures that were not relevant.
- Added accessibility material for: Alderfer's ERG Theory and Satisfying Existence, Relatedness, and Growth Needs.
- Fixed heading levels.
- Changed References to a numbered list for accessibility.

Caution

After the fork operation, Exhibits were replaced by LibreTexts with irrelevant material. This needed to be repaired. Remixed in the original article as an appendix in order to retrieve the correct material. Created new Figure using template, bringing in the correct figure.

This page titled 4.6.2.1: Content Theories of Motivation is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax.

4.6.2.2: Leadership

Note

The source document for this section on Leadership is Chapter 13 of the following textbook:

Principles of Management (OpenStax) - Business LibreTexts.

This page titled 4.6.2.2: Leadership is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

4.6.2.2.1: Introduction

Learning Objectives

After reading this chapter, you should be able to answer these questions:

1. What is the nature of leadership and the leadership process?
2. What are the processes associated with people coming to leadership positions?
3. How do leaders influence and move their followers to action?
4. What are the trait perspectives on leadership?
5. What are the behavioral perspectives on leadership?
6. What are the situational perspectives on leadership?
7. What does the concept “substitute for leadership” mean?
8. What are the characteristics of transactional, transformational, and charismatic leadership?
9. How do different approaches and styles of leadership impact what is needed now?

Exploring Managerial Careers

John Arroyo: Springfield Sea Lions

John Arroyo is thrilled with his new position as general manager of the Springfield Sea Lions, a minor league baseball team in. Arroyo has been a baseball fan all of his life, and now his diligent work and his degree in sports management are paying off.

Arroyo knew he had a hard act to follow. The general manager whom John replaced, “T.J.” Grevin, was a much-loved old-timer who had been with the Sea Lions since their inception 14 years ago. John knew it would be difficult for whoever followed T.J., but he didn’t realize how ostracized and powerless he would feel. He tried a pep talk: “I’m the general manager—the CEO of this ball club! In time, the staff will respect me.” [Not a very good pep talk!]

After his first season ends, Arroyo is discouraged. Ticket and concession sales are down, and some longtime employees are rumored to be thinking about leaving. If John doesn’t turn things around, he knows his tenure with the Sea Lions will be short.

Questions: Is John correct in assuming that the staff will learn to respect him in time? What can John do to earn the loyalty of his staff and improve the ball club’s performance?

Outcomes: During the winter, John thinks long and hard about how he can earn the respect of the Sea Lions staff. Before the next season opener, John announces his plan: “So I can better understand what your day is like, I’m going to spend one day in each of your shoes. I’m trading places with each of you. I will be a ticket taker, a roving hot dog vendor, and a janitor. And I will be a marketer, and an accountant—for a day. You in turn will have the day off so you can enjoy the game from the general manager’s box.” The staff laughs and whistles appreciatively. Then the Springfield mascot, Sparky the Sea Lion, speaks up: “Hey Mr. Arroyo, are you going to spend a day in my flippers?” “You bet!” says John, laughing. The entire staff cheers.

John continues. “At the close of the season, we will honor a staff member with the T.J. Grevin Award for outstanding contributions to the Sea Lions organization. T.J. was such a great guy, it’s only right that we honor him.” The meeting ends, but John’s staff linger to tell him how excited they are about his ideas. Amidst the handshakes, he hopes that this year may be the best year yet for the Sea Lions.

Sarah Elizabeth Roisland is the manager of a district claims office for a large insurance company. Fourteen people work for her. The results of a recent attitude survey indicate that her employees have extremely high job satisfaction and motivation. Conflict is rare in Sarah’s office. Furthermore, productivity measures place her group among the most productive in the entire company. Her success has brought the company’s vice president of human resources to her office in an attempt to discover the secret to her success. Sarah’s peers, superiors, and workers all give the same answer: she is more than a good manager—she is an outstanding leader. She continually gets high performance from her employees and does so in such a way that they enjoy working for her.

There is no magic formula for becoming a good leader. There are, however, many identifiable reasons why some people are better and more effective leaders. Leaders, especially effective leaders, are not created by simply attending a one-day leadership workshop. Yet effective leadership skills are not something most people are born with. You can become an effective leader if you are willing to invest the time and energy to develop all of the “right stuff.”

According to Louise Axon, director of content strategy, and her colleagues at Harvard Business Publishing, in seeking management talent, *leadership* is an urgently needed quality in all managerial roles.[1] Good leaders and good leadership are rare. Harvard management professor John P. Kotter notes that “there is a leadership crisis in the U.S. today,”[2] and the late USC Professor Warren Bennis states that many of our organizations are overmanaged and underled.[3]

References

1. Louise Axon, Elisa Friedman, and Kathy Jordan. 2015 (July). Leading Now: Critical Capabilities for a Complex World. *Harvard Business Publishing*, (Accessed July 25, 2017) www.harvardbusiness.org/leadi...-complex-world.
2. K. Labich. 1988 (Oct. 24). The seven keys to business leadership.Fortune,58.
3. W. Bennis. 1989. *Why leaders can’t lead*. San Francisco: Jossey-Bass.

Note

Changes made to original OER material:

- Forked only to make accessibility fixes (changed References to a numbered list).

This page titled 4.6.2.2.1: Introduction is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 16.1: Curve Fitting by Troy Siemers is licensed CC BY-NC 3.0. Original source: https://drive.google.com/file/d/0B2Pdg3LqVPNPLUhLbks0QzF5ZVhmWGp3YW5hYlVBd3RDN2x3/view?resourcekey=0-fbyowylmkO1jBbS2ehDP0g&usp=sharing.

4.6.2.2.2: The Nature of Leadership

Learning Objectives

1. What is the nature of leadership and the leadership process?

The many definitions of leadership each have a different emphasis. Some definitions consider leadership an act or behavior, such as initiating structure so group members know how to complete a task. Others consider a leader to be the center or nucleus of group activity, an instrument of goal achievement who has a certain personality, a form of persuasion and power, and the art of inducing compliance.[4] Some look at leadership in terms of the management of group processes. In this view, a good leader develops a vision for the group, communicates that vision,[5] orchestrates the group's energy and activity toward goal attainment, "[turns] a group of individuals into a team," and "[transforms] good intentions into positive actions."[6]

Leadership is frequently defined as a social (interpersonal) influence relationship between two or more persons who depend on each other to attain certain mutual goals in a group situation.[7] Effective leadership helps individuals and groups achieve their goals by focusing on the group's *maintenance needs* (the need for individuals to fit and work together by having, for example, shared norms) and *task needs* (the need for the group to make progress toward attaining the goal that brought them together).



Figure 4.6.2.2.2.1: Joe Maddon, manager of the Chicago Cubs, at pitcher's mound making a pitching change (Joe Maddon at the mound)

Joe Maddon, manager of the Chicago Cubs baseball team shown in Figure 4.6.2.2.2.1, is lauded for both his managerial and leadership skills. Maddon is a role model for managers competing in the business world. Managers can learn and profit from the Cubs skipper's philosophy of instilling an upbeat attitude with the team, staying loose but staying productive, and avoiding being the center of attention

Leader versus Manager

The two dual concepts, leader and manager, leadership and management, are not interchangeable, nor are they redundant. The differences between the two can, however, be confusing. In many instances, to be a good manager one needs to be an effective leader. Many CEOs have been hired in the hope that their leadership skills, their ability to formulate a vision and get others to "buy into" that vision, will propel the organization forward. In addition, effective leadership often necessitates the ability to manage—to set goals; plan, devise, and implement strategy; make decisions and solve problems; and organize and control. For our purposes, the two sets of concepts can be contrasted in several ways.

First, we define the two concepts differently. Management is defined as a process consisting of planning, organizing, directing, and controlling. Leadership is defined as a social (interpersonal) influence relationship between two or more people who are dependent on each another for goal attainment.

Second, managers and leaders are commonly differentiated in terms of the processes through which they initially come to their position. Managers are generally appointed to their role. Even though many organizations appoint people to positions of leadership, leadership per se is a relationship that revolves around the followers' acceptance or rejection of the leader.[8] Thus, leaders often emerge out of events that unfold among members of a group.

Third, managers and leaders often differ in terms of the types and sources of the power they exercise.

- Managers commonly derive their power from the larger organization. Virtually all organizations legitimize the use of certain "carrots and sticks" (rewards and punishments) as ways of securing the compliance of their employees. In other words, by virtue of the position that a manager occupies (president, vice president, department head, supervisor), certain "rights to act" (schedule production, contract to sell a product, hire and fire) accompany the position and its place within the hierarchy of authority.
- Leaders can also secure power and the ability to exercise influence using carrots and sticks; however, it is much more common for leaders to derive power from followers' perception of their knowledge (expertise), their personality and attractiveness, and the working relationship that has developed between leaders and followers.

From the perspective of those who are under the leader's and manager's influence, the motivation to comply often has a different base. The subordinate to a manager frequently complies because of the role authority of the manager, and because of the carrots and sticks that managers have at their disposal. The followers of a leader comply because they want to. Thus, leaders motivate primarily through intrinsic processes, while managers motivate primarily through extrinsic processes.

Finally, it is important to note that while managers may be successful in directing and supervising their subordinates, they often succeed or fail because of their ability or inability to lead.[9] As noted above, effective leadership often calls for the ability to manage, and effective management often requires leadership.

Concept Check

1. What is the nature of leadership and the leadership process?

References

4. B.M. Bass. 1990. *Bass & Stogdill's handbook of leadership: Theory, research, and managerial applications*. New York: The Free Press.
5. W. Bennis. 1989. *Why leaders can't lead*. San Francisco: Jossey-Bass; W. Bennis, & B. Nanus. 1985. *Leaders: The strategies for taking charge*. New York: Harper & Row.
6. T.B. Pickens, Jr. 1992 (Fall/Winter). Pickens on leadership.Hyatt Magazine,21.
7. E.P. Hollander & J.W. Julian. 1969. Contemporary trends in the analysis of leadership process. *Psychological Bulletin* 7(5): 387–397.
8. E.P. Hollander. 1964. Emergent leadership and social influence. In E.P. Hollander (ed.), *Leaders, groups, & influence*. New York: Oxford University Press.
9. F.E. Fiedler. 1996. Research on leadership selection and training: One view of the future. *Administrative Science Quarterly* 41:241–250.

Note

Changes made:

- Changed Exhibit to a Figure using LibreTexts template.
- Deleted reference to "Chapter 1" as it is not applicable to this textbook.
- Converted some text into a bulleted list for readability.

- Changed References to a numbered list per accessibility check.

This page titled 4.6.2.2.2: The Nature of Leadership is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

4.6.2.2.3: The Leadership Process

Learning Objectives

1. What are the processes associated with people coming to leadership positions?

Leadership is a process, a complex and dynamic exchange relationship built over time between leader and follower and between leader and the group of followers who depend on each other to attain a mutually desired goal.[10] There are several key components to this “working relationship”: the leader, the followers, the context (situation), the leadership process per se, and the consequences (outcomes) (see Figure 4.6.2.2.3.1).[11] Across time, each component interacts with and influences the other components, and whatever consequences (such as leader-follower trust) are created influence future interactions. As any one of the components changes, so too will leadership.[12]

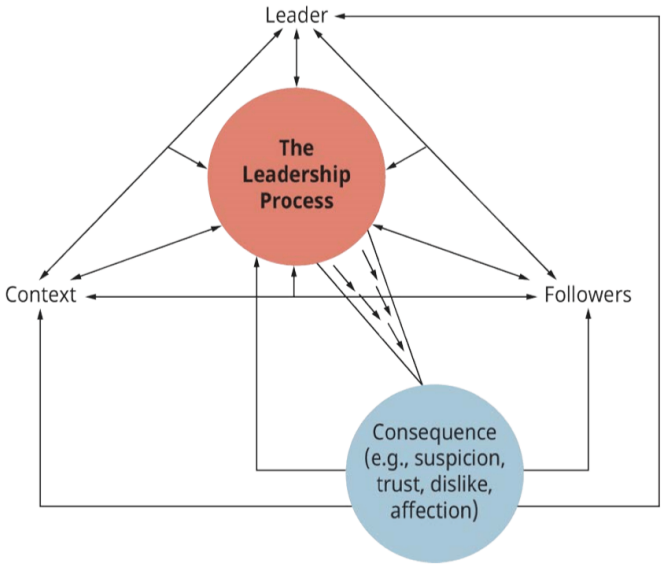


Figure 4.6.2.2.3.1: The Leadership Process (Attribution: Copyright Rice University, OpenStax, under CC-BY 4.0 license)

Accessibility Note

Figure 4.6.2.2.3.1 entitled "The Leadership Process" depicts the following five entities with interconnecting lines as appropriate: Leader, Followers, Context (situation), "The Leadership Process" in a blue circle, and Consequences (e.g., suspicion, trust, dislike, affection) in a blue circle.

All entities have directed arrows to The Leadership Process. There are directed arrows between Leader and Context, Leader and Followers, between Consequence and Leader, Consequence and Context, Context and Followers. There is a stream of directed arrows from "The Leadership Process" and "Consequence."

The Leader

Leaders are people who take charge of or guide the activities of others. They are often seen as the focus or orchestrater of group activity, the people who set the tone of the group so that it can move forward to attain its goals. Leaders provide the group with what is required to fulfill its maintenance and task-related needs. (Later in the chapter, we will return to the “leader as a person” as part of our discussion of the trait approach to leadership.)



Figure 4.6.2.2.3.2: New York Philharmonic @ UN The New York Philharmonic, conducted by Music Director Alan Gilbert, paid special tribute in the General Assembly Hall to UN Secretary-General Ban Ki-moon as a tribute to his 10-year term. Gilbert is the formal leader of the New York Philharmonic. (New York Philharmonic picture.)

The Follower

The follower is not a passive player in the leadership process. Edwin Hollander, after many years of studying leadership, suggested that the follower is the most critical factor in any leadership event.[13] It is, after all, the follower who perceives the situation and comes to define the needs that the leader must fulfill. In addition, it is the follower who either rejects leadership or accepts acts of leadership by surrendering his power to the leader to diminish task uncertainty, to define and manage the meaning of the situation to the follower, and to orchestrate the follower's action in pursuit of goal attainment.

The follower's personality and readiness to follow determine the style of leadership that will be most effective. For example, individuals with an internal locus of control are much more responsive to participative styles of leadership than individuals with an external locus of control.[14] Individuals with an authoritarian personality are highly receptive to the effectiveness of directive acts of leadership.[15] It is the followers' expectations, as well as their performance-based needs, that determine what a leader must do in order to be effective.

The strength of the follower's self-concept has also been linked to the leadership process.

- High-self-esteem individuals tend to have a strong sense of self-efficacy, that is, a generalized belief they can be successful in difficult situations. They therefore tend to be strongly motivated to perform and persist in the face of adversity.[16] The high-self-esteem follower tends to be responsive to participative styles of leadership.
- Low-self-esteem individuals, who doubt their competence and worthiness and their ability to succeed in difficult situations, function better with supportive forms of leadership. This helps them deal with the stress, frustration, and anxiety that often emerge with difficult tasks. Followers without a readiness to follow, limited by their inability to perform and lack of motivation and commitment, usually need more directive forms of leadership.[17]

Follower behavior plays a major role in determining what behaviors leaders engage in. For example, followers who perform at high levels tend to cause their leaders to be considerate in their treatment and to play a less directive role. Followers who are poor performers, on the other hand, tend to cause their leaders to be less warm toward them and to be more directive and controlling in their leadership style.[18]

The Context

Situations make demands on a group and its members, and not all situations are the same. Context refers to the situation that surrounds the leader and the followers. Situations are multidimensional. We discuss the context as it pertains to leadership in greater detail later in this chapter, but for now let's look at it in terms of the task and task environment that confront the group. Is the task structured or unstructured? Are the goals of the group clear or ambiguous? Is there agreement or disagreement about goals? Is there a body of knowledge that can guide task performance? Is the task boring? Frustrating? Intrinsically satisfying? Is the environment complex or simple, stable or unstable? These factors create different contexts within which leadership unfolds,

and each factor places a different set of needs and demands on the leader and on the followers.

The Process

The process of leadership is separate and distinct from the leader (the person who occupies a central role in the group). The process is a complex, interactive, and dynamic working relationship between leader and followers. This working relationship, built over time, is directed toward fulfilling the group's maintenance and task needs. Part of the process consists of an exchange relationship between the leader and follower. The leader provides a resource directed toward fulfilling the group's needs, and the group gives compliance, recognition, and esteem to the leader. To the extent that leadership is the exercise of influence, part of the leadership process is captured by the surrender of power by the followers and the exercise of influence over the followers by the leader.[19] Thus, the leader influences the followers and the followers influence the leader, the context influences the leader and the followers, and both leader and followers influence the context.

The Consequences

A number of outcomes or consequences of the leadership process unfold between leader, follower, and situation. At the group level, two outcomes are important:

- Have the group's maintenance needs been fulfilled? That is, do members of the group like and get along with one another, do they have a shared set of norms and values, and have they developed a good working relationship? Have individuals' needs been fulfilled as reflected in attendance, motivation, performance, satisfaction, citizenship, trust, and maintenance of the group membership?
- Have the group's task needs been met? That is, there are also important consequences of the leadership process for individuals: attendance, motivation, performance, satisfaction, citizenship, trust, and maintenance of their group membership.

The **leader-member exchange (LMX)** theory of the leadership process focuses attention on consequences associated with the leadership process. The theory views leadership as consisting of a number of dyadic relationships linking the leader with a follower. A leader-follower relationship tends to develop quickly and remains relatively stable over time. The quality of the relationship is reflected by the degree of mutual trust, loyalty, support, respect, and obligation. High- and low-quality relationships between a leader and each of his followers produce in and out groups among the followers. Members of the in group come to be key players, and high-quality exchange relationships tend to be associated with higher levels of performance, commitment, and satisfaction than are low-quality exchange relationships.[20] Attitudinal similarity and extroversion appear to be associated with a high-quality leader-member relationship.[21]

The nature of the leadership process varies substantially depending on the leader, the followers, and the situation and context. Thus, leadership is the function of an interaction between the leader, the follower, and the context.

The leadership context for the leader of a group of assembly line production workers differs from the context for the leader of a self-managing production team and from the context confronted by the lead scientists in a research laboratory. The leadership tactics that work in the first context might fail miserably in the latter two.

Catching the Entrepreneurial Spirit

How a Start-Up Finds the Right Leader

Start-ups, by their very nature, require innovation to bring new products and services to market. Along with establishing a new brand or product, the leader has to develop the relationships and processes that make a company succeed, or risk its early demise. While leading an established firm has its challenges, a start-up requires even more from a leader.

How critical is leadership to a start-up? Ask the four cofounders of the now-defunct PYP (Pretty Young Professionals), a website founded as a source of information for young professional women. What began as four young professional women working on a new start-up ended with hurt feelings and threats of legal action. In 2010, Kathryn Minshew, Amanda Pouchot, Caroline Ghosn, and Alex Cavoulacos decided to create the website and Minshew was named CEO (Cohan 2011a). Lines

blurred about Minshew's authority and the ultimate look, feel, and direction of the website. Ideals about shared leadership, where the company was going, and how it was going to get there ultimately got lost in the power shuffle. By June 2011, passwords were changed and legal actions began, and in August Minshew and Cavoulacos left altogether (Cohan 2011b).

When the legal haggling from PYP was over, Alex Cavoulacos and Kathryn Minshew, joined by Melissa McCreery, tried again. But this time, rather than hoping for the best, they put a leadership plan in place. Minshew was named CEO of the new start-up, The Daily Muse, with Cavoulacos as chief operating officer and McCreery as editor in chief. Rather than trusting to luck, the three cofounders based their team positions on strengths and personalities. Cavoulacos and McCreery agreed that Minshew's outgoing personality and confidence made her the proper choice as CEO (Casserly 2013).

No single trait will guarantee that a person can lead a start-up from idea to greatness, but a survey of successful entrepreneurs does show some common traits. According to David Barbash, a partner at Boston-based law firm Posternak Blankstein & Lund LLP, personality is paramount: "You can have great technology but if you're not a great communicator it may die in the lab" (Casserly 2013 n.p.). A start-up needs a leader who is confident and willing, if not eager, to face the future. According to Michelle Randall, a principal of Enriching Leadership International, start-up CEOs have to be willing to fundraise and not be too proud to beg (Casserly 2013). Peter Shankman, an entrepreneur and angel investor, says leaders have to be willing to make the hard decisions, even risking being the bad guy (Casserly 2013).

Gary Vaynerchuk credits his success to six factors. Angel investor, social media marketer, and early social media adopter, Vaynerchuk leveraged YouTube in its early years to market wine from the family's liquor store, eventually increasing sales from $3 million to $60 million a year (Clifford 2017). Gary believes good leaders recognize that they don't dictate to the market, but rather respond to where it is going. They have respect for and believe in other people, and have a strong work ethic, what Vaynerchuk called a "lunch pail work ethic": they are willing to put in long hours because they love the work, not the perks. He also stresses that he loves technology and doesn't fear it, is obsessed with the youth of today, and is optimistic about people and the future of humanity (Vaynerchuk 2017).

Leading a startup requires more than simple management. It requires the right leader for the right company at the right time, which means matching the right management skills with the proper flexibility and drive to keep it all together and moving in the right direction.

Sources:

1. Casserly, Meghan. 2013. "Rocks, Paper, CEO: Finding The Best Leader For Your Startup." *Forbes*. Forbes article.
2. Clifford, Catherine. 2017. "Self-made millionaire Gary Vaynerchuk: This is the real secret to success." CNBC. CNBC article.
3. Cohan, Peter. 2011a. "A Cautionary Tale: Friendship, Business Ethics, and Bad Breakups (Acts I and II). *Forbes*. August 9, 2011. Forbes Cohan article.
4. Cohan, Peter. 2011b. "A Cautionary Tale: Friendship, Business Ethics, and Bad Breakups (Acts III and IV).Forbes.August 9, 2011. Forbes Cohan article.
5. Vaynerchuk, Gary. 2017. "What Makes Me a Great CEO."

Concept Check

1. What are the processes associated with people coming to leadership positions?
2. Why would start-up leaders need different leadership qualities than someone managing an established firm?
3. Vaynerchuk has been quoted as saying that if you live for Friday, get a different job. How does this apply to successful entrepreneurs?

References

10. Hollander & Julian, 1969.
11. R.M. Stogdill. 1948. Personal factors associated with leadership: A survey of the literature. *Journal of Psychology* 28: 35–71.
12. A.J. Murphy. 1941. A study of the leadership process. *American Sociological Review* 6:674–687.
13. Hollander, 1964.
14. R.J. House & T.R. Mitchell. 1974 (Autumn). Path-goal theory of leadership. *Journal of Contemporary Business* 81–97.
15. G. Yukl. 1971. Toward a behavioral theory of leadership. *Organizational Behavior and Human Performance* 6:414–440.

16. D.G. Gardner & J.L. Pierce. 1998. Self-esteem and self-efficacy within the organizational context. *Group & Organization Management* 23(1):48–70.
17. P. Hersey & K.H. Blanchard. 1988. *Management of organizational behavior utilizing human resources*. Englewood Cliffs, NJ: Prentice-Hall.
18. C.N. Greene. 1975. The reciprocal nature of influence between leader and subordinate. *Journal of Applied Psychology* 60: 187–193.
19. Hollander & Julian, 1969.
20. B.B. Graen & M. Wakabayashi. 1994. Cross-cultural leadership-making: Bridging American and Japanese diversity for team advantage. In M. D. Dunnette (ed.), *Handbook of industrial and organizational psychology*, 4 (2nd ed.): 415–446. Palo Alto: Consulting Psychologists Press.
21. C.A. Schriesheim, S.L. Castro, & F.J. Yammarino. 2000. Investigating contingencies: An examination of the impact of span of supervision and upward controlling on leader-member exchange using traditional and multivariate within- and between-entities analysis. *Journal of Applied Psychology* 85:659–677; A.S. Phillips & A.G. Bedeian. 1994. Leader-follower exchange quality: The role of personality and interpersonal attributes. *Academy of Management Journal* 37:990–1001.

Note

Changes made to this section:

- Changed "Exhibit" to "Figure" using Libretexts template as used elsewhere in this textbook. Added accessibility as needed.
- Converted some text to a bulleted list to improve readability.
- Added Accessibility information.
- Changed References to a numbered list per accessibility review.

This page titled 4.6.2.2.3: The Leadership Process is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

4.6.2.2.4: Types of Leaders and Leader Emergence

Learning Objectives

1. How do leaders influence and move their followers to action?

Leaders hold a unique position in their groups, exercising influence and providing direction. Leonard Bernstein was part of the symphony, but his role as the New York Philharmonic conductor differed dramatically from that of the other symphony members. Besides conducting the orchestra, he created a vision for the symphony. In this capacity, leadership can be seen as a differentiated role and the nucleus of group activity.

Organizations have two kinds of leaders: formal and informal. A **formal leader** is that individual who is recognized by those outside the group as the official leader of the group. Often, the formal leader is appointed by the organization to serve in a formal capacity as an agent of the organization. Jack Welch was the formal leader of General Electric, and Leonard Bernstein was the formal leader of the symphony. Practically all managers act as formal leaders as part of their assigned role. Organizations that use self-managed work teams allow members of the team to select the individual who will serve as their team leader. When this person's role is sanctioned by the formal organization, these team leaders become formal leaders. Increasingly, leaders in organizations will be those who "best sell" their ideas on how to complete a project—persuasiveness and inspiration are important ingredients in the leadership equation, especially in high-involvement organizations.[22]

Informal leaders, by contrast, are not assigned by the organization. The **informal leader** is that individual whom members of the group acknowledge as their leader. Athletic teams often have informal leaders, individuals who exert considerable influence on team members even though they hold no official, formal leadership position. In fact, most work groups contain at least one informal leader. Just like formal leaders, informal leaders can benefit or harm an organization depending on whether their influence encourages group members to behave consistently with organizational goals.

As we have noted, the terms *leader* and *manager* are not synonymous. Grace Hopper, retired U.S. Navy admiral, draws a distinction between leading and managing: "You don't manage people, you manage *things*. You lead *people*."[23] Informal leaders often have considerable leverage over their colleagues. Traditionally, the roles of informal leaders have not included the total set of management responsibilities because an informal leader does not always exercise the functions of planning, organizing, directing, and controlling. However, high-involvement organizations frequently encourage their formal and informal leaders to exercise the full set of management roles. Many consider such actions necessary for self-managing work teams to succeed. Informal leaders are acknowledged by the group, and the group willingly responds to their leadership.

Paths to Leadership

People come to leadership positions through two dynamics. In many instances, people are put into positions of leadership by forces outside the group. University-based ROTC programs and military academies (like West Point) formally groom people to be leaders. We refer to this person as the **designated leader** (in this instance the designated and formal leader are the same person). **Emergent leaders**, on the other hand, arise from the dynamics and processes that unfold within and among a group of individuals as they endeavor to achieve a collective goal.

A variety of processes help us understand how leaders emerge. Gerald Salancik and Jeffrey Pfeffer observe that power to influence others flows to those individuals who possess the critical and scarce resources (often knowledge and expertise) that a group needs to overcome a major problem.[24] They note that the dominant coalition and leadership in American corporations during the 1950s was among engineers, because organizations were engaged in competition based on product design. The power base in many organizations shifted to marketing as competition became a game of advertising aimed at differentiating products in the consumer's mind. About 10–15 years ago, power and leadership once again shifted, this time to people with finance and legal backgrounds, because the critical contingencies facing many organizations were mergers, acquisitions, hostile takeovers, and creative financing. Thus, Salancik and Pfeffer reason that power and thus leadership flow to those individuals who have the ability to help an organization or group [overcome its critical contingencies]. As the challenges facing a group change, so too may the flow of power and leadership.

Many leaders emerge out of the needs of the situation. Different situations call for different configurations of knowledge, skills, and abilities. A group often turns to the member who possesses the knowledge, skills, and abilities that the group requires to achieve its goals.[25] People surrender their power to individuals whom they believe will make meaningful contributions to attaining group goals.[26] The individual to whom power is surrendered is often a member of the group who is in good standing. As a result of this member's contributions to the group's goals, he has accumulated *idiosyncrasy credits* (a form of competency-based status). These credits give the individual a status that allows him to influence the direction that the group takes as it works to achieve its goals.[27]

It is important to recognize that the traits possessed by certain individuals contribute significantly to their emergence as leaders. Research indicates that people are unlikely to follow individuals who, for example, do not display drive, self-confidence, knowledge of the situation, honesty, and integrity.

Leadership as an Exercise of Influence

As we have noted, leadership is the exercise of influence over those who depend on one another for attaining a mutual goal in a group setting. But *how* do leaders effectively exercise this influence? *Social or (interpersonal) influence* is one's ability to effect a change in the motivation, attitudes, and/or behaviors of others. *Power*, then, essentially answers the "how" question: How do leaders influence their followers? The answer often is that a leader's social influence is the source of his power.

French and Raven provide us with a useful typology that identifies the sources and types of power that may be at the disposal of leaders:

- *Reward power*—the power a person has because people believe that he can bestow rewards or outcomes, such as money or recognition that others desire
- *Coercive power*—the power a person has because people believe that he can punish them by inflicting pain or by withholding or taking away something that they value
- *Referent power*—the power a person has because others want to associate with or be accepted by him
- *Expert power*—the power a person has because others believe that he has and is willing to share expert knowledge that they need. (The concept of *resource power* extends the idea of expert power to include the power that a person has because others believe that he possesses and is willing to share resources, such as information, time, or materials that are needed.)
- *Legitimate power*—the power a person has because others believe that he possesses the "right" to influence them and that they ought to obey. This right can originate in tradition; in the charisma or appeal of the person; and in laws, institutional roles within society, moralistic appeal, and rationality (that is, logical arguments, factual evidence, reason, and internally consistent positions).[28]

Not all forms of power are equally effective (see Figure 4.6.2.2.4.1), nor is a leader's total power base the simple sum of the powers at his disposal. Different types of power elicit different forms of compliance: Leaders who rely on coercive power often alienate followers who resist their influence attempts. Leaders who rely on reward power develop followers who are very measured in their responses; the use of rewards often leads people to think in terms of "How much am I getting?" or "How much should I give?" or "Am I breaking even?" The use of referent power produces identification with the leader and his cause. The use of rationality, expert power, and/or moralistic appeal generally elicits commitment and the internalization of the leader's goals.[29]

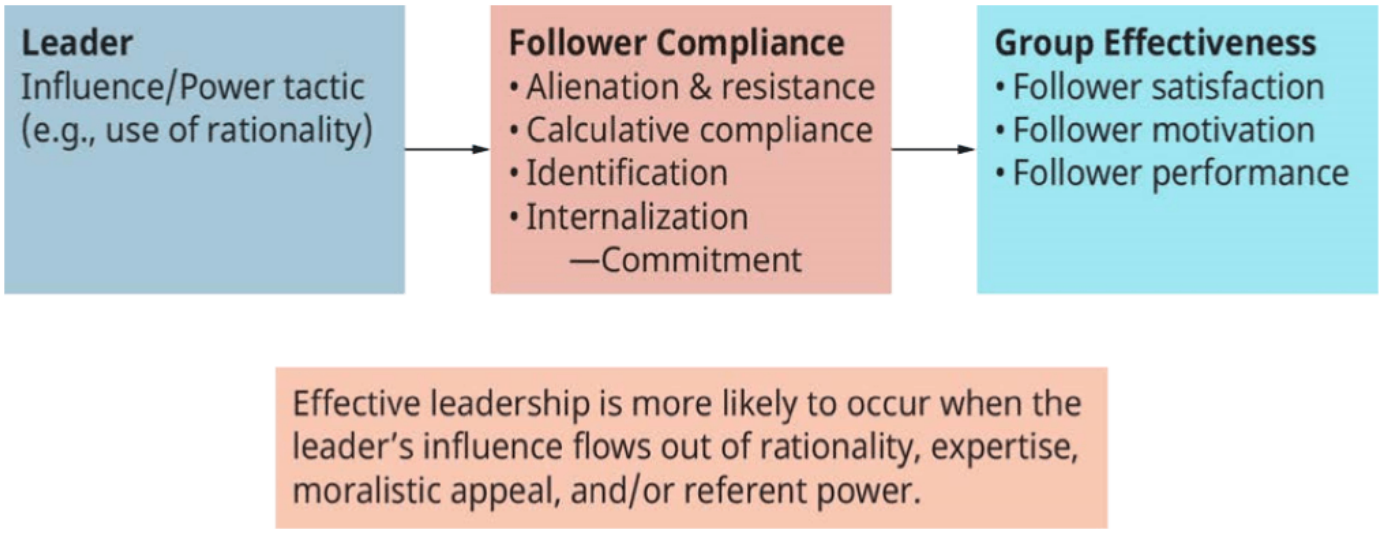


Figure 4.6.2.2.4.1: The Leader-Follower Power Relationship. (Leader-Follower diagram)

Accessibility Note

Figure 4.6.2.2.4.1 has a blue box Entitled "Leader" Influence/Power tactic (e.g., use of rationality) pointing to an orange box entitled "Follower Compliance" with four bullets:

- Alienation & resistance
- Calculative compliance
- Identification
- Internalization - Commitment.

The orange box in turn has an arrow to a light blue box entitled "Group Effectiveness" with three bulleted points:

- Follower satisfaction
- Follower motivation
- Follower performance.

At the bottom is a centered orange box with the following text:

"Effective leadership is more likely to occur when the leader;s influence flows out of rationality, expertise, moralistic appeal, and/or referent power."

Leaders who use referent and expert power commonly experience a favorable response in terms of follower satisfaction and performance. Research suggests that rationality is the most effective influence tactic in terms of its impact on follower commitment, motivation, performance, satisfaction, and group effectiveness.[30]

Reward and legitimate power (that is, relying on one's position to influence others) produce inconsistent results. Sometimes these powers lead to follower performance and satisfaction, yet they also sometimes fail. Coercive power can result in favorable performance, yet follower and resistance dissatisfaction are not uncommon.

Good leaders, whether formal or informal, develop many sources of power. Leaders who rely solely on their legitimate power and authority seldom generate the influence necessary to help their organization and its members succeed. In the process of building their power base, effective leaders have discovered that the use of coercive power tends to dilute the effectiveness of other powers, while the development and use of referent power tends to magnify the effectiveness of other forms of power. A compliment or reward from a person we like generally has greater value than one from someone we dislike, and punishment from someone we love (such as "tough love" from a parent) is less offensive than the pain inflicted by someone we dislike.[31]

In sum, one key to effective leadership, especially as it pertains to the exercise of social and interpersonal influence, relates to the

type of power employed by the leader. Overall leader effectiveness will be higher when people follow because they want to follow. This is much more likely to happen when the leader's influence flows out of intrinsic such as rationality, expertise, moralistic appeal, and/or referent power.

Leadership is also about having a vision and communicating that vision to others in such a way that it provides meaning for the follower.[32] Language, ritual, drama, myths, symbolic constructions, and stories are some of the tools leaders use to capture the attention of their "followers to be" to evoke emotion and to manage the meaning "of the task (challenges) facing the group."[33] These tools help the leader influence the attitudes, motivation, and behavior of their followers.

Influence-Based Leadership Styles

Many writers and researchers have explored how leaders can use power to address the needs of various situations. One view holds that in traditional organizations members expect to be told what to do and are willing to follow highly structured directions. Individuals attracted to high-involvement organizations, however, want to make their own decisions, expect their leaders to allow them to do so, and are willing to accept and act on this responsibility. This suggests that a leader may use and employ power in a variety of ways.

The Tannenbaum and Schmidt Continuum

In the 1950s, Tannenbaum and Schmidt created a continuum (see Figure 4.6.2.2.4.2) along which leadership styles range from authoritarian to extremely high levels of worker freedom.[34] Subsequent to Tannenbaum and Schmidt's work, researchers adapted the continuum by categorizing leader power styles as *autocratic* (boss-centered),participative(workers are consulted and involved), or *free-rein* (members are assigned the work and decide on their own how to do it; the leader relinquishes the active assumption of the role of leadership).[35]

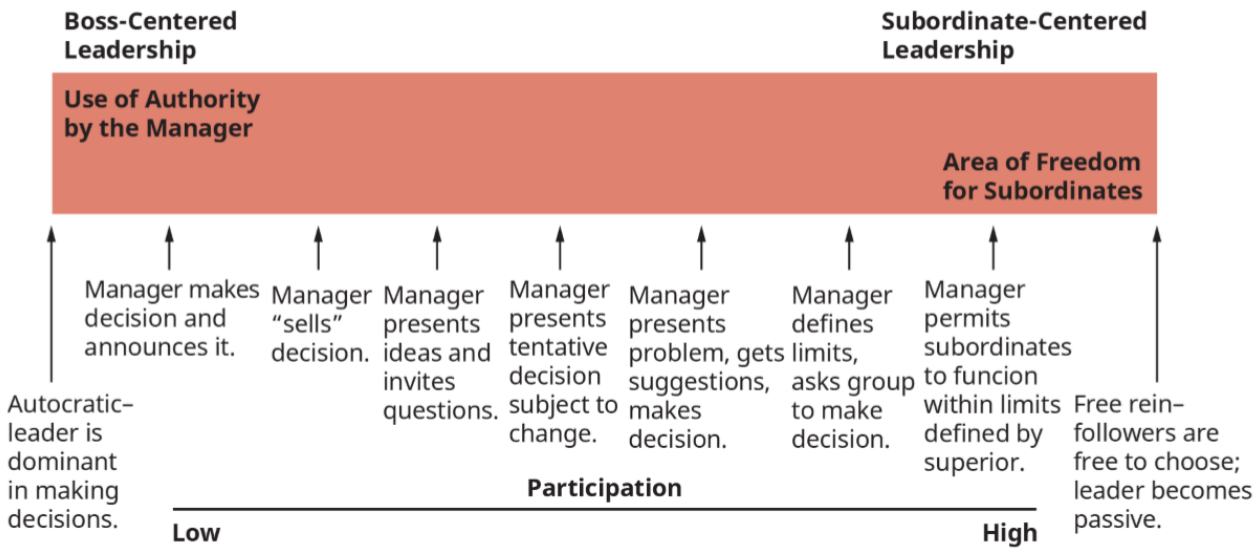


Figure 4.6.2.2.4.2:Tannenbaum and Schmidt's Leadership Continuum (Source:Modified from R. Tannenbaum and W. H. Schmidt. May—June 1971. How to choose a leadership pattern.Harvard Business Review, 167.)

Accessibility Note

Figure 4.6.2.2.4.2 Tannenbaum and Schmidt's Leadership Continuum has an orange box with "Use of Authority by the Manager" on the left and "Area of Freedom for Subordinates on the right. On top of the orange box the words "Boss-centered Leadership" are on the left, and "Subordinate-Centered Leadership" on the right.

Underneath the orange box are many directed arrows pointing to the orange box, with the following text for each arrow, going from left to right:

- Autocratic-leader is dominant in making decisions.
- Manager makes decision and announces it.
- Manager "sells" decision.
- Manager presents ideas and invites questions.
- Manager presents tentative decision subject to change.
- Manager presents problem, gets suggestions, makes decision.
- Manager defines limits, asks group to make decision.
- Manager permits subordinates to function within limits defined by superior.
- Free-rein-followers are free to choose; leader becomes passive.

At the bottom of the figure is a line labeled "Participation" with "Low" on the left, and "High" on the right.

Theory X and Theory Y Leaders

McGregor's Theory X and Theory Y posits two different sets of attitudes about the individual as an organizational member.[36] Theory X and Y thinking gives rise to two different styles of leadership.

- The *Theory X leader* assumes that the average individual dislikes work and is incapable of exercising adequate self-direction and self-control. As a consequence, they exert a highly controlling leadership style.
- In contrast,*Theory Y leaders* believe that people have creative capacities, as well as both the ability and desire to exercise self-direction and self-control. They typically allow organizational members significant amounts of discretion in their jobs and encourage them to participate in departmental and organizational decision-making. Theory Y leaders are much more likely to adopt involvement-oriented approaches to leadership and organically designed organizations for their leadership group.

Theory X and Theory Y thinking and leadership are not strictly an American phenomenon. Evidence suggests that managers from different parts of the global community commonly hold the same view. A study of 3,600 managers from 14 countries reveals that most of them held assumptions about human nature that could best be classified as Theory X.[37] Even though managers might publicly endorse the merits of participatory management, most of them doubted their workers' capacities to exercise self-direction and self-control and to contribute creatively.[38]

Directive/Permissive Leadership Styles

Contemplating the central role of problem-solving in management and leadership, Jan P. Muczyk and Bernard C. Reimann of Cleveland State University offer an interesting perspective on four different leadership styles (see Exhibit 13.7) that revolve around decision-making and implementation processes.[39]

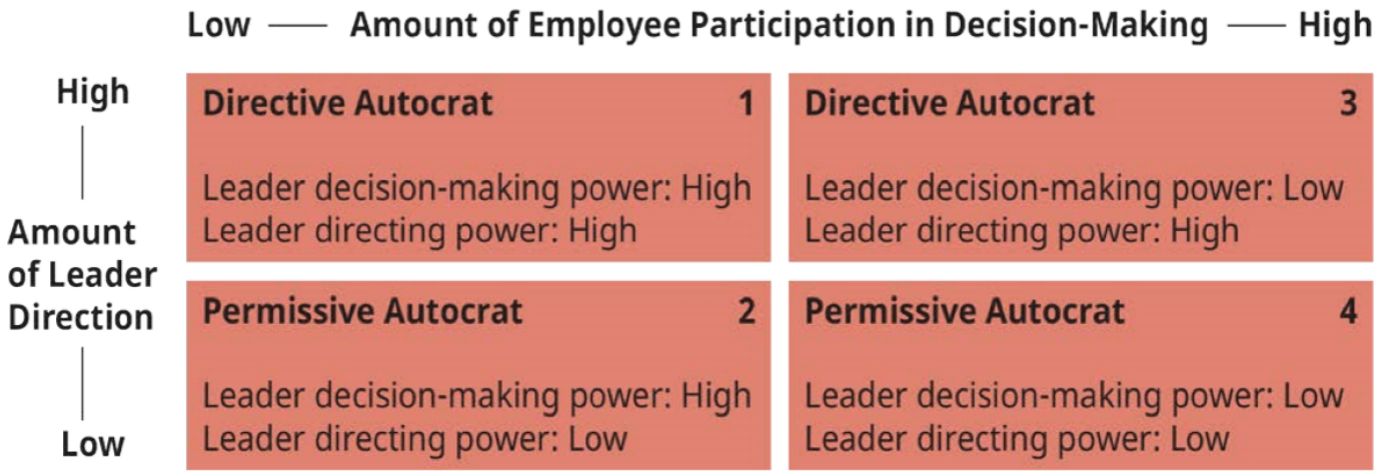


Figure 4.6.2.2.4.3: Leadership Behavior and the Uses of Power Source:Modified from J. P. Muczyk and B. C. Reimann. 1987. The case for directive leadership. Academy of Management Executive, 1:304.

Accessibility Note

Figure 4.6.2.2.4.3: Leadership Behavior and the Uses of Power is a quadrant diagram with a horizontal label across the top stating "Low --- Amount of Employee Participation in Decision Making -- High" and a vertical label stating "High -- Amount of Leader Direction -- Low".

Two orange boxes are stacked on the left going from top to bottom:

- Directive Autocrat 1: Leader decision-making power: High. Leader directing power: High.
- Permissive Autocrat 2 Leader decision-making power: High. Leader directing power: Low.

Two orange boxes are stacked on the right going from top to bottom:

- Directive Autocrat 3: Leader decision-making power: Low. Leader directing power: High.
- Permissive Autocrat 4: Leader decision-making power: Low. Leader directing power: Low.

A *directive autocrat* retains power, makes unilateral decisions, and closely supervises workers' activities. This style of leadership is seen as appropriate when circumstances require quick decisions and organizational members are new, inexperienced, or under-qualified. A doctor in charge of a hastily constructed shelter for victims of a tornado may use this style to command nonmedical volunteers.

The *permissive autocrat* mixes his or her use of power by retaining decision-making power but permitting organizational members to exercise discretion when executing those decisions. This leader behavior is recommended when decision-making time is limited, when tasks are routine, or when organizational members have sufficient expertise to determine appropriate role behaviors.

Also sharing power is the *directive democrat*, who encourages participative decision-making but retains the power to direct team members in the execution of their roles. This style is appropriate when followers have valuable opinions and ideas, but one person needs to coordinate the execution of the ideas. A surgeon might allow the entire surgical team to participate in developing a plan for a surgical procedure. Once surgery begins, however, the surgeon is completely in charge.

Finally, the *permissive democrat* shares power with group members, soliciting involvement in both decision-making and execution. This style is appropriate when participation has both informational and motivational value, when time permits group decision-making, when group members are capable of improving decision quality, and when followers are capable of exercising self-management in their performance of work.

The permissive democratic approach to leadership is characteristic of leadership in high-involvement organizations. Here, leaders act as facilitators, process consultants, network builders, conflict managers, inspirationalists, coaches, teachers/mentors, and cheerleaders.[40] Such is the role of Ralph Stayer, founder, owner, and CEO of Johnsonville Foods. He defines himself as his company's philosopher. At Quad/Graphics, president Harry V. Quadracci is a permissive democrat because he encourages all Quad employees to play a major role in decision-making and execution as they manage their teams as independent profit centers.



Figure 4.6.2.2.4.4: Jeff Bezos, founder and CEO of Amazon, used to bring an empty chair to meetings to signal and remind participants of the most important people that did not have a seat at the table: the customers. He has now replaced the empty chair with Amazon employees with the job title Customer Experience Bar Raisers. (Bezos picture)

Concept Check

1. What is the role of the leader and follower in the leadership process?
2. How do the theories of Tannenbaum and Schmidt's leadership continuum and McGregor's Theory X and Theory Y attempt to define leadership?

References

22. J.A. Conger. 1993. The brave new world of leadership training. *Organizational Dynamics* 21(3):46–59.
23. Pickens, 1992, 21.
24. G.R. Salancik & J. Pfeffer. 1977 (Winter). Who gets power and how they hold on to it: A strategic contingency model of power. *Organizational Dynamics*, 3–21.
25. A.J. Murphy. 1941. A study of the leadership process. *American Sociological Review* 6:674–687.
26. L. Smircich & G. Morgan. 1982. Leadership: The management of meaning. *Journal of Applied Behavioral Science* 18(3): 257–273; Stogdill, 1948.
27. Hollander, 1964.
28. J. R. P. French, Jr. & B. Raven. 1959. The bases of social power. In D. Cartwright (ed.), *Studies in social power*. Ann Arbor, MI: Institute for Social Research, University of Michigan, 150– 167.
29. A. Etzioni. 1961. *A comparative analysis of complex organizations, on power, involvement, and their correlates*. New York: Free Press of Glenco; H. C. Kelman. 1958. Compliance, identification, and internalization: Three processes of attitude change.Journal of Conflict Resolution, 51–61.
30. G. Yukl & J. B. Tracey. 1992. Consequences of influence tactics used with subordinates, peers, and the boss. *Journal of Applied Psychology* 77:525–535; T.R. Hinkin & CA. Schriesheim. 1990. Relationships between subordinate perceptions of supervisor influence tactics and attributed bases of supervisory power.Human Relations43:221–237; P.M. Podsakoff & C.A. Schriesheim. 1985. Field studies of French and Raven's bases of power: Critique, reanalysis, and suggestions for future research. *Psychological Bulletin* 97:398–411.
31. T.R. Hinkin & C.A. Schriesheim. 1990. Relationships between subordinate perceptions of supervisor influence tactics and attributed based of supervisory power. *Human Relations* 43:221–237.
32. Bennis, 1989.
33. L. Smircich & G. Morgan. 1982. Leadership: The management of meaning. *Journal of Applied Behavioral Sciences* 18(3): 257–273.
34. R. Tannenbaum & W.H. Schmidt. 1958 (Mar.–Apr.). How to choose a leadership pattern. *Harvard Business Review*, 95–101; R.

Tannenbaum & W.H. Schmidt. 1973 (May–June). How to choose a leadership pattern. *Harvard Business Review*, 162–175.
35. K. Davis & J.W. Newstrom. 1985. *Human behavior at work: Organization behavior*. New York: McGraw-Hill.
36. D. McGregor. 1957. The human side of enterprise, *Management Review* 46:22–28, 88–92; D. McGregor. 1960. *The human side of enterprise*. New York: McGraw-Hill.
37. M. Haire, E.E. Ghiselli, & L.W. Porter. 1966. *Managerial thinking: An international study*. New York: Wiley.
38. R.E. Miles. 1975. *Theories of management: Implications for organizational behavior and development*. New York: McGraw-Hill.
39. J.P. Muczyk & B.C. Reimann. 1987. The case for directive leadership. *The Academy of Management Executive* 1:301–311.
40. W. A. Pasmore. 1988. *Designing effective organizations: The sociotechnical systems perspective*. New York: Wiley; T. J. Peters & R.H. Waterman, Jr. 1982. *In search of excellence: Lessons from America's best-run companies*. New York: Harper & Row

Note

Changes made:

- Converted Exhibits to Figures using Libretexts templates. Added accessibility information as needed.
- Hyphenated a few words.
- Changed References to a numbered list per accessibility review.

This page titled 4.6.2.2.4: Types of Leaders and Leader Emergence is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

4.6.2.2.5: The Trait Approach to Leadership

Learning Objectives

1. What are the trait perspectives on leadership?

Ancient Greek, Roman, Egyptian, and Chinese scholars were keenly interested in leaders and leadership. Their writings portray leaders as heroes. Homer, in his poem *The Odyssey*, portrays Odysseus during and after the Trojan War as a great leader who had vision and self-confidence. His son Telemachus, under the tutelage of Mentor, developed his father's courage and leadership skills.[41] Out of such stories there emerged the "great man" theory of leadership, and a starting point for the contemporary study of leadership.

The **great man theory of leadership** states that some people are born with the necessary attributes to be great leaders. Alexander the Great, Julius Caesar, Joan of Arc, Catherine the Great, Napoleon, and Mahatma Gandhi are cited as naturally great leaders, born with a set of personal qualities that made them effective leaders. Even today, the belief that truly great leaders are born is common. For example, Kenneth Labich, writer for Fortune magazine, commented that "the best leaders seem to possess a God-given spark."[42]

During the early 1900s, scholars endeavored to understand leaders and leadership. They wanted to know, from an organizational perspective, what characteristics leaders hold in common in the hope that people with these characteristics could be identified, recruited, and placed in key organizational positions. This gave rise to early research efforts and to what is referred to as the *trait approach to leadership*. Prompted by the great man theory of leadership and the emerging interest in understanding what leadership is, researchers focused on the leader—Who is a leader? What are the distinguishing characteristics of the great and effective leaders? The great man theory of leadership holds that some people are born with a set of personal qualities that make truly great leaders. Mahatma Gandhi is often cited as a naturally great leader.

Leader Trait Research

Ralph Stogdill, while on the faculty at The Ohio State University, pioneered our modern (late 20th century) study of leadership.[43] Scholars taking the trait approach attempted to identify physiological (appearance, height, and weight), demographic (age, education, and socioeconomic background), personality (dominance, self-confidence, and aggressiveness), intellective (intelligence, decisiveness, judgment, and knowledge), task-related (achievement drive, initiative, and persistence), and social characteristics (sociability and cooperativeness) with leader emergence and leader effectiveness. After reviewing several hundred studies of leader traits, Stogdill in 1974 described the successful leader this way:

The [successful] leader is characterized by a strong drive for responsibility and task completion, vigor and persistence in pursuit of goals, venturesomeness and originality in problem solving, drive to exercise initiative in social situations, self-confidence and sense of personal identity, willingness to accept consequences of decision and action, readiness to absorb interpersonal stress, willingness to tolerate frustration and delay, ability to influence other person's behavior, and capacity to structure social interaction systems to the purpose at hand.[44]

The last three decades of the 20th century witnessed continued exploration of the relationship between traits and both leader emergence and leader effectiveness. Edwin Locke from the University of Maryland and a number of his research associates, in their recent review of the trait research, observed that successful leaders possess a set of core characteristics that are different from those of other people.[45] Although these core traits do not solely determine whether a person will be a leader—or a successful leader—they are seen as preconditions that endow people with leadership potential. Among the core traits identified are:

- *Drive*—a high level of effort, including a strong desire for achievement as well as high levels of ambition, energy, tenacity, and initiative
- *Leadership motivation*—an intense desire to lead others
- *Honesty and integrity*—a commitment to the truth (nondeceit), where word and deed correspond
- *Self-confidence*—an assurance in one's self, one's ideas, and one's ability
- *Cognitive ability*—conceptually skilled, capable of exercising good judgment, having strong analytical abilities, possessing

the capacity to think strategically and multidimensionally
- *Knowledge of the business*—a high degree of understanding of the company, industry, and technical matters
- *Other traits*—charisma, creativity/originality, and flexibility/adaptiveness[46]

While leaders may be "people with the right stuff," effective leadership requires more than simply possessing the correct set of motives and traits. Knowledge, skills, ability, vision, strategy, and effective vision implementation are all necessary for the person who has the "right stuff" to realize their leadership potential.[47] According to Locke, people endowed with these traits engage in behaviors that are associated with leadership. As followers, people are attracted to and inclined to follow individuals who display, for example, honesty and integrity, self-confidence, and the motivation to lead.

Personality psychologists remind us that behavior is a result of an interaction between the person and the situation—that is, Behavior =f[(Person) (Situation)]. To this, psychologist Walter Mischel adds the important observation that personality tends to get expressed through an individual's behavior in "weak" situations and to be suppressed in "strong" situations.[48] A strong situation is one with strong behavioral norms and rules, strong incentives, clear expectations, and rewards for a particular behavior. Our characterization of the mechanistic organization with its well-defined hierarchy of authority, jobs, and standard operating procedures exemplifies a strong situation. The organic social system exemplifies a weak situation. From a leadership perspective, a person's traits play a stronger role in their leader behavior and ultimately leader effectiveness when the situation permits the expression of their disposition. Thus, personality traits prominently shape leader behavior in weak situations.

Finally, about the validity of the "great person approach to leadership": Evidence accumulated to date does not provide a strong base of support for the notion that leaders are born. Yet, the study of twins at the University of Minnesota leaves open the possibility that part of the answer might be found in our genes. Many personality traits and vocational interests (which might be related to one's interest in assuming responsibility for others and the motivation to lead) have been found to be related to our "genetic dispositions" as well as to our life experiences.[49] Each core trait recently identified by Locke and his associates traces a significant part of its existence to life experiences. Thus, a person is not born with self-confidence. Self-confidence is developed, honesty and integrity are a matter of personal choice, motivation to lead comes from within the individual and is within his control, and knowledge of the business can be acquired. While cognitive ability does in part find its origin in the genes, it still needs to be developed. Finally, drive, as a dispositional trait, may also have a genetic component, but it too can be self- and other-encouraged. It goes without saying that none of these ingredients are acquired overnight.

Other Leader Traits

Sex and gender, disposition, and self-monitoring also play an important role in leader emergence and leader style.

Sex and Gender Role

Much research has gone into understanding the role of sex and gender in leadership.[50] Two major avenues have been explored: sex and gender roles in relation to leader emergence, and whether style differences exist across the sexes.

Evidence supports the observation that men emerge as leaders more frequently than women.[51] Throughout history, few women have been in positions where they could develop or exercise leadership behaviors. In contemporary society, being perceived as experts appears to play an important role in the emergence of women as leaders. Yet, gender role is more predictive than sex. Individuals with "masculine" (for example, assertive, aggressive, competitive, willing to take a stand) as opposed to "feminine" (cheerful, affectionate, sympathetic, gentle) characteristics are more likely to emerge in leadership roles.[52] In our society males are frequently socialized to possess the masculine characteristics, while females are more frequently socialized to possess the feminine characteristics.

Recent evidence, however, suggests that individuals who are androgynous (that is, who simultaneously possess both masculine and feminine characteristics) are as likely to emerge in leadership roles as individuals with only masculine characteristics. This suggests that possessing feminine qualities does not distract from the attractiveness of the individual as a leader.[53]

With regard to leadership style, researchers have looked to see if male-female differences exist in task and interpersonal styles, and whether or not differences exist in how autocratic or democratic men and women are. The answer is, when it comes to interpersonal versus task orientation, differences between men and women appear to be marginal. Women are somewhat more concerned with meeting the group's interpersonal needs, while men are somewhat more concerned with meeting the group's task

needs. Big differences emerge in terms of democratic versus autocratic leadership styles. Men tend to be more autocratic or directive, while women are more likely to adopt a more democratic/participative leadership style.[54] In fact, it may be because men are more directive that they are seen as key to goal attainment and they are turned to more often as leaders.[55]

Dispositional Trait

Psychologists often use the terms disposition and mood to describe and differentiate people. Individuals characterized by a positive affective state exhibit a mood that is active, strong, excited, enthusiastic, peppy, and elated. A leader with this mood state exudes an air of confidence and optimism and is seen as enjoying work-related activities.

Recent work conducted at the University of California-Berkeley demonstrates that leaders (managers) with positive affectivity (a positive mood state) tend to be more competent interpersonally, to contribute more to group activities, and to be able to function more effectively in their leadership role.[56] Their enthusiasm and high energy levels appear to be infectious, transferring from leader to followers. Thus, such leaders promote group cohesiveness and productivity. This mood state is also associated with low levels of group turnover and is positively associated with followers who engage in acts of good group citizenship.[57]

Self-Monitoring

Self-monitoring as a personality trait refers to the strength of an individual's ability and willingness to read verbal and nonverbal cues and to alter one's behavior so as to manage the presentation of the self and the images that others form of the individual. "High self-monitors" are particularly astute at reading social cues and regulating their self-presentation to fit a particular situation. "Low self-monitors" are less sensitive to social cues; they may either lack motivation or lack the ability to manage how they come across to others.

Some evidence supports the position that high self-monitors emerge more often as leaders. In addition, they appear to exert more influence on group decisions and initiate more structure than low self-monitors. Perhaps high self-monitors emerge as leaders because in group interaction they are the individuals who attempt to organize the group and provide it with the structure needed to move the group toward goal attainment.[58]

Concept Check

1. What are the trait perspectives on leadership?

References

41. F. A. Kramer. 1992 (Summer). Perspectives on leadership from Homer's Odyssey. *Business and the Contemporary World* 168–173.
42. K. Labich. 1988 (Oct. 24). The seven keys to business leadership.Fortune,58.
43. Stogdill, 1948; R. M. Stogdill. 1974. *Handbook of leadership: A survey of theory and research*. New York: Free Press.
44. Ibid., 81. See also Stogdill, 1948.
45. S.A. Kirkpatrick & E.A. Locke. 1991. Leadership: Do traits matter? *The Executive* 5(2):48–60. E.A. Locke, S. Kirkpatrick, J.K. Wheeler, J. Schneider, K. Niles, H. Goldstein, K. Welsh, & D.-O. Chad. 1991. *The essence of leadership: The four keys to leading successfully*. New York: Lexington.
46. Kirkpatrick & Locke. 1991. The best managers: What it takes. 2000 (Jan. 10). *Business Week*,158.
47. Locke et al., 1991; T.A. Stewart. 1999 (Oct. 11). Have you got what it takes? *Fortune* 140 (7):318–322.
48. W. Mischel. 1973. Toward a cognitive social learning reconceptualization of personality. *Psychological Review* 80:252– 283.
49. R.J. House & R.N. Aditya. 1997. The social scientific study of leadership: Quo vadis? Journal of Management 23:409– 473; T.J. Bouchard, Jr., D.T. Lykken, M. McGue, N.L. Segal, & A. Tellegen. 1990. Sources of human psychological differences: The Minnesota study of twins reared apart. *Science* 250:223–228.
50. S. Helgesen. 1990. *The female advantage*. New York: Doubleday/Currency; J. Fierman. 1990 (Dec. 17). Do women manage differently? *Fortune* 122:115–120; J.B. Rosener. 1990 (Nov.–Dec.). Ways women lead.Harvard Business Review68(6): 119–125.
51. J.B. Chapman. 1975. Comparison of male and female leadership styles. *Academy of Management Journal* 18:645–650; E.A. Fagenson 1990. Perceived masculine and feminine attributes examined as a function of individual's sex and level in the organizational power hierarchy: A test of four theoretical perspectives. *Journal of Applied Psychology* 75:204–211.

52. R.L. Kent & S.E. Moss. 1994. Effects of sex and gender role on leader emergence. *Academy of Management Journal* 37: 1335–1346.
53. Ibid.
54. A.H. Early & B.T. Johnson. 1990. Gender and leadership style: A meta-analysis. *Psychological Bulletin* 108:233–256.
55. G.H. Dobbins, W.S. Long, E. Dedrick, & T.C. Clemons. 1990. The role of self-monitoring and gender on leader emergence: A laboratory and field study. *Journal of Management* 16:609–618.
56. B.M. Staw & S.G. Barsade. 1993. Affect and managerial performance: A test of the sadder-but-wiser vs happier-and-smarter hypothesis. *Administrative Science Quarterly* 38:304–331.
57. J.M. George & K. Bellenhausen. 1990. Understanding prosocial behavior, sales performance, and turnover: A group-level analysis in a service context. *Journal of Applied Psychology* 75:698–709.
58. Dobbins et al., 1990.

Note

Changes made:

- Corrected typo: Changed: "termsdispositionandmoodto" to "terms disposition and mood to."
- Changed References to a numbered list per accessibility review.
- Corrected typo: Changed "forFortunemagazine" to "for Fortune magazine."

This page titled 4.6.2.2.5: The Trait Approach to Leadership is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

4.6.2.2.6: Behavioral Approaches to Leadership

Learning Objectives

1. What are the behavioral perspectives on leadership?

The nearly four decades of research that focused on identifying the personal traits associated with the emergence of leaders and leader effectiveness resulted in two observations.

1. First, leader traits are important—people who are endowed with the "right stuff" (drive, self-confidence, honesty, and integrity) are more likely to emerge as leaders and to be effective leaders than individuals who do not possess these characteristics.
2. Second, traits are only a part of the story. Traits only account for part of why someone becomes a leader and why they are (or are not) effective leaders.

Still under the influence of the great man theory of leadership, researchers continued to focus on the leader in an effort to understand leadership—who emerges and what constitutes effective leadership. Researchers then began to reason that maybe the rest of the story could be understood by looking at what it is that leaders do. Thus, we now turn our attention to leader behaviors and the behavioral approaches to leadership.

It is now common to think of effective leadership in terms of what leaders do. CEOs and management consultants agree that effective leaders display trust in their employees, develop a vision, keep their cool, encourage risk, bring expertise into the work setting, invite dissent, and focus everyone's attention on that which is important.[59] William Arruda, in a *Fortune* article, noted that "organizations with strong coaching cultures report their revenue to be above average, compared to their peer group." Sixty-five percent of employees "from strong coaching cultures rated themselves as highly engaged," compared to 13 percent of employees worldwide."[60] Jonathan Anthony calls himself an intrapreneur and corporate disorganizer, because same-old, same-old comms practices are dying in front of our eyes.[61] Apple founder Steve Jobs believed that the best leaders are coaches and team cheerleaders. Similar views have been frequently echoed by management consultant Tom Peters.

During the late 1940s, two major research programs—The Ohio State University and the University of Michigan leadership studies—were launched to explore leadership from a behavioral perspective.

The Ohio State University Studies

A group of Ohio State University researchers, under the direction of Ralph Stogdill, began an extensive and systematic series of studies to identify leader behaviors associated with effective group performance. Their results identified two major sets of leader behaviors: consideration and initiating structure.

Consideration is the "relationship-oriented" behavior of a leader. It is instrumental in creating and maintaining good relationships (that is, addressing the group's maintenance needs) with organizational members. Consideration behaviors include being supportive and friendly, representing people's interests, communicating openly with group members, recognizing them, respecting their ideas, and sharing concern for their feelings.

Initiating structure involves "task-oriented" leader behaviors. It is instrumental in the efficient use of resources to attain organizational goals, thereby addressing the group's task needs. Initiating structure behaviors include scheduling work, deciding what is to be done (and how and when to do it), providing direction to organizational members, planning, coordinating, problem-solving, maintaining standards of performance, and encouraging the use of uniform procedures.

After consideration and initiating structure behaviors were first identified, many leaders believed that they had to behave one way or the other. If they initiated structure, they could not be considerate, and vice versa. It did not take long, however, to recognize that leaders can simultaneously display any combination of both behaviors.

The Ohio State studies are important because they identified two critical categories of behavior that distinguish one leader from another. Both consideration and initiating structure behavior can significantly impact work attitudes and behaviors. Unfortunately,

the effects of consideration and initiating structure are not consistent from situation to situation.[62] In some of the organizations studied, for example, high levels of initiating structure increased performance. In other organizations, the amount of initiating structure seemed to make little difference. Although most organizational members reported greater satisfaction when leaders acted considerately, consideration behavior appeared to have no clear effect on performance.

Initially, these mixed findings were disappointing to researchers and managers alike. It had been hoped that a profile of the most effective leader behaviors could be identified so that leaders could be trained in the best ways to behave. Research made clear, however, that there is no one best style of leader behavior for all situations.

The University of Michigan Studies

At about the same time that the Ohio State studies were underway, researchers at the University of Michigan also began to investigate leader behaviors. As at Ohio State, the Michigan researchers attempted to identify behavioral elements that differentiated effective from ineffective leaders.[63]

The two types of leader behavior that stand out in these studies are job centered and organizational member centered. *Job-centered behaviors* are devoted to supervisory functions, such as planning, scheduling, coordinating work activities, and providing the resources needed for task performance. *Employee-member-centered* behaviors include consideration and support for organizational members. These dimensions of behavior, of course, correspond closely to the dimensions of initiating structure and consideration identified at Ohio State. The similarity of the findings from two independent groups of researchers added to their credibility. As the Ohio State researchers had done, the Michigan researchers also found that any combination of the two behaviors was possible.

The studies at Michigan are significant because they reinforce the importance of leader behavior. They also provide the basis for later theories that identify specific, effective matches of work situations and leader behaviors. Subsequent research at Michigan and elsewhere has found additional behaviors associated with effective leadership: support, work facilitation, goal emphasis, and interaction facilitation.[64]

These four behaviors are important to the successful functioning of the group in that support and interaction facilitation contribute to the group's maintenance needs, and goal emphasis and work facilitation contribute to the group's task needs. The Michigan researchers also found that these four behaviors do not need to be brought to the group by the leader. In essence, the leader's real job is to set the tone and create the climate that ensure these critical behaviors are present.[65]

The Leadership Grid®

Much of the credit for disseminating knowledge about important leader behaviors must go to Robert R. Blake and Jane S. Mouton, who developed a method for classifying styles of leadership compatible with many of the ideas from the Ohio State and Michigan studies.[66] In their classification scheme, *concern for results* (production) emphasizes output, cost effectiveness, and (in for-profit organizations) a concern for profits. *Concern for people* involves promoting working relationships and paying attention to issues of importance to group members. As shown in Figure 4.6.2.2.6.1, the Leadership Grid® demonstrates that any combination of these two leader concerns is possible, and five styles of leadership are highlighted here.

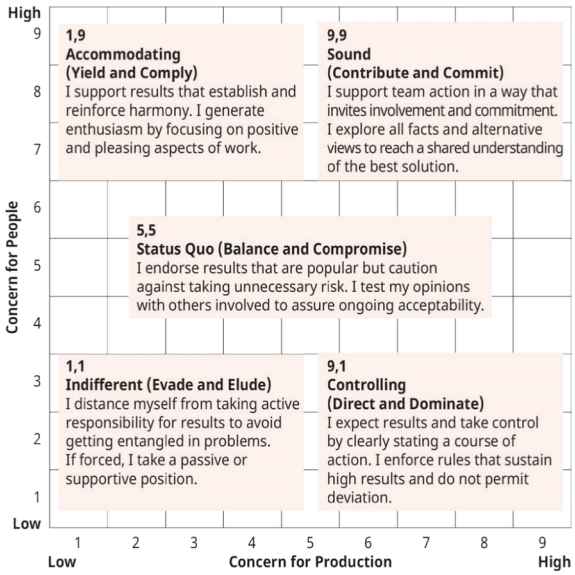


Figure 4.6.2.2.6.1: Blake and Mouton's Managerial Grid® Source: Adapted from R. McKee and B. Carlson. 1999.The Power to Change, p.16.

Accessibility Note

Figure 4.6.2.2.6.1: Blake and Mouton's Managerial Grid® depicts five styles of leadership in five light brown boxes.

The horizontal axis is labeled "Concern for Production" and ranges from 1 Low to 9 High. The vertical axis is labeled "Concern for People" and ranges from 1 Low to 9 High.

The five brown boxes are labeled with their X-Y coordinates as follows:

- Upper left corner: 1,9 Accommodating (Yield and Comply). I support results that establish and reinforce harmony. I generate enthusiasm by focusing on positive and pleasing aspects of work.
- Upper right corner: 9,9 Sound (Contribute and Commit), I support team action in a way that invites involvement and commitment. I explore all facts and alternative views to reach a share understanding of the best solution.
- Middle: 5,5 Status Quo (Balance and Compromise). I endorse results that are popular but caution against taking unnecessary risk. I test my opinions with others involved to assure ongoing acceptability.
- Lower left corner: 1,1 Indifferent (Evade and Elude:, I distance myself from taking active responsibility for results to avoid getting entangled in problems. If forced, I take a passive or supportive position.
- Lower right corner: 9.1 Controlling (Direct and Dominate). I expect results and take control by clearly stating a course of action. I enforce rules that sustain high results and do not permit deviation.

Blake and Mouton contend that the sound (contribute and commit) leader (a high concern for results and people, or 9,9) style is universally the most effective.[67] While the Leadership Grid® is appealing and well structured, research to date suggests that there is no universally effective style of leadership (9,9 or otherwise).[68] There are, however, well-identified situations in which a 9,9 style is unlikely to be effective. Organizational members of high-involvement organizations who have mastered their job duties require little production-oriented leader behavior. Likewise, there is little time for people-oriented behavior during an emergency. Finally, evidence suggests that the "high-high" style may be effective when the situation calls for high levels of initiating structure. Under these conditions, the initiation of structure is more acceptable, favorably affecting follower satisfaction and performance, when the leader is also experienced as warm, supportive, and considerate.[69]

Concept Check

1. What are the behavioral approaches to defining leadership?
2. What roles do gender and the popular perceptions of gender roles have on views of leadership traits?

References

59. K. Labich, 1988, 58–66.
60. C. Williams. 2017 (June 23). Leadership: Coaching has a Role to Play in Business. *Central Penn Business Journal*. Central Penn Business Journal article.
61. J. Anthony. 2017. *This Much We Know.* (Accessed August 4, 2017). J. Anthony article. 10-ideas-and-concepts-that-describe-me-really-well/
62. E.A. Fleishman. 1953. The description of supervisory behavior. *Personnel Psychology* 37:1–6; E.A. Fleishman & E.F. Harris. 1962. Patterns of leadership behavior related to employee grievances and turnover. *Personnel Psychology* 15:43–56; A. W. Halpin & B. J. Winer. 1957. A factorial study of the leader behavior descriptions. In R.M. Stogdill & A.C. Coons (eds.)*, Leader behavior: Its description and measurement*. Columbus: Bureau of Business Research, Ohio State University; J.K. Hemphill & A.E. Coons. 1975. Development of the leader behavior description questionnaire. In R. M. Stogdill & A. E. Coons (eds.), *Leader behavior*; S. Kerr & C. Schriesheim. 1974. Consideration, initiating structure, and organizational criteria—an update of Korman's 1966 review. *Personnel Psychology* 27:555–568.
63. D. Katz & R.L. Kahn. 1952. Some recent findings in human relations research. In E. Swanson, T. Newcomb, & E. Hartley (eds.), *Readings in social psychology*, New York: Holt, Rinehart, & Winston; D. Katz, N. Macoby, & N. Morse. 1950. *Productivity, supervision, and morale in an office situation,* Ann Arbor, MI: Institute for Social Research; F.C. Mann & J. Dent. 1954. The supervisor: Member of two organizational families. *Harvard Business Review* 32:103–112.
64. D. G. Bowers & S. C. Seashore. 1966. Pretesting organizational effectiveness with a four-factor theory of leadership. *Administrative Science Quarterly* 11:238–262; Yukl, 1971; D.A. Nadler, G.D. Jenkins, Jr., C. Cammonn, and E.E. Lawler, III. 1975. *The Michigan organizational assessment package progress report*. Ann Arbor: Institute for Social Research, University of Michigan.
65. Bowers & Seashore, 1966.
66. R.R. Blake & J.S. Mouton. 1964. *The managerial grid*. Houston: Gulf; R.R. Blake & J.S. Mouton. 1981. *The versatile manager: A grid profile*, Homewood, IL: Dow Jones-Irwin; R.R. Blake & J.S. Mouton. 1984. *The new managerial grid III*. Houston: Gulf.
67. R.R. Blake & J.S. Mouton. 1981. *Management by grid® principles or situationalism: Which?* Group and Organization Studies6:439–455.
68. L. L. Larson, J. G. Hunt, & R. N. Osborn. 1976. The great hi-hi leader behavior myth: A lesson from Occam's razor. *Academy of Management Journal* 19:628–641.
69. D. Tjosvold. 1984. Effects of warmth and directiveness on subordinate performance on a subsequent task. Journal of Applied Psychology69:422–427; A.W. Halpin. 1957. The leader behavior and effectiveness of aircraft commanders. In R.M. Stogdill & A. E. Coons (eds.). *Leader Behavior: Its description and measurement*. Columbus, OH: The Ohio State University, Bureau of Business Research.; E.A. Fleishman & J. Simmons. 1970. Relationship between leadership patterns and effectiveness ratings among Israeli foremen. *Personnel Psychology* 23:169–172.

Note

Changes made:

- Changed Exhibit to Figure using Libretexts template, adding accessibility information as needed.
- Converted some text to numbered list."
- Fixed typo from "leadersdo" to "leaders do".
- Accessibility fixes (References numbered list, fixed links).

This page titled 4.6.2.2.6: Behavioral Approaches to Leadership is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

4.6.2.2.7: Situational (Contingency) Approaches to Leadership

Learning Objectives

1. What are the situational perspectives on leadership?

As early as 1948, Ralph Stogdill stated that "the qualities, characteristics, and skills required in a leader are determined to a large extent by the demands of the situation in which he is to function as a leader."[70] In addition, it had been observed that two major leader behaviors, initiating structure and consideration, didn't always lead to equally positive outcomes. That is, there are times when initiating structure results in performance increases and follower satisfaction, and there are times when the results are just the opposite. Contradictory findings such as this lead researchers to ask "Under what conditions are the results positive in nature?" and "When and why are they negative at other times?" Obviously, situational differences and key contingencies are at work.

Several theories have been advanced to address this issue. These are:

- Fiedler's contingency theory of leadership,
- the path-goal theory of leader effectiveness,
- Hersey and Blanchard's life cycle theory,
- cognitive resource theory,
- the decision tree, and
- the decision process theory.[71]

We explore two of the better-known situational theories of leadership, Fred Fiedler's contingency model and Robert J. House's path-goal theory, here. Victor Vroom, Phillip Yetton, and Arthur Jago's decision tree model also applies.

Fiedler's Contingency Model

One of the earliest, best-known, and most controversial situation-contingent leadership theories was set forth by Fred E. Fiedler from the University of Washington.[72] This theory is known as the **contingency theory of leadership**.

According to Fiedler, organizations attempting to achieve group effectiveness through leadership must:

- assess the leader according to an underlying trait,
- assess the situation faced by the leader, and
- construct a proper match between the two.

The Leader's Trait

Leaders are asked about their **least-preferred coworker (LPC)**, the person with whom they *least* like to work. The most popular interpretation of the LPC score is that it reflects a leader's underlying disposition toward others—for example: pleasant/unpleasant, cold/warm, friendly/unfriendly, and untrustworthy/trustworthy.

Fiedler states that leaders with high LPC scores are *relationship oriented*—they need to develop and maintain close interpersonal relationships. They tend to evaluate their least-preferred coworkers in fairly favorable terms. Task accomplishment is a secondary need to this type of leader and becomes important only after the need for relationships is reasonably well satisfied. In contrast, leaders with low LPC scores tend to evaluate the individuals with whom they least like to work fairly negatively. They are *task-oriented* people, and only after tasks have been accomplished are low-LPC leaders likely to work on establishing good social and interpersonal relations.

The Situational Factor

Some situations favor leaders more than others do. To Fiedler, *situational favorableness* is the degree to which leaders have control and influence and therefore feel that they can determine the outcomes of a group interaction.[73]

Several years later, Fiedler changed his situational factor from situational favorability to situational control—where situational control essentially refers to the degree to which a leader can influence the group process.[74] Three factors work together to determine how favorable a situation is to a leader. In order of importance, they are

1. *leader-member relations*—the degree of the group's acceptance of the leader, their ability to work well together, and members' level of loyalty to the leader;
2. *task structure*—the degree to which the task specifies a detailed, unambiguous goal and how to achieve it; and
3. *position power*—a leader's direct ability to influence group members.

The situation is most favorable for a leader when the relationship between the leader and group members is good, when the task is highly structured, and when the leader's position power is strong (cell or column 1 in Figure 4.6.2.2.7.1). The least-favorable situation occurs under poor leader-member relations, an unstructured task, and weak position power (cell or column 8).

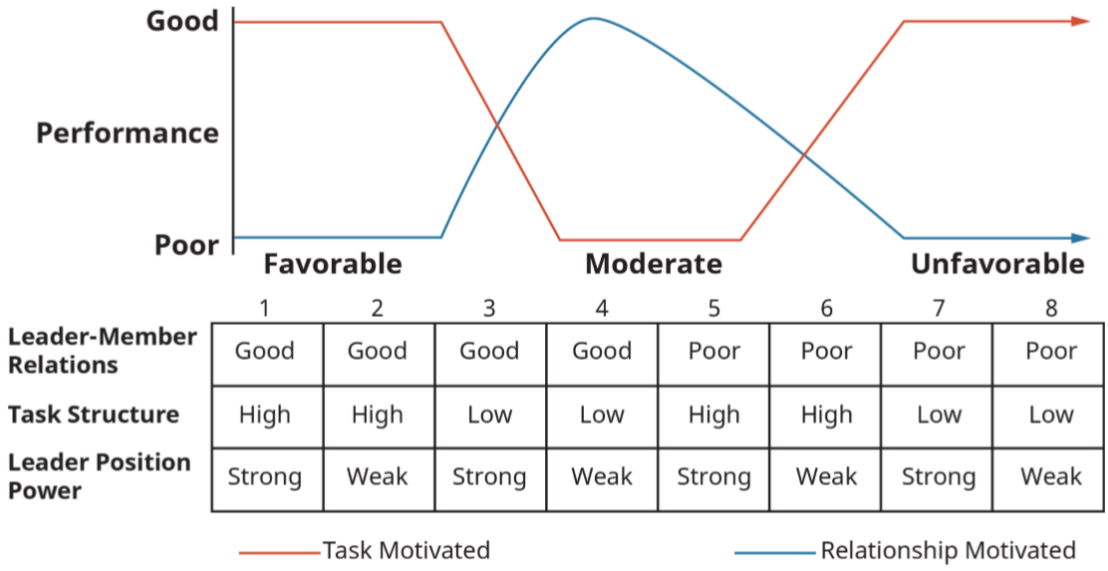


Figure 4.6.2.2.7.1: Fiedler's Contingency Model of Leader-Situation Matches (Source:Adapted from F. E. Fiedler and M. M. Chemers. 1974. Leadership and effective management. Glenview, IL: Scott, Foresman.)

Accessibility Note

Figure 4.6.2.2.7.1: Fiedler's Contingency Model of Leader-Situation Matches shows three metrics with a scale from 1 to 8. The metrics are:

- Leader-Member Relations: 1 Good, 2 Good, 3 Good, 4 Good, 5 Poor, 6 Poor, 7 Poor, 8 Poor.
- Task Structure: 1 High, 2 High, 3 Low, 4 Low, 5 High, 6 High, 7 Low, 8 Low.
- Leader Position Power: 1 Strong, 2 Weak, 3 Strong, 4 Weak, 5 Strong, 6 Weak, 7 Strong, 8 Weak.

Superimposed on it at the top is an orange line depicting "Task Motivated" and a blue line depicting "Relationship Motivated."

Leader-Situation Matches

Some combinations of leaders and situations work well; others do not. In search of the best combinations, Fiedler examined a large number of leadership situations. He argued that most leaders have a relatively unchangeable or dominant style, so organizations need to design job situations to fit the leader.[75]

While the model has not been fully tested and tests have often produced mixed or contradictory findings,[76] Fiedler's research indicates that relationship-oriented (high-LPC) leaders are much more effective under conditions of intermediate favorability than under either highly favorable or highly unfavorable situations. Fiedler attributes the success of relationship-oriented leaders in situations with intermediate favorability to the leader's nondirective, permissive attitude; a more directive attitude could lead to anxiety in followers, conflict in the group, and a lack of cooperation.

For highly favorable and unfavorable situations, task-oriented leaders (those with a low LPC) are very effective. As tasks are accomplished, a task-oriented leader allows the group to perform its highly structured tasks without imposing more task-directed behavior. The job gets done without the need for the leader's direction. Under unfavorable conditions, task-oriented behaviors, such as setting goals, detailing work methods, and guiding and controlling work behaviors, move the group toward task accomplishment.

As might be expected, leaders with mid-range LPC scores can be more effective in a wider range of situations than high- or low-LPC leaders.[77] Under conditions of low favorability, for example, a middle-LPC leader can be task oriented to achieve performance, but show consideration for and allow organizational members to proceed on their own under conditions of high situational favorability.

Controversy over the Theory

Although Fiedler's theory often identifies appropriate leader-situation matches and has received broad support, it is not without critics. Some note that it characterizes leaders through reference to their attitudes or personality traits (LPC) while it explains the leader's effectiveness through their behaviors—those with a particular trait will behave in a particular fashion. The theory fails to make the connection between the least-preferred coworker attitude and subsequent behaviors. In addition, some tests of the model have produced mixed or contradictory findings.[78] Finally, what is the true meaning of the LPC score—exactly what is being revealed by a person who sees their least-preferred coworker in positive or negative terms? Robert J. House and Ram N. Aditya recently noted that, in spite of the criticisms, there has been substantial support for Fiedler's theory.[79]

Path-Goal Theory

Robert J. House and Martin Evans, while on the faculty at the University of Toronto, developed a useful leadership theory. Like Fiedler's, it asserts that the type of leadership needed to enhance organizational effectiveness depends on the situation in which the leader is placed. Unlike Fiedler, however, House and Evans focus on the leader's observable behavior. Thus, managers can either match the situation to the leader or modify the leader's behavior to fit the situation.

The model of leadership advanced by House and Evans is called the **path-goal theory of leadership** because it suggests that an effective leader provides organizational members with a *path* to a valued *goal*. According to House, the motivational function of the leader consists of increasing personal payoffs to organizational members for work-goal attainment, and making the path to these payoffs easier to travel by clarifying it, reducing roadblocks and pitfalls, and increasing the opportunities for personal satisfaction en route.[80]

Effective leaders therefore provide rewards that are valued by organizational members. These rewards may be pay, recognition, promotions, or any other item that gives members an incentive to work hard to achieve goals. Effective leaders also give clear instructions so that ambiguities about work are reduced and followers understand how to do their jobs effectively. They provide coaching, guidance, and training so that followers can perform the task expected of them. They also remove barriers to task accomplishment, correcting shortages of materials, inoperative machinery, or interfering policies.

An Appropriate Match

According to the path-goal theory, the challenge facing leaders is basically twofold.

- First, they must analyze situations and identify the most appropriate leadership style. For example, experienced employees who work on a highly structured assembly line don't need a leader to spend much time telling them how to do their jobs—they already know this. The leader of an archeological expedition, though, may need to spend a great deal of time telling inexperienced laborers how to excavate and care for the relics they uncover.
- Second, leaders must be flexible enough to use different leadership styles as appropriate. To be effective, leaders must

engage in a wide variety of behaviors. Without an extensive repertoire of behaviors at their disposal, a leader's effectiveness is limited.[81] All team members will not, for example, have the same need for autonomy. The leadership style that motivates organizational members with strong needs for autonomy (participative leadership) is different from that which motivates and satisfies members with weaker autonomy needs (directive leadership). The degree to which leadership behavior matches situational factors will determine members' motivation, satisfaction, and performance (see Figure 4.6.2.2.7.2).[82]

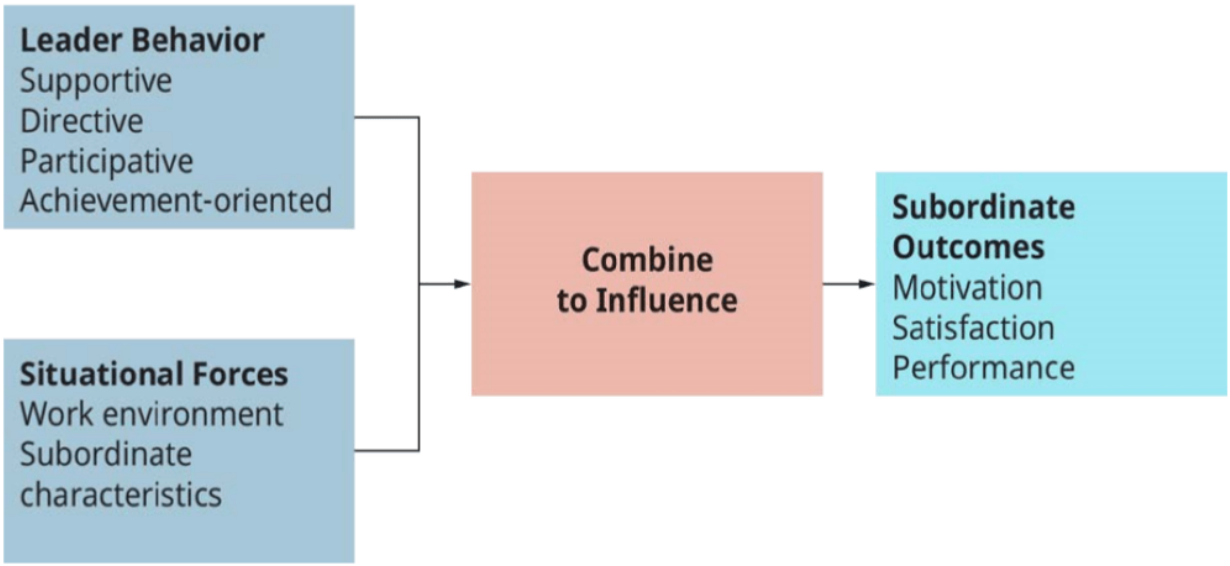


Figure 4.6.2.2.7.2: The Path-Goal Leadership Model (Attribution: Copyright Rice University, OpenStax, under CC-BY 4.0 license)

Accessibility Note

Figure 4.6.2.2.7.2: The Path-Goal Leadership Model shows two blue boxes feeding into an orange box (labeled "Combine to Influence" which has an arrow pointing to a light blue box (labeled "**Subordinate Outcomes**, Motivation, Satisfaction, Performance.)

The two blue boxes starting from the top are labeled:

- **Leader Behavior:** Supportive, Directive, Participative, Achievement-oriented.
- **Situational Forces:** Work environment, Subordinate characteristics.

Behavior Dimensions

According to path-goal theory, there are four important dimensions of leader behavior, each of which is suited to a particular set of situational demands.[83]

- *Supportive leadership*—At times, effective leaders demonstrate concern for the well-being and personal needs of organizational members. Supportive leaders are friendly, approachable, and considerate to individuals in the workplace. Supportive leadership is especially effective when an organizational member is performing a boring, stressful, frustrating, tedious, or unpleasant task. If a task is difficult and a group member has low self-esteem, supportive leadership can reduce some of the person's anxiety, increase his confidence, and increase satisfaction and determination as well.
- *Directive leadership*—At times, effective leaders set goals and performance expectations, let organizational members know what is expected, provide guidance, establish rules and procedures to guide work, and schedule and coordinate the activities of members. Directive leadership is called for when role ambiguity is high. Removing uncertainty and providing needed guidance can increase members' effort, job satisfaction, and job performance.
- *Participative leadership*—At times, effective leaders consult with group members about job-related activities and consider their opinions and suggestions when making decisions. Participative leadership is effective when tasks are unstructured. Participative leadership is used to great effect when leaders need help in identifying work procedures and where followers have the expertise to provide this help.

- *Achievement-oriented leadership*—At times, effective leaders set challenging goals, seek improvement in performance, emphasize excellence, and demonstrate confidence in organizational members' ability to attain high standards. Achievement-oriented leaders thus capitalize on members' needs for achievement and use goal-setting theory to great advantage.

Cross-Cultural Context

Gabriel Bristol, the CEO of Intelifluence Live, a full-service customer contact center offering affordable inbound customer service, outbound sales, lead generation and consulting services for small to mid-sized businesses, notes "diversity breeds innovation, which helps businesses achieve goals and tackle new challenges."[84] *Multiculturalism* is a new reality as today's society and workforce become increasingly diverse. This naturally leads to the question "Is there a need for a new and different style of leadership?"

The vast majority of the contemporary scholarship directed toward understanding leaders and the leadership process has been conducted in North America and Western Europe. Westerners have "developed a highly romanticized, heroic view of leadership."[85] Leaders occupy center stage in organizational life. We use leaders in our attempts to make sense of the performance of our groups, clubs, organizations, and nations. We see them as key to organizational success and profitability, we credit them with organizational competitiveness, and we blame them for organizational failures. At the national level, recall that President Reagan brought down Communism and the Berlin Wall, President Bush won the Gulf War, and President Clinton brought unprecedented economic prosperity to the United States during the 1990s.

This larger-than-life role ascribed to leaders and the Western romance with successful leaders raise the question "How representative is our understanding of leaders and leadership across cultures?" That is, do the results that we have examined in this chapter generalize to other cultures?

Geert Hofstede points out that significant value differences (individualism-collectivism, power distance, uncertainty avoidance, masculinity-femininity, and time orientation) cut across societies. Thus, leaders of culturally diverse groups will encounter belief and value differences among their followers, as well as in their own leader-member exchanges.

There appears to be consensus that a universal approach to leadership and leader effectiveness does not exist. Cultural differences work to enhance and diminish the impact of leadership styles on group effectiveness. For example, when leaders empower their followers, the effect for job satisfaction in India has been found to be negative, while in the United States, Poland, and Mexico, the effect is positive.[86] The existing evidence suggests similarities as well as differences in such areas as the effects of leadership styles, the acceptability of influence attempts, and the closeness and formality of relationships. The distinction between task and relationship-oriented leader behavior, however, does appear to be meaningful across cultures.[87] Leaders whose behaviors reflect support, kindness, and concern for their followers are valued and effective in Western and Asian cultures. Yet it is also clear that democratic, participative, directive, and contingent-based rewards and punishment do not produce the same results across cultures. The United States is very different from Brazil, Korea, New Zealand, and Nigeria. The effective practice of leadership necessitates a careful look at, and understanding of, the individual differences brought to the leader-follower relationship by cross-cultural contexts.[88]

Concept Check

1. Identify and describe the variables presented in Fiedler's theory of leadership.
2. What are the leadership behaviors in the path-goal theory of leadership?
3. What role does culture have in how leadership is viewed?
4. What are the differences between the trait, behavioral, and situational approaches to defining leadership?

References

70. Stogdill, 1948, 63.
71. House & Aditya, 1997.
72. F.E. Fiedler & M.M. Chemers. 1974. *Leadership and effective management.* Glenview, IL: Scott, Foresman.
73. F.E. Fiedler. 1976. The leadership game: Matching the men to the situation. *Organizational Dynamics,* 4, 9.

74. Personal conversation between Robert J. House and Fred Fiedler in September 1996, as reported in House & Aditya, 1997.
75. F.E. Fiedler. Sept.–Oct. 1965. Engineering the job to fit the manager. *Harvard Business Review,* 115–122.
76. See, for example, the supporting results of M.M. Chemers & G.J. Skrzypek. 1972. Experimental test of the contingency model of leadership effectiveness. *Journal of Personality and Social Psychology* 24:172–177; and the contradictory results of R.P. Vecchio. 1977. An empirical examination of the validity of Fiedler's model of leadership effectiveness. *Organizational Behavior and Human Performance* 19:180–206.
77. R.B. Dunham. 1984. [Interview with Fred E. Fiedler.] *Organizational behavior: People and processes in management*. Homewood, IL: Irwin, 368; J. L. Kennedy, Jr. 1982. Middle LPC leaders and the contingency model of leadership effectiveness. *Organizational Behavior and Human Performance* 30:1–14.
78. Chemens & Skrzpek, 1972; Vecchio, 1977.
79. House & Aditya. 1997; L.H. Peters, D.D. Hartke, & J.T. Pohlman. 1985. Fiedler's contingency model of leadership: An application of the meta-analysis procedure of Schmidt and Hunter. *Psychological Bulletin* 97:274–285.
80. R.J. House. 1971. A path goal theory of leader effectiveness. *Administrative Science Quarterly* 16:324.
81. R. Hoojiberg. 1996. A multidimensional approach toward leadership: An extension of the concept of behavioral complexity. *Human Relations* 49(7):917–946.
82. R.J. House & T.R. Mitchell. 1974 (Autumn). Path-goal theory of leadership, *Journal of Contemporary Business*, 86; R.J. House & G. Dessler. 1974. The path-goal theory of leadership: Some post hoc and a priori tests. In J. Hunt & L. Larson (eds.). *Contingency approaches to leadership*. Carbondale, IL: Southern Illinois University Press.
83. House & Mitchell, 1974; House & Dessler, 1974; R.T. Keller. 1989. A test of the path-goal theory of leadership with need for clarity as a moderator in research and development organizations. *Journal of Applied Psychology* 74:208–212.
84. G. Bristol. 2016. Why Diversity in the Workplace is Imperrative. *Entrepreneur*, March 25. (Accessed august 4, 2017). Entrepreneur article.
85. J.R. Meindl, S.B. Ehrlich, & J.M. Dukerich. 1985. The romance of leadership. *Administrative Science Quarterly* 30:78–102.
86. C. Robert, T. M. Probst, J. J. Martocchion, F. Drasgow, & J. J. Lawler. 2000. Empowerment and continuous improvement in the United States, Mexico, Poland, and India: Predicting fit on the basis of the dimensions of power distance and individualism. *Journal of Applied Psychology* 85:643–658.
87. P.W. Dorfman & S. Roonen. 1991. *The universality of leadership theories: Challenges and paradoxes*. Paper presented at the Academy of Management Meetings, Miami.
88. P.W. Dorfman, J.P. Howell, S. Hiblino, J.K. Lee, U. Tate, & A. Bautista. 1997. Leadership in Western and Asian countries: Commonalities and differences in effective leadership processes across cultures. *Leadership Quarterly* 8(3):233–274.

Note

Changes made:

- Converted Exhibits to Figures using Libretexts template and added accessibility commentary as needed.
- Changed "leadermember" to "leader-member" for consistency with the rest of the text.
- Converted some text to bulleted lists or numbered lists for readability (and accessibility).
- Deleted "(You can examine your own LPC score by completing the LPC self-assessment on the following page.)"
- Under "Situational Factors: changed text to also reference column (the original text referred only to cell).
- Changed References to a numbered list.

This page titled 4.6.2.2.7: Situational (Contingency) Approaches to Leadership is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 16: Mathcad Curve Fitting by Troy Siemers is licensed CC BY-NC 3.0. Original source: https://drive.google.com/file/d/0B2Pdg3LqVPNPLUhLbks0QzF5ZVhmWGp3YW5hYlVBd3RDN2x3/view?resourcekey=0-fbyowylmkO1jBbS2ehDP0g&usp=sharing.

4.6.2.2.8: Substitutes for and Neutralizers of Leadership

Learning Objectives

1. What does the concept “substitute for leadership” mean?

Several factors have been discovered that can substitute for or neutralize the effects of leader behavior.[89] *Substitutes* for leadership behavior can clarify role expectations, motivate organizational members, or satisfy members (making it unnecessary for the leader to attempt to do so). In some cases, these substitutes supplement the behavior of a leader. Sometimes it is a group member’s characteristics that make leadership less necessary, as when a master craftsperson or highly skilled worker performs up to his or her own high standards without needing outside prompting. Sometimes the task’s characteristics take over, as when the work itself—solving an interesting problem or working on a familiar job—is intrinsically satisfying. Sometimes the characteristics of the organization make leadership less necessary, as when work rules are so clear and specific that workers know exactly what they must do without help from the leader.

Substitutes for and Neutralizers of Leader Behavior. Source: Adapted from Leadership in organizations by G. A. Yukl.

Supportive or Neutralizer	Substitute Leadership	Instrumental Leadership
A. Subordinate Characteristics:		
1. Experience, ability, training		Substitute
2. "Professional" orientation	Substitute	Substitute
3. Indifference toward rewards offered by organization	Neutralizer	Neutralizer
B. Task Characteristics:		
1. Structured, routine, unambiguous task		Substitute
2. Feedback provided by task		Substitute
3. Intrinsically satisfying task	Substitute	
C. Organization Characteristics:		
1. Cohesive work group	Substitute	Substitute
2. Low position power (leader lacks control over organizational rewards)	Neutralizer	Neutralizer
3. Formalization (explicit plans, goals, areas of responsibility)		Substitute
4. Inflexibility (rigid, unyielding rules and procedures)		Neutralizer
5. Leader located apart from subordinates with only limited communication possible	Neutralizer	Neutralizer

Neutralizers of leadership, on the other hand, are not helpful; they prevent leaders from acting as they wish. A computer-paced assembly line, for example, prevents a leader from using initiating structure behavior to pace the line. A union contract that specifies that workers be paid according to seniority prevents a leader from dispensing merit-based pay. Sometimes, of course, neutralizers can be beneficial. Union contracts, for example, clarify disciplinary proceedings and identify the responsibilities of both management and labor. Leaders must be aware of the presence of neutralizers and their effects so that they can eliminate troublesome neutralizers or take advantage of any potential benefits that accompany them (such as the clarity of responsibilities provided by a union contract). If a leader’s effectiveness is being neutralized by a poor communication system, for example, the leader might try to remove the neutralizer by developing (or convincing the organization to develop) a more effective system.

Followers differ considerably in their *focus of attention* while at work, thereby affecting the effectiveness of the act of leadership. Focus of attention is an employee’s cognitive orientation while at work. It reflects what and how strongly an individual thinks about various objects, events, or phenomena while physically present at work. Focus of attention reflects an individual difference in that

not all individuals have the same cognitive orientation while at work—some think a great deal about their job, their coworkers, their leader, or off-the-job factors, while others daydream.[90] An employee's focus of attention has both "trait" and "state" qualities. For example, there is a significant amount of minute-by-minute variation in an employee's focus of attention (the "state" component), and there is reasonable consistency in the categories of events that employees think about while they are at work (the "trait" component).

Research suggests that the more followers focus on off-job (nonleader) factors, the less they will react to the leader's behaviors. Thus, a strong focus on one's life "away from work" (for example, time with family and friends) tends to neutralize the motivational, attitudinal, and/or behavioral effects associated with any particular leader behavior. It has also been observed, however, that a strong focus on the leader, either positive or negative, enhances the impact that the leader's behaviors have on followers.[91]

Managerial Leadership

You Are Now the Leader

Leading and managing are two very different things. Being a manager means something more than gaining authority or charge over former colleagues. With the title does come the power to affect company outcomes, but it also comes with something more: the power to shape the careers and personal growth of subordinates.

According to Steve Keating, a senior manager at the Toro Company, it is important not to assume that being made a manager automatically makes you a leader. Rather, being a manager means having the *opportunity* to lead. Enterprises need managers to guide processes, but the employees—the people—need a leader. Keating believes that leaders need a mindset that emphasizes people, and the leader's job is to help the people in the organization to be successful. According to Keating, "If you don't care for people, you can't lead them" (Hakim 2017 n.p.).

For someone who has been promoted over his peers, ground rules are essential. "Promotion doesn't mean the end of friendship but it does change it," according to Keating. If a peer has been promoted, rather than grouse and give in to envy, it is important to step back and look at the new manager; take a hard look at why the peer was promoted and what skill or characteristic made you a less appealing fit for the position (Hakim 2017).

Carol Walker, president of Prepared to Lead, a management consulting firm, advises new managers to develop a job philosophy. She urges new managers to develop a core philosophy that provides a guide to the day-to-day job of leading. She urges managers to build up the people they are leading and work as a "servant leader." The manager's perspective should be on employee growth and success. Leaders must bear in mind that employees don't work for the manager; they work for the organization—and for themselves. Managers coordinate this relationship; they are not the center of it. Work should not be assigned haphazardly, but with the employee's skills and growth in mind. "An employee who understands why she has been asked to do something is far more likely to assume true ownership for the assignment," Walker says (Yakowicz 2015 n.p.). A leader's agenda should be on employee success, not personal glory. Employees are more receptive when they recognize that their leader is working not for their own success, but for the employee's success.

A survey from HighGround revealed one important item that most new managers and even many seasoned managers overlook: asking for feedback. Everyone has room for growth, even managers. Traditional management dictates a top-down style in which managers review subordinates. But many companies have found it beneficial to turn things around and ask employees, "How can I be a better manager?" Of course, this upward review only works if employees believe that their opinion will be heard. Managers need to carefully cultivate a rapport where employees don't fear reprisals for negative feedback. Listening to criticism from those you are leading builds trust and helps ensure that as a manager, you are providing the sort of leadership that employees need to be successful (Kauflin 2017). Showing respect and caring for employees by asking this simple question is *inspiring*—an important aspect of leadership itself. Whether asking for feedback or focusing on an employee's fit with a particular job description, a leader helps guide employees through the day-to-day, builds a positive culture, and helps employees improve their skills.

Sources

1. Hakim, Amy C. 2017. "When a Manager Becomes a Leader." Psychology Today. Psychology Today article.
2. Yakowicz, Will. 2015. "How to Help a New Manager Become a Great Leader." *Inc*. Inc Yakowicz article.
3. Kauflin, Jeff. 2017. "Every Manager Can Become A Better Leader By Asking This One Question." *Forbes*. Forbes Kauflin article.

Questions

1. What do you think are the most important qualities in a leader? In a manager? Are your two lists mutually exclusive? Why?
2. How do you think a leader can use feedback to model the growth process for employees?

Concept Check

1. Identify and describe substitutes of leadership.

References

89. P.M. Podsakoff, B.P. Niehoff, S.B. MacKenzie, & M.L. Williams. 1993. Do substitutes for leadership really substitute for leadership: An empirical examination of Kerr and Jermier's situational leadership model. *Organizational Behavior and Human Decision Processes* 54:1–44; S. Kerr. 1977. Substitutes for leadership: Some implications for organizational design. *Organization and Administrative Sciences* 8:135–146; S. Kerr & J.M. Jermier. 1978. Substitutes for leadership: Their meaning and measurement. *Organizational Behavior and Human Performance* 22:375– 403; J. P. Howell & P. W. Dorfman. 1981. Substitutes for leadership: Test of a construct. *Academy of Management Journal* 24:714– 728; J.L. Pierce, R.B. Dunham, & L.L. Cummings. 1984. Sources of environmental structuring and participant responses. *Organizational Behavior and Human Performance* 33:214–242.
90. D.G. Gardner, R.B. Dunham, L.L. Cummings, & J.L. Pierce. 1989. Focus of attention at work: Construct definition and empirical validation. *Journal of Occupational Psychology* 62:61–77.
91. D.G. Gardner, R.B. Dunham, L.L. Cummings, & J.L. Pierce. 1987. Focus of attention at work and leader-follower relationships. *Journal of Occupational Behaviour* 8:277–294.

Note

Changes made to the original OER source material:

- Removed label from table.
- Inserted spaces between words as needed.
- Deleted blank lines as needed.
- Changed References to numbered list per accessibility review.

This page titled 4.6.2.2.8: Substitutes for and Neutralizers of Leadership is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 17.1: Calculus by Troy Siemers is licensed CC BY-NC 3.0. Original source: https://drive.google.com/file/d/0B2Pdg3LqVPNPLUhLbks0QzF5ZVhmWGp3YW5hYlVBd3RDN2x3/view?resourcekey=0-fbyowylmkO1jBbS2ehDP0g&usp=sharing.

4.6.2.2.9: Transformational, Visionary, and Charismatic Leadership

Learning Objectives

1. What are the characteristics of transactional, transformational, and charismatic leadership?

Many organizations struggling with the need to manage chaos, to undergo a culture change, to empower organizational members, and to restructure have looked for answers in "hiring the right leader." Many have come to believe that the transformational, visionary, and charismatic leader represents the style of leadership needed to move organizations through chaos.

The Transformational and Visionary Leader

Leaders who subscribe to the notion that "if it ain't broke, don't fix it" are often described as **transactional leaders**. They are extremely task oriented and instrumental in their approach, frequently looking for incentives that will induce their followers into a desired course of action.[92] These reciprocal exchanges take place in the context of a mutually interdependent relationship between the leader and the follower, frequently resulting in interpersonal bonding.[93] The transactional leader moves a group toward task accomplishment by initiating structure and by offering an incentive in exchange for desired behaviors.

The **transformational leader**, on the other hand, moves and changes (fixes) things "in a big way"! Unlike transactional leaders, they don't cause change by offering inducements. Instead, they inspire others to action through their personal values, vision, passion, and belief in and commitment to the mission.[94] Through charisma (idealized influence), individualized consideration (a focus on the development of the follower), intellectual stimulation (questioning assumptions and challenging the status quo), and/or inspirational motivation (articulating an appealing vision), transformational leaders move others to follow.

The transformational leader is also referred to as a visionary leader. **Visionary leaders** are those who influence others through an emotional and/or intellectual attraction to the leader's dreams of what "can be." Vision links a present and future state, energizes and generates commitment, provides meaning for action, and serves as a standard against which to assess performance.[95] Evidence indicates that vision is positively related to follower attitudes and performance.[96] As pointed out by Warren Bennis, a vision is effective only to the extent that the leader can communicate it in such a way that others come to internalize it as their own.[97]

As people, transformational leaders are engaging. They are characterized by extroversion, agreeableness, and openness to experience.[98] They energize others. They increase followers' awareness of the importance of the designated outcome.[99] They motivate individuals to transcend their own self-interest for the benefit of the team and inspire organizational members to self-manage (become self-leaders).[100] Transformational leaders move people to focus on higher-order needs (self-esteem and self-actualization). When organizations face a turbulent environment, intense competition, products that may die early, and the need to move fast, managers cannot rely solely on organizational structure to guide organizational activity. In these situations, transformational leadership can motivate followers to be fully engaged and inspired, to internalize the goals and values of the organization, and to move forward with dogged determination!

Transformational leadership is positively related to follower satisfaction, performance, and acts of citizenship. These effects result from the fact that transformational leader behaviors elicit trust and perceptions of procedural justice, which in turn favorably impact follower satisfaction and performance.[101] As R. Pillai, C. Schriesheim, and E. Williams note, "when followers perceive that they can influence the outcomes of decisions that are important to them and that they are participants in an equitable relationship with their leader, their perceptions of procedural justice [and trust] are likely to be enhanced."[102] Trust and experiences of organizational justice promote leader effectiveness, follower satisfaction, motivation, performance, and citizenship behaviors.

Charismatic Leadership

Ronald Reagan, Jesse Jackson, and Queen Elizabeth I have something in common with Martin Luther King Jr., Indira Gandhi, and Winston Churchill. The effectiveness of these leaders originates in part in their charisma, a special magnetic charm and appeal that arouses loyalty and enthusiasm. Each exerted considerable personal influence to bring about major events.

It is difficult to differentiate the charismatic and the transformational leader. True transformational leaders may achieve their results through the magnetism of their personality. In this case, the two types of leaders are essentially one and the same, yet it is important to note that not all transformational leaders have a personal "aura."

Sociologist Max Weber evidenced an interest in charismatic leadership in the 1920s, calling **charismatic leaders** people who possess legitimate power that arises from "exceptional sanctity, heroism, or exemplary character."[103] Charismatic leaders "single-handedly" effect changes even in very large organizations. Their personality is a powerful force, and the relationship that they forge with their followers is extremely strong.

The charismatic leadership phenomenon involves a complex interplay between the attributes of the leader and followers' needs, values, beliefs, and perceptions.[104] At its extreme, leader-follower relationships are characterized by followers' unquestioning acceptance; trust in the leader's beliefs; affection; willing obedience to, emulation of, and identification with the leader; emotional involvement with his mission; and feelings of self-efficacy directed toward the leader's mission.[105] This can work to better the welfare of individuals, such as when Lee Iacocca saved thousands of jobs through his dramatic turnaround of a failing corporate giant, the Chrysler Corporation. It also can be disastrous, as when David Koresh led dozens and dozens of men, women, and children to their fiery death in Waco, Texas. Individuals working for charismatic leaders often have higher task performance, greater task satisfaction, and lower levels of role conflict than those working for leaders with considerate or structuring behaviors.[106] What are the characteristics of these people who can exert such a strong influence over their followers? Charismatic leaders have a strong need for power and the tendency to rely heavily on referent power as their primary power base.[107] Charismatic leaders also are extremely self-confident and convinced of the rightness of their own beliefs and ideals. This self-confidence and strength of conviction make people trust the charismatic leader's judgment, unconditionally following the leader's mission and directives for action.[108] The result is a strong bond between leader and followers, a bond built primarily around the leader's personality.

Although there have been many effective charismatic leaders, those who succeed the most have coupled their charismatic capabilities with behaviors consistent with the same leadership principles followed by other effective leaders. Those who do not add these other dimensions still attract followers but do not meet organizational goals as effectively as they could. They are (at least for a time) the pied pipers of the business world, with lots of followers but no constructive direction.

Concept Check

1. What are the defining characteristics of transformational and charismatic leaders?

References

92. G.A. Yukl. 1981. *Leadership in organizations*. Englewood Cliffs, NJ: Prentice-Hall.
93. B. Kellerman. 1984. *Leadership: Multidisciplinary perspectives*. Englewood Cliffs, NJ: Prentice-Hall; F. L. Landy. 1985. *Psychology of work behavior*. Homewood, IL: Dorsey Press.
94. J.M. Burns. 1978.Leadership.New York: Harper & Row; B. M. Bass. 1985. *Leadership and performance beyond expectations*. New York: Free Press.
95. R.L. Daft. 2018.The Leadership Experience7th edition.Mason, OH: Cengage Learning.
96. J. R. Baum, E. A. Locke, & S. A. Kirkpatrick. 1998. A longitudinal study of the relation of vision and vision communication to venture growth in entrepreneurial firms. *Journal of Applied Psychology* 83:43–54; J.M. Howell & P.J. Frost. 1989. A laboratory study of charismatic leadership. *Organizational Behavior and Human Decision Processes* 43:243–269.
97. Bennis, 1989.
98. T. A. Judge & J.E. Bono. 2000. Five-factor model of personality and transformational leadership. *Journal of Applied Psychology* 85:751–765.
99. R. Pillai, C.A. Schriesheim, & E.S. Williams. 1999. Fairness perceptions and trust as mediators for transformational and

transactional leadership: A two-sample study. *Journal of Management* 25:897–933.
90. C.C. Manz & H.P. Sims, Jr. 1987. Leading workers to lead themselves: The external leadership of selfmanaged work teams. *Administrative Science Quarterly* 32:106–129.
91. Pillai, Schriesheim, & Williams, 1999.
92. Ibid., 901.
93. S.N. Eisenstadt. 1968. *Max Weber: On charisma and institution building*. Chicago: University of Chicago Press, 46.
94. J.A. Conger & R.N. Kanungo. 1987. Toward a behavioral theory of charismatic leadership in organizational settings. *Academy of Management Review* 12:637–647; Howell & Frost, 1989.
95. R.J. House & M.L. Baetz. 1979. Leadership: Some empirical generalizations and new research directions. *Research in Organizational Behavior* 1:341–423; Conger and Kanungo, 1987.
96. Howell & Frost, 1989.
97. R. J. House. 1977. A 1976 theory of charismatic leadership. In J. G. Hunt & L. L. Larson (eds.). *Leadership: The cutting edge*. Carbondale, IL: Southern Illinois University Press.
98. A. R. Willner. 1984. *The spellbinders: Charismatic political leadership*. New Haven, CT: Yale University Press.

Note

Changes made:

- Modified "exhibit" to "Figure" using Libretexts Template and added any needed accessibility commentary. Decided to remove Uber material.
- Removed italics from "transactional leaders" and put it in bold for consistency with the rest of the text.
- Added hyphen to self-confident.
- Changed References to numbered list.
- Removed blank line.

This page titled 4.6.2.2.9: Transformational, Visionary, and Charismatic Leadership is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

- 17.2: Symbolics by Troy Siemers is licensed CC BY-NC 3.0. Original source: https://drive.google.com/file/d/0B2Pdg3LqVPNPLUhLbks0QzF5ZVhmWGp3YW5hYlVBd3RDN2x3/view?resourcekey=0-fbyowylmkO1jBbS2ehDP0g&usp=sharing.

4.6.2.2.10: Leadership Needs in the 21st Century

Learning Objectives

1. How do different approaches and styles of leadership impact what is needed now?

Frequent headlines in popular business magazines like *Fortune* and *Business Week* call our attention to a major movement going on in the world of business. Organizations are being re-engineered and restructured, and network, virtual, and modular corporations are emerging. People talk about the transnational organization, the boundary-less company, the post-hierarchical organization. By the end of the decade, the organizations that we will be living in, working with, and competing against are likely to be vastly different from what we know today.

The transition will not be easy; uncertainty tends to breed resistance. We are driven by linear and rational thinking, which leads us to believe that "we can get there from here" by making some incremental changes in who we are and what we are currently doing. Existing paradigms frame our perceptions and guide our thinking. Throwing away paradigms that have served us well in the past does not come easily.

A look back tells most observers that the past decade has been characterized by rapid change, intense competition, an explosion of new technologies, chaos, turbulence, and high levels of uncertainty. A quick scan of today's business landscape suggests that this trend is not going away anytime soon. According to Professor Jay A. Conger from Canada's McGill University, "In times of great transition, leadership becomes critically important. Leaders, in essence, offer us a pathway of confidence and direction as we move through seeming chaos. The magnitude of today's changes will demand not only *more* leadership, but *newer forms* of leadership."[109]

According to Conger, two major forces are defining for us the genius of the next generation of leaders. The first force is the organization's external environment. Global competitiveness is creating some unique leadership demands. The second force is the growing diversity in organizations' internal environments. Diversity will significantly change the relationship between organizational members, work, and the organization in challenging, difficult, and also very positive ways.

What will the leaders of tomorrow be like? Professor Conger suggests that the effective leaders of the 21st century will have to be many things.[110] They will have to be *strategic opportunists*; only organizational visionaries will find strategic opportunities before competitors. They will have to be *globally aware*; with 80 percent of today's organizations facing significant foreign competition, knowledge of foreign markets, global economics, and geopolitics is crucial. They will have to be *capable of managing a highly decentralized organization*; movement toward the high-involvement organization will accelerate as the environmental demands for organizational speed, flexibility, learning, and leanness increase. They will have to be *sensitive to diversity*; during the first few years of the 21st century, fewer than 10 percent of those entering the workforce in North America will be white, Anglo-Saxon males, and the incoming women, minorities, and immigrants will bring with them a very different set of needs and concerns. They will have to be *interpersonally competent*; a highly diverse workforce will necessitate a leader who is extremely aware of and sensitive to multicultural expectations and needs. They will have to be *builders of an organizational community*; work and organizations will serve as a major source of need fulfillment, and in the process leaders will be called on to help build this community in such a way that organizational members develop a sense of ownership for the organization and its mission.

Finally, it is important to note that leadership theory construction and empirical inquiry are an ongoing endeavor. While the study of traits, behavior, and contingency models of leadership provide us with a great deal of insight into leadership, the mosaic is far from complete. During the past 15 years, several new theories of leadership have emerged; among them are leader-member exchange theory, implicit leadership theory, neo-charismatic theory, value-based theory of leadership, and visionary leadership,[111] each of which over time will add to our bank of knowledge about leaders and the leadership process.

Leaders of the 21st-century organization have a monumental challenge awaiting them and a wealth of self-enriching and fulfilling opportunities. The challenge and rewards awaiting effective leaders are awesome!

Concept Check

1. What is the role of leadership in the 21st century?

References

9. Conger, 1993.
10. Ibid.
11. House & Aditya, 1997.

Note

Changes to original OER:

- Converted References to a numbered list for accessibility.
- Fixed a few typos.

This page titled 4.6.2.2.10: Leadership Needs in the 21st Century is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

4.6.2.2.11: Summary

Note: Key terms used in this section

Charisma: A special personal magnetic charm or appeal that arouses loyalty and enthusiasm in a leader-follower relationship.

Charismatic leader: A person who possesses legitimate power that arises from "exceptional sanctity, heroism, or exemplary character."

Consideration: A "relationship-oriented" leader behavior that is supportive, friendly, and focused on personal needs and interpersonal relationships.

Contingency theory of leadership: A theory advanced by Dr. Fred E. Fiedler that suggests that different leadership styles are effective as a function of the favorableness of the leadership situation least preferred.

Designated leader: The person placed in the leadership position by forces outside the group.

Emergent leader: The person who becomes a group's leader by virtue of processes and dynamics internal to the group.

Formal leader: That individual who is recognized by those outside the group as the official leader of the group.

Great man theory of leadership: The belief that some people are born to be leaders and others are not.

Informal leader: That individual whom members of the group acknowledge as their leader.

Initiating structure: A "task-oriented" leader behavior that is focused on goal attainment, organizing and scheduling work, solving problems, and maintaining work processes.

Leadership: A social (interpersonal) influence relationship between two or more persons who depend on each other to attain certain mutual goals in a group situation.

Least-preferred coworker (LPC): The person with whom the leader least likes to work.

Path-goal theory of leadership: A theory that posits that leadership is path- and goal-oriented, suggesting that different leadership styles are effective as a function of the task confronting the group.

Transformational leader: A leader who moves and changes things "in a big way" by inspiring others to perform the extraordinary.

Visionary leader: A leader who influences others through an emotional and/or intellectual attraction to the leader's dreams of what "can be."

Summary of Learning Outcomes

The Nature of Leadership

Q. What is the nature of leadership and the leadership process?

A. Leadership is a primary vehicle for fulfilling the directing function of management. Because of its importance, theorists, researchers, and practitioners have devoted a tremendous amount of attention and energy to unlocking the secrets of effective leadership. They have kept at this search for perhaps a greater period of time than for any other single issue related to management.

The Leadership Process

Q. What are the processes associated with people coming to leadership positions?

A. Organizations typically have both formal and informal leaders. Their leadership is effective for virtually identical reasons. Leadership and management are not the same. Although effective leadership is a necessary part of effective management, the overall management role is much larger than leadership alone. Managers plan, organize, direct, and control. As leaders, they are engaged primarily in the directing function.

Types of Leaders and Leader Emergence

Q. How do leaders influence and move their followers to action?

A. There are many diverse perspectives on leadership. Some managers treat leadership primarily as an exercise of power. Others believe that a particular belief and attitude structure makes for effective leaders. Still others believe it is possible to identify a collection of leader traits that produces a leader who should be universally effective in any leadership situation. Even today, many believe that a profile of behaviors can universally guarantee successful leadership. Unfortunately, such simple solutions fall short of the reality.

The Trait Approach to Leadership

Q. What are the trait perspectives on leadership?

A. Core traits include the following:

- Drive
- Leadership motivation
- Honesty and integrity
- Self-confidence
- Cognitive ability
- Knowledge of the business
- Other traits: charisma, creativity/originality, and flexibility/adaptiveness

Behavioral Approaches to Leadership

Q. What are the behavioral perspectives on leadership?

A. It is clear that effective leaders are endowed with the “right stuff,” yet this “stuff” is only a precondition to effective leadership. Leaders need to connect with their followers and bring the right configuration of knowledge, skills, ability, vision, and strategy to the situational demands confronting the group.

Situational (Contingency) Approaches to Leadership

Q. What are the situational perspectives on leadership?

A. We now know that there is no one best way to be an effective leader in all circumstances. Leaders need to recognize that how they choose to lead will affect the nature of their followers’ compliance with their influence tactics, and ultimately impacts motivation, satisfaction, performance, and group effectiveness. In addition, the nature of the situation—contextual demands and characteristics of the follower—dictates the type of leadership that is likely to be effective. Fiedler focuses on leader traits and argues that the favorableness of the leadership situation dictates the type of leadership approach needed. He recommends selecting leaders to match the situation or changing the situation to match the leader. Path-goal theory focuses on leader behavior that can be adapted to the demands of a particular work environment and organizational members’ characteristics. Path-goal theorists believe both that leaders can be matched with the situation and that the situation can be changed to match leaders. Together, these theories make clear that leadership is effective when the characteristics and behavior of the leader match the demands of the situation.

Substitutes for and Neutralizers of Leadership

Q. What does the concept of “substitute for leadership” mean?

A. Characteristics of followers, tasks, and organizations can substitute for or neutralize many leader behaviors. Leaders must remain aware of these factors, no matter which perspective on leadership they adopt. Such awareness allows managers to use substitutes for, and neutralizers of, leadership to their benefit, rather than be stymied by their presence.

Transformational, Visionary, and Charismatic Leadership

Q. What are the characteristics of transactional, transformational, and charismatic leadership?

A. In recent years, there has been a renewed interest in key leader traits and behaviors. As organizations face increasing amounts of chaos in their external environments, searches for “the right leader” who can bring about major organizational transformations has intensified. This search once again focuses our attention on a set of “key” motives, knowledge, skills, and personality attributes. Emerging from this search has been the identification of the charismatic and transformational leader.

Leadership Needs in the 21st Century

Q. How do different approaches and styles of leadership impact what is needed now?

A. Leadership in the high-involvement organization differs dramatically from that in the traditional and control-oriented organization. Leaders external to the team have as one of their primary roles empowering group members and the teams themselves to self-lead and self-manage. Leaders internal to the team are peers; they work alongside and simultaneously facilitate planning, organizing, directing, controlling, and the execution of the team’s work.

Although we know a great deal about the determinants of effective leadership, we have much to learn. Each theory presented in this chapter is put into practice by managers every day. None provides the complete answer to what makes leaders effective, but each has something important to offer.

Finally, our understanding of leadership has many shortcomings and limitations. The existing literature is largely based on observations from a Western industrialized context. The extent to which our theories of leadership are bound by our culture, limiting generalization to other cultures, is largely unknown. Cross-cultural leadership research will no doubt intensify as the global economy becomes an ever more dominant force in the world

Section Review Questions

1. Define leadership and distinguish between leadership and management.
2. Discuss the processes associated with people coming to positions of leadership.
3. Discuss the different forms of power available to leaders and the effects associated with each.
4. It has been observed that effective leaders have the “right stuff.” What traits are commonly associated with leader emergence and effective leaders?
5. Both the Ohio State University and University of Michigan leadership studies identified central leader behaviors. What are these behaviors, and how are they different from one another?
6. Blake and Mouton’s work with the Leadership Grid® identified several leadership types. What are they, and how does this leadership model look from the perspective of situation theories of leadership?
7. Identify and describe the three situational variables presented in Fiedler’s contingency theory of leadership.
8. What are the four leadership behaviors in the path-goal theory of leadership?
9. Discuss the differences between the internal and external leadership roles surrounding self-managed work teams.
10. What are substitutes for leadership? What are neutralizers? Give an example of each.
11. What are the distinguishing features of the transformational and the charismatic leader?

Note

Changes made:

- Changed the "Key Terms" into a construct that works in the Libretexts system, and corrected typo.
- Removed hard-coded chapter numbers
- Changed Summary of Learning Outcomes so that instead of a numbered set of questions, it is structured as Q and A.
- Since it was missing, added a list of core traits as an answer to the question "What are the trait perspectives on leadership?".
- Changed Exhibit to Figure using Libretexts template and adding any necessary accessibility text.
- Removed the Critical Thinking case about United Airlines.
- Deleted Management Skills Application Exercise.
- Deleted Managerial Decision Exercises.
- Converted bold text to headings per accessibility review.
- Changed "Chapter Review Questions" to "Section Review Questions"

This page titled 4.6.2.2.11: Summary is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

4.7: Legal Considerations

In this section we will cover some of the intellectual property considerations associated with technology projects. These topics include:

- Patents
- Trade secrets
- Trademarks

Note that we will not be covering copyright considerations.

Patents

Technology companies place a premium on innovation and one of the ways that innovations are codified with special legal status is with patents.

Technology companies usually offer patent bonuses to innovative engineers associated with their work on a product that is being developed and brought to market. Many engineers think that their work is innovative and they may be motivated by the patent bonuses or by their professional desire to have patents on their resumes.

Companies typically have a budget for patent work for each calendar year. Filing for patents is subject to the oversight of an executive panel (patent review board) and budget constraints. Thus, even though many ideas may be worthy of being patented, that may not be possible due to budgetary considerations.

Trade Secrets

When some ideas are not patented, there is some protection provided by protecting the intellectual property through the use of trade secrets.

Additionally, some material may be fundamentally so important to a company that it does not want to make disclosures through the patent process and wanhts to use the trade secret mechanism instead.

Trademarks

Early in the product development cycle, it is important to perform a trademark search for the planned potential trademarks, such as the product name or product family name. One of the reasons for this is that the software for the product may need to make checks for the product platform on which it is running and do different things based on which platform it is. The product name or family is often printed on the product itself and appears in product collateral such as data sheets, white papers, etc.

If the trademark search is not performed, or the results of it are ignored or the company decides to take a risk, the downstream results can be very undesirable. For example, after the product ships into the market, receiving a cease and desist request from another company's legal department for trademark infringement can have very negative consequences.

This page titled 4.7: Legal Considerations is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

4.7.1: Patents

In this section we will cover:

- the patenting process
- patent disclosures and
- patent cross-licensing agreements.

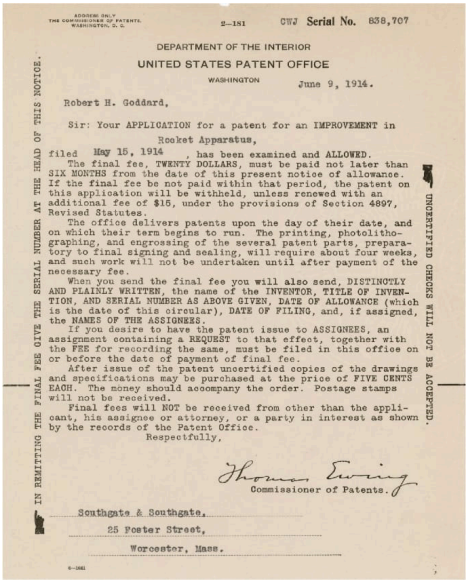
This page titled 4.7.1: Patents is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

4.7.1.1: The Patenting Process

Learning Objectives

After completing this section, you will be able to

- Describe the steps in the patent application process.
- Discuss the importance of proper claims drafting.



ADDRESS ONLY THE COMMISSIONER OF PATENTS, WASHINGTON, D. C.

2-181 CWJ Serial No. 838,707

DEPARTMENT OF THE INTERIOR
UNITED STATES PATENT OFFICE
WASHINGTON June 9, 1914.

Robert H. Goddard,

Sir: Your APPLICATION for a patent for an IMPROVEMENT in Rocket Apparatus, filed May 15, 1914, has been examined and ALLOWED.

The final fee, TWENTY DOLLARS, must be paid not later than SIX MONTHS from the date of this present notice of allowance. If the final fee be not paid within that period, the patent on this application will be withheld, unless renewed with an additional fee of $15, under the provisions of Section 4897, Revised Statutes.

The office delivers patents upon the day of their date, and on which their term begins to run. The printing, photolithographing, and engrossing of the several patent parts, preparatory to final signing and sealing, will require about four weeks, and such work will not be undertaken until after payment of the necessary fee.

When you send the final fee you will also send, DISTINCTLY AND PLAINLY WRITTEN, the name of the INVENTOR, TITLE OF INVENTION, AND SERIAL NUMBER AS ABOVE GIVEN, DATE OF ALLOWANCE (which is the date of this circular), DATE OF FILING, and, if assigned, the NAMES OF THE ASSIGNEES.

If you desire to have the patent issue to ASSIGNEES, an assignment containing a REQUEST to that effect, together with the FEE for recording the same, must be filed in this office on or before the date of payment of final fee.

After issue of the patent uncertified copies of the drawings and specifications may be purchased at the price of FIVE CENTS EACH. The money should accompany the order. Postage stamps will not be received.

Final fees will NOT be received from other than the applicant, his assignee or attorney, or a party in interest as shown by the records of the Patent Office.

Respectfully,

Thomas Ewing
Commissioner of Patents.

Southgate & Southgate,
25 Foster Street,
Worcester, Mass.

IN REMITTING THE FINAL FEE GIVE THE SERIAL NUMBER AT THE HEAD OF THIS NOTICE.

UNCERTIFIED CHECKS WILL NOT BE ACCEPTED.

Figure 4.7.1.1.1: Notice of Examination and Approval of Patent Application. (credit: National Archives and Records Administration via Wikimedia Commons / Public domain)

Accessibility Note

Figure 4.7.1.1.1 is an image of an approved patent letter. The beginning of the letter to Robert H. Goggard reads "Your application for improvement in Rocket Apparatus dated May 15, 1914 has been examined and allowed."

Applying for a Patent

Much is at stake in the process of applying for a patent. Depending upon how you draft the claims and write the specification of your application, you could win or lose patent rights at any point in the examination process. In addition, any patent rights you win can be worth a considerable amount of money, and can be enforced by the U.S. federal courts.

An applicant must first determine whether to file an application for a utility, design, or plant patent. Then, the applicant must determine their filing status: large entity, small entity, or the new category of micro entity created by the America Invents Act of 2011 (AIA)(opens in new window). Small entities, which are universities, non-profits, and small businesses with fewer than 500 employees who also meet certain other criteria, receive a 50 percent discount on the application fees paid by large entities. Micro entities, which are small entities who have a gross income less than three times U.S. median household income or meet certain other criteria, receive a 75 percent discount. These discounts on United States Patent and Trademark Office (USPTO) fees for small and micro entities do not apply to the attorneys' fees often involved in applying for a patent, which can be significant.

Definition: Utility patent

A utility patent is a type of patent that may be granted to anyone who invents or discovers a new and useful process, machine, article of manufacture, or composition of matter, or any new and useful improvements of these.[1]

Definition: Design patent

A design patent is a type of patent that may be granted to anyone who has invented a new, original ornamental design for an article of manufacture. The ornamental characteristics must be embodied in or applied to such an article. The subject matter may relate to the configuration or shape of an article, surface ornamentation applied to it, or the combination of both. A surface ornamentation design is inseparable from the article to which it is applied and cannot exist alone. It must be a definite pattern applied to an article of manufacture.[1]

Provisional vs. Nonprovisional Patent Application

Finally, the applicant must decide whether to file an abbreviated, "provisional" patent application or a complete, "nonprovisional" one. A provisional application consists only of the specification describing the invention for which a patent is sought, as well as any drawings that might be necessary to understand the invention. A provisional application is not subject to examination, and is viable for one year.

The chief benefits of a provisional application are twofold. It is much less expensive to file than a regular nonprovisional application—only $1,000 in official fees for large entities, $500 for small entities, and $250 for micro entities. It also gives the inventor the benefit of the earlier provisional filing date of the nonprovisional application based on the same specification, while measuring the term of the patent from the nonprovisional application's filing date.

In a sense, the provisional application serves as a placeholder for up to one year while the inventor does all the prior art searching, claims drafting, and other work required of a full and complete nonprovisional application. This priority date placeholder function can be important given that many companies and inventors are working on new products and services aimed at the same markets. This is especially so under the United States' new "first inventor to file" rather than "first to invent" priority regime, which some fear may give an edge to large companies with the legal and financial resources to file early and often. To the extent this concern may be valid, provisional applications can serve to mitigate any large company filing advantage.

The main disadvantage of a provisional application is that because the written description cannot be changed when filing the follow-on, nonprovisional application, all the subsequent claims in that follow-on application must be completely consistent with the earlier description language. As most inventors can attest, the understanding of an invention and its potential market—as well as the claims best suited to protect it—inevitably evolves and matures over time.

Here's how an experienced entrepreneur describes the value of a provisional application in the "first inventor to file" environment:

> "What I recommend to every entrepreneur is that if you write a new, non-obvious, and useful line of code, put it in the specification of a provisional patent application and file it. It only costs $1,000 at most. You then have a year in which to honestly evaluate where the industry is going, where the market is going, and whether your invention is truly valuable enough to pursue a patent for it."

A year might not seem very long. But a lot can happen to markets and to technology in that amount of time. Thanks to the rapid rise of smartphones, for example, it took merely a year for netbooks to go from the "next big thing" to "who cares?"

Note

The specification of a nonprovisional utility patent application should have the following sections in order:

1. **Title of the invention** — should be as short and specific as possible (no more than 500 characters), should appear as a heading on the first page of the specification (if not otherwise appearing at the beginning of the application).

2. **Cross-reference to related application**, if any (may be listed on a data sheet, either instead of or along with being listed in the specification).
3. **Statement of federally sponsored research or development** (if any) — for an invention made with U.S. government support and for which the U.S. government has certain rights, **35 U.S.C. 202**(c)(6) and **37 CFR 401.14**(f)(4) require the specification to contain a statement that the invention was made with such support. Here's a sample wording: "This invention was made with government support under [*identify the contract*] awarded by [*identify the federal agency*]. The government has certain rights in the invention."
4. Names of the parties to a joint research agreement, if the claimed invention resulted from activities within the scope of such an agreement.
5. Reference to a "sequence listing," table, or computer program listing appendix submitted on an optical disc or via the USPTO electronic filing system and an incorporation by reference of the material on the optical disc. The total number of discs, including duplicates and the files on each disc, must be specified.
6. **Background on the invention**.
7. **Brief summary of the invention** — indicating its nature and substance (which may include a statement of the object of the invention), preceding the detailed description. The summary should be commensurate with the invention as claimed, and any object recited should be that of the invention as claimed.
8. **Brief description of views of the drawing** (if any) — a description of the drawing views by specifying the numbers of the figures.
9. **Detailed description of the invention** — description of the invention including the manner and process of its making and use;
    1. Required in full, clear, concise, and exact terms so that anyone skilled in the technology area to which the invention belongs or is most nearly connected to, can make and use it.
    2. Set forth the precise invention for which a patent is requested, distinguishing it from other inventions and from what is old. It must describe completely a specific embodiment of the process, machine, manufacture, composition of matter, or improvement invented; and must explain the best mode of carrying out the invention.
    3. For an improvement, the specification must point out the part(s) of the process, machine, manufacture, or composition of matter to which the improvement relate(s). The description should be confined to the specific improvement and to the parts that necessarily cooperate with it or are necessary to a complete understanding or description.
10. **At least one claim** — particularly pointing out and distinctly claiming the subject matter regarded as the invention. The portion of the application setting forth the claims defines the scope of the patent protection and which questions of infringement are judged by the courts.
    1. More than one claim may be presented (they must differ substantially and not unduly multiplied). One or more claims may be presented in dependent form, referring back to and further limiting other claims in the same application. Any dependent claim referring back to more than one other claim is called a "multiple dependent claim."
    2. Multiple dependent claims shall refer to such other claims in the alternative only. A multiple dependent claim shall not serve as a basis for another multiple dependent claim. Claims in dependent form will be understood to include all limitations of the claim referenced by the dependent claim. A multiple dependent claim will be considered to incorporate all the limitations of each of the particular claims in relation to which it is being considered.
    3. The claims must conform to the invention as set forth in the rest of the specification, and any wording in the description must be clearly supported so that their meaning may be understood in reference to the description.
11. **Brief abstract of the disclosure** — must appear on a separate page (preferably following the claims), should be a single paragraph (150 words or less) under the heading "Abstract of the Disclosure."
12. Sequence listing (if any).[1]

The Critical Importance of the Claims

When it comes to filing a full nonprovisional application, the most critical task of the inventor (and patent attorney) is to draft the claims. This is an art in and of itself, one that will determine the inventor's rights and an infringer's liability.

Each claim serves as a stand-alone definition of the patent coverage—as a sort of mini-patent unto itself. After a patent is granted, a claim is the only thing that someone can be accused of infringing. Accordingly, broad claim language is essential if a commercially meaningful patent is to be obtained.

On the other hand, the broader the claim, the more likely it is that prior art may exist somewhere that anticipates the claim or renders it obvious and, therefore, invalidates the claim. So the competent draftsperson submits a dozen or more claims, typically moving from broader to narrower scope, in case the broader claims turn out to be disallowed by the USPTO or later invalidated

by a court.

Some applicants make the opposite mistake and draft claims that are too narrow—"a programmable multifunction computer in the shape of a metal clamshell," for example. Such a claim is of little value because others can easily design around it simply by using plastic or other nonmetal materials in the design. Some claims are so narrow they are even referred to as "picture claims" because they paint a literal and limiting picture of the invention. You may get a patent with such picture claims, but it probably won't be worth the paper it's printed on.

The strongest claims usually define powerful functionality. They capture something fundamental or seminal in the functioning of the invention.

A good example is the claims in the aforementioned Amazon one-click patent. The algorithms for "one-click" shopping are not very complicated. Yet the claims capture a very profound functionality—the idea that a user should not have to type in credit card and shipping information every time he or she wants to buy something. The consumer-friendly functionality captured in that patent's claims has won Amazon a lot of customers and made it a lot of money.

Reviewing the Patent Application

Once the patent application with all necessary forms is filed and all fees are paid, the **patent examiner** will review the patent application to determine if the invention meets the statutory requirements for patentability. The examiner will also conduct a search of patent and other databases to determine if the invention appears to be truly novel and non-obvious. Once a determination has been made, the patent examiner will send the applicant a document known as a **first office action**, in which the patent examiner approves, rejects, or requires additional information about the claims and/or other elements of the application.

If all the claims are allowed, which is rare in a first office action, the patent prosecution process is complete. A more common next step requires the applicant to respond to the patent examiner's rejections and requests for more information. In this response, the applicant will address all of the examiner's concerns and either rebut them or amend the application, typically by revising some or all of the claims.

But according to the Manual of Patenting Examination Procedure (MPEP) Section 1.121, "No amendment may introduce new matter into the disclosure of an application."[li] Thus, an applicant cannot file an application disclosing a new compound and how to make it, and then later, after discovering that the compound's structure and method are incorrect, amend the description.

After responding to the office action by perhaps amending the application, the examination process continues until the patent examiner either allows all of the claims or finally rejects them. If all claims are allowed, then the patent prosecution process is complete. If some claims are finally rejected, then the applicant's options at this point are more limited.

According to MPEP Section 706.07, "In the second or any subsequent examination or consideration by the examiner, the rejection can be made final."[lii] At that point, the applicant must usually choose one of three approaches: Cancel the rejected claims, leaving only allowed claims to issue as a patent; continue the fight by filing what is called a "continuation application" and drafting new claims; or appeal to the Patent Trial and Appeals Board arguing that the patent examiner erred in not allowing the claims.[liii] Any appeal to the Patent Trial and Appeals Board, as with any other office action, must occur within six months.[liv]

Finally, if the patent has been vetted through the patent prosecution process and is ready for issuance, the patent holder must pay the applicable fee. If the applicant pays the issue fee, the USPTO will issue the patent in due course. Additionally, utility patents are subject to maintenance fees. These fees are due 3 years and 6 months, 7 years and 6 months, and 11 years and 6 months from the date of the original patent grant. Plant and design patents are also required to pay maintenance fees.[lv]

Footnotes

li. United States Patent and Trademark Office. (2010, July 21). *Manual of Patent Examining Procedure* § 1.121. Retrieved from USPTO 1.121.

lii. United States Patent and Trademark Office. (2010, July 21). *Manual of Patent Examining Procedure* §.0 Retrieved from USPTO 706.07.

liii. 37 CFR Part 41. Department of Commerce, United States Patent and Trademark Office. (2008). *Rules of practice before the*

board of patent appeals and interferences in ex parte appeals. Retrieved from website: USPTO rules of practice.

liv. 37 CFR § 1.134

lv. 37 CFR § 1.362

References

1. Applying for Patents | USPTO. Accessed March 14, 2026.

Terms of Use for USPTO websites | USPTO.

Pursuant to federal law, most government-produced materials appearing on this website are not subject to copyright restrictions within the United States and are therefore in the public domain. Public domain information may be freely distributed and copied, but it is requested that in any subsequent use the United States Patent and Trademark Office (USPTO) be given appropriate acknowledgement (e.g., “Source: United States Patent and Trademark Office, USPTO link”). The USPTO reserves the right to assert copyright protection internationally.

Note

Changes made:

- Converted Figures to Libtretexts Template format.
- Spelled out United States Patent and Trademark Office (USPTO.
- Added definitions of utility and design patents.
- Added the structure of a patent.
- Deleted the image of the board game patent.
- Changed Reference list to a numbered list per accessibility review.

This page titled 4.7.1.1: The Patenting Process is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

4.7.1.2: Patent Disclosures

608 Disclosure [R-11.2013]

To obtain a valid patent, a patent application as filed must contain a full and clear disclosure of the invention in the manner prescribed by **35 U.S.C. 112(a)**. The requirement for an adequate disclosure ensures that the public receives something in return for the exclusionary rights that are granted to the inventor by a patent. All amendments and claims must find descriptive basis in the original disclosure. No new matter may be introduced into an application after its filing date. Applicant may rely for disclosure upon the specification with original claims and drawings, as filed. See also **37 CFR 1.121(f)** and **MPEP § 608.04**.[1]

This refers to disclosures within the patent application itself.

Defensive Disclosures

The **IBM Technical Disclosure Bulletin** was a technical publication produced by IBM between 1958 and 1998. The purpose of the IBM Technical Disclosure Bulletin was to disclose inventions that IBM did not want their competitors to get patents on. The Bulletin was a form of defensive publication. By publishing the details of how to make and use the invention, patent examiners could have a searchable source of prior art that they could cite against subsequent patent applications filed by others on the same or similar inventions.[2]

A **defensive publication**, or **defensive disclosure**, is an intellectual property strategy that can be employed used to prevent another party from obtaining a patent on a product, apparatus or method for instance. The strategy consists in disclosing a description and/or drawing of the product, apparatus or method so that it enters the public domain and becomes prior art.

Another method for making defensive disclosures is by the use of Technical Disclosure Commons (TDCommons).[4] This mechanism allows the disclosure of information to prevent patenting of the ideas in the disclosure as they would be in prior art that patent examiners assess for novelty.

References

1. 608-Disclosure. The United States Patent and Trademark Office. Accessed February 16, 2026.
2. Wikipedia - IBM Technical Disclosure Bulletin. Accessed February 16, 2026. Text is available under the Creative Commons Attribution-ShareAlike 4.0 License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.
3. Wikipedia - Defensive Publication. Accessed March 13, 2026. Text is available under the Creative Commons Attribution-ShareAlike 4.0 License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.
4. Technical Disclosure Commons. Accessed March 13, 2026.

This page titled 4.7.1.2: Patent Disclosures is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

4.7.1.3: Patent cross-licensing

Definition: Cross-licensing agreement

A **cross-licensing** agreement is a contract between two or more parties where each party grants rights to their intellectual property to the other parties.[1]

Why do cross-licensing?

Suppose that you are the dominant supplier in a particular product market with a substantial patent portfolio. A start-up company could enter the market having a very small patent portfolio, but if it sues the dominant supplier for patent infringement and succeeds, it could secure an injunction against the dominant supplier from further sales of its products. This scenario is unlikely, but is a risk. A more likely scenario is that the dominant supplier sues the start-up for patent infringement after the start-up company gains enough market share.

To avoid these scenarios, the companies will work out a patent cross-licensing agreement. As part of the agreement, royalties may flow from the company with the smaller, less valuable patent portfolio to the company with the larger, more valuable patent portfolio that is relevant to their area of competition.

Patent law

In patent law, a **cross-licensing agreement** is an agreement according to which two or more parties grant a license to each other for the exploitation of the subject-matter claimed in one or more of the patents each owns. Usually, this type of agreement happens between two parties in order to avoid litigation or to settle an infringement dispute. Very often, the patents that each party owns covers different essential aspects of a given commercial product. Thus by cross licensing, each party maintains their freedom to bring the commercial product to market.[1]

Each party, therefore, is able to practice the inventions covered by the patents included in the agreement. This benefits competition by allowing each company more freedom to design products covered by the other's patents without provoking a patent infringement lawsuit.[1]

Caution

Companies that enter into cross-licensing agreements must be careful not to violate antitrust laws and regulations. This can easily become a complex issue, involving (as far as the European Union is concerned) Art. 101 and 102 of the Treaty on the Functioning of the European Union (TFEU), previously Art. 81 and 82 of the EC Treaty, (abuse of dominant position, etc.) as well as licensing directives, cartels, etc.[1]

Some companies file patent applications primarily to be able to cross license the resulting patents, as opposed to trying to stop a competitor from bringing a product to market.[1]

One of the limitations of cross licensing is that it is ineffective against patent holding companies. The primary business of a patent holding company is to license patents in exchange for a monetary royalty. Thus, they have no need for rights to practice other companies' patents.[1]

References

1. Wikipedia - Cross-licensing. Accessed February 16, 2026. Text is available under the Creative Commons Attribution-ShareAlike 4.0 License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.

This page titled 4.7.1.3: Patent cross-licensing is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

4.7.2: Trade Secrets

In this section we will cover:

- Trade secret protections
- Foundations of trade secret law
- Elements of a trade secret
- Secrecy requirement
- Misappropriation of trade secrets
- Remedies for the misappropriation of trade secrets

This page titled 4.7.2: Trade Secrets is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

4.7.2.1: Trade Secret Protections

Learning Objectives

After completing this section, you will be able to

- Define trade secrets.
- Identify the advantages and disadvantages of protecting intellectual property through trade secrets.

What Is a Trade Secret?

Trade secret law is a source of protection for intellectual property that serves as an alternative to patent or trademark law. Whereas patent and trademark law require that the intellectual property to be protected be publicly "disclosed" (for example, through a patent application), trade secret law requires precisely the opposite—that is, that the intellectual property to be protected *not* be publicly disclosed. The subject matter of a trade secret may be virtually any information that is of value as a result of not being generally known.

There is enormous **economic value** in trade secrets—for companies and countries. A study released in March of 2014 estimated that trade secret misappropriation costs the world's top 40 economies between 1 percent and 3 percent of their gross domestic product each year.[i]

Unlike copyright and patent law, there is no formal requirement of "novelty" or "tangibility" under trade secret law. Rather, trade secret protection is extended to information that has independent economic value by virtue of not being generally known or readily ascertainable by others, and which has been subject to reasonable efforts to avoid public disclosure.

Advantages and Disadvantages of Trade Secret Law

There are both advantages and disadvantages to protecting intellectual property through the use of trade secret law. The decision of whether to use trade secret law as opposed to other legal means of protecting knowledge or information is governed primarily by the nature of the intellectual property to be protected.

Advantages

One of the main advantages of trade secret law is that it provides indefinite future protection. Unlike the 20-year time limit that underlies the protection granted to a patent holder, there is no definite time limit placed on the protection granted to a trade secret owner. This indefinite future protection granted to trade secrets is appealing, assuming that the intellectual property to be protected is likely to remain generally unknown well into the future. For example, if the intellectual property to be protected is a method of manufacturing a product that can be determined from examining the product, then the best way to protect the new method of manufacture might be through a patent. On the other hand, if the new method of manufacturing the product cannot be determined from examination of the product, then the best way to protect the intellectual property might be to treat it as a trade secret.

Disadvantages

Although there is no definite expiration of the protection granted to trade secrets, the protection prohibits only the disclosure or use of the trade secret by one to whom the secret was disclosed in confidence. Unlike patent law, trade secret law does not offer any affirmative protection against the use of the same intellectual property that is independently derived or reverse engineered by a competitor. Consequently, if the intellectual property to be protected is a new method of manufacturing a product, and if that method of manufacture is ascertainable from an examination of the product, then the best form of protection for the intellectual property is most likely that offered by patent law.

The same analysis applies to intellectual property that is likely to be independently derived by a competitor. For as the U.S. Supreme Court noted in Bonito Boats v Thunder Craft Boats,[ii] state trade secret law has never given the holder of a trade secret protection against reverse engineering by the public or a competitor. The Supreme Court commented that in order to receive protection from reverse engineering, the holder of the intellectual property must seek the protection of federal patent law.

Table 4.7.2.1.1*: Pros and Cons of Trade Secret Law*

Pros	Cons
Trade secret law provides indefinite future protection, so long as the trade secret stays a secret. Trade secret protection has no expiration date.	If someone came up with the same idea on their own, the trade secret is no longer protected by law.
Trade secret protection prevents the disclosure or use of the trade secret by one to whom the secret was disclosed in confidence.	Trade secret law does not give the holder of a trade secret protection from reverse engineering by the public or a competitor.

Whereas patent and trademark law involve an affirmative assertion of an intellectual property right, trade secret law only actively operates after a misappropriation or threatened misappropriation of the trade secret has occurred. Due to this difference in the way trade secret law protects intellectual property, the protections given to trade secrets are largely dependent on classifications of information made by courts. Thus, the foundation and development of trade secret law in the United States is inherently tied to the statutes protecting trade secrets, and the court decisions defining and interpreting the statutory protections created for trade secrets.

Notes

i. *Economic Impact of Trade Secret Theft*," PricewaterhouseCoopers.
ii. 489 U.S. 141 (1989).

Note

Changes to the original OER material:

- Accessibility audit called out link issues with roman numeral footnotes. Forked the page in an attempt to correct the issues.
- Created a numbered list using lower case roman numerals to match the needed numbering. Transferred content from the original "Footnotes" to "Notes." Deleted "original "Footnotes" section.
- Removed the link from note i since it was a bad link. The link for note ii is valid.
- Needed to correctly label the Table using the chapter numbering in this textbook using Libretexts constructs. Changed to embed it in the table's caption.

This page titled 4.7.2.1: Trade Secret Protections is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

4.7.2.2: The Foundations of Trade Secrets Law

Learning Objectives

After completing this section, you will be able to

- Identify the roots of trade secret law in the United States.
- Explain the purpose of the Uniform Trade Secrets Act.

Unlike patents and copyrights that are issued in accordance with federal law, trade secret protection until very recently emanated from state law. But on May 11, 2016, the federal Defend Trade Secrets Act (DTSA) was signed into law, thereby creating a federal cause of action for trade secret misappropriation that largely mirrors the current state of the law under the Uniform Trade Secrets Act (UTSA), which has been adopted by 48 states.

Trade secret law in the United States is largely rooted in the English legal tradition, and was first incorporated through the court-made common law. The general consensus is that trade secret law originated in England in the early 1800s.[iii] Early cases dealt with trade secret concepts not in modern trade secret vernacular but instead through the guise of jurisdictional or restraint of trade terms. The trade secret concept began in American jurisprudence in the case of Vickery v, Welch, 36 Mass. 523 (1837). The Massachusetts Supreme Court first expounded on trade secret law in modern terms in *Peabody v. Norfolk* , 98 Mass. 452 (1868). In the 145 years since the Peabody decision, the courts, including numerous federal courts, have addressed trade secret issues and have contributed to the current body of trade secret law. However, because each state may regulate trade secret law independently, each state has developed its own legal regime, involving a mixture of both statutes and common law, to regulate and protect trade secrets.

Uniform Trade Secrets Act

In an effort to harmonize and standardize trade secret law in the United States, the Uniform Law Commission[iv] published the Uniform Trade Secrets Act (UTSA) in 1979. The UTSA sought to create a uniform trade secret law regime from state to state. However, the UTSA itself was not binding upon states at publication. In order for the UTSA to be successful, the states themselves needed to adopt and enact the model UTSA statute created by the commission. In 1980, Minnesota became the first state to adopt the UTSA. Since 1980, most states have adopted the UTSA, many with the 1985 amendments to the act. New York and Massachusetts remain the last states that have yet to adopt the UTSA. Although the UTSA harmonized various state law approaches to trade secrets, each state enacts the UTSA with any modifications or customizations the state legislature desires to make. Thus, although most states have adopted the UTSA, each state may have a slightly different adaptation of the UTSA.

One important provision of the UTSA is section 7, which deals with the interaction of the UTSA and state law. Section 7 preempts certain state law causes of action. Section 7 "displaces conflicting tort, restitutionary, and other law...providing civil remedies for misappropriation of a trade secret."

Trade Secret Misappropriation

Although this section seems to indicate that the UTSA broadly replaces all state laws dealing with **misappropriation** of trade secrets, there is debate as to what constitutes "conflicting tort, restitutionary, and other law." Additionally, what it means for a common law remedy to be "based on" misappropriation of a trade secret is often contested. Owners seeking to recover damages for misappropriation generally argue that only common law causes of action that have elements identical to the UTSA are preempted, while defendants argue that any cause of action used to protect trade secrets is preempted because it is duplicative of the UTSA. Despite this potential conflict between the UTSA and state common law, this chapter focuses on the main body of trade secret law as governed by the UTSA.

Notes

iii. Michael Risch, Why Do We Have Trade Secrets?, 11 Intellectual Property L. Rev. 1 (2007) (citing Restatement (Third) of Unfair

Competition § 39 cmt. a (1995).

iv. The Uniform Law Commission, also known as the National Conference of Commissioners on Uniform State Laws, is a nonpartisan, nonprofit organization that promotes the uniformity of law from state to state. The Commission publishes “model laws” that states are then able to adopt and enact.

Note

Changes to the original OER material:

- Accessibility audit called out link issues with roman numeral footnotes. Forked the page in an attempt to correct the issues.
- Created a numbered list using lower case roman numerals to match the needed numbering. Transferred content from the original "Footnotes" to "Notes." Deleted "original "Footnotes" section.

This page titled 4.7.2.2: The Foundations of Trade Secrets Law is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

4.7.2.3: Elements of a Trade Secret

Learning Objectives

After completing this section, you will be able to

- Identify the key criteria used to qualify as a trade secret.
- Explain how ideas can be protected as trade secrets.

To qualify as a trade secret, the information in question must meet two essential criteria. First, the information must attain its value from the fact that it is not generally known.

Second, the owner of that information must take reasonable efforts to maintain its secrecy. Absent these two elements, information does not qualify as a trade secret and is not entitled to trade secret protections that prevent or remedy misappropriation under the UTSA.[v]

Trade secrets are not limited to particular subject matters, although the knowledge or information to be protected need not take any particular form. A trade secret can be information or knowledge in the form of a formula, pattern, compilation, program, device, method, technique, or process. Although the UTSA defines what may be a trade secret, various courts interpreting the UTSA have come to different conclusions on what type of material should be considered a trade secret. For example, whereas courts have found that an insurer's database is a trade secret under Wisconsin law, a similar database was found not to be a trade secret in Rhode Island.[vi]

The Value of Secrets

For information or knowledge in any of the above forms to be considered a trade secret, it must derive independent economic value because it is not generally known to, nor readily ascertainable by, other persons who can profit from its disclosure or use. Some examples of information courts have found to be trade secrets include computer software, sales information, customer information, and manufacturing formulas. However, merely because information is not known to the public does not necessarily make the information a trade secret. For example, information that is not generally known by the public but is known by different manufacturers in the same industry likely does not qualify as a trade secret.[vii]

Material that is ascertainable through public sources generally does not derive independent economic value justifying trade secret protection. If the information in question cannot be quickly or easily ascertained from examining or testing the publicly available product, then it is more likely to be a trade secret.[viii] Conversely, if the information may be discovered through public observation of the product or the company's displays, it is unlikely to be a trade secret.[ix] For example, customer lists are not considered trade secrets when the identity of the customers could potentially be discovered through public sources. However, identical information if not available through public sources could constitute a trade secret.[x] Additionally, processes or systems that are simply a compilation of known information generally do not rise to the level of trade secrets. The courts, however, are still quite divided on this issue.

Protecting Ideas as Trade Secrets

In addition to information, certain ideas may constitute trade secrets. Although ideas were not protected as trade secrets under common law, the UTSA provides protection for certain novel and concrete ideas. For example, remedies for misappropriation of an idea may be permitted based on the existence of a confidential relationship. The elements that must be proven under this common theory are: (1) a novel or original idea; (2) reduced to concrete form; (3) disclosed to the defendant in a confidential relationship, i.e., one in which it would be implied that use would not be made without compensation; and, (4) used by the defendant.

Protection of an idea may also be permitted on the basis of a contractual relationship with the defendant. Parties may enter into a contract to protect an idea being submitted or disclosed to another person or company. The major benefit of a contract to the

submitter of an idea is evidence of a protected disclosure that is deserving of trade secret protections. The major benefit of a contract to the recipient of an idea is the establishment of the ground rules under which compensation, if any, will be payable to the submitter.

Corporations that receive unsolicited ideas many times grapple with how to best protect use of those unsolicited ideas. To protect themselves from the uncertainties of trade secrets law, the idea must be converted from an unsolicited idea to a solicited idea. This may be accomplished by contacting the submitter and alerting them that the idea will be considered only with the understanding that the use of the idea, and decisions regarding compensation for that idea, rest solely with the corporation.

Whether the trade secret takes the form of information or an idea, the owner of that trade secret must show not only that the trade secret was not generally known or readily ascertainable, but also that the lack of general knowledge of the trade secret conferred some value to the owner. When the owner derives a profit from licensing the use of the protected information, the value to the owner is quite apparent. However, in cases where the owner is seeking to show he was injured in the marketplace or through the loss of profits after costly research and development, the value may only become apparent by inference. But even if information is not generally known and gives the owner some value, the information is not a trade secret unless the owner takes steps to protect the secrecy of the information.

Notes

v. The UTSA defines a trade secret as “information, including a formula, pattern, compilation, program, device, method, technique, or process, that (i) derives independent economic value, actual or potential, from not being generally known to, and not being readily ascertainable by proper means by, other persons who can obtain value from its disclosure or use, and (ii) is the subject of efforts that are reasonable under the circumstances to maintain its secrecy.”
vi. Compare *American Fam. Mut. Ins. Co. v. Roth* 485 F.3d 930, 933 (7th Cir. 2007); *APG Inc. v. MCI Telecommunications Corp.* 436 F.3d 294, 307 (1st Cir 2006).
vii. See e.g. *Ruckelshaus v. Monsanto Co.*, 467 U.S. 986, 1002 (1984); *Speedry Chemical Products, Inc. v. Carter’s Ink Co.*, 306 F.2d 328, 331 (2d Cir 1962).
viii. See *Morlife, Inc. v. Perry*, 56 Cal. App. 4th 1514 (1st Dist. 1997).
ix. See *MicroStrategy Inc. v. Business Objects*, S.A., 331 F. Supp. 2d 396,417 (E.D. Va. 2004).
x. Compare *Inflight Newspapers, Inc. v. Magazines In-Flight, LLC,* 990 F. Supp. 119, 129-30 (E.D.N.Y. 1997) (holding that the plaintiff’s customer lists were not trade secrets because the customer identity could be easily found through publicly available means, such as the Internet, trade shows, and trade directories) with *North Atl. Instruments, Inc. v. Haber*, 188 F.3d 38, 44 (2d Cir. N.Y. 1999) (finding that a customer list kept in confidence may be treated as a trade secret).

Note

Changes to the original OER material:

- Accessibility audit called out link issues with roman numeral footnotes. Forked the page in an attempt to correct the issues.
- Created a numbered list using lower case roman numerals to match the needed numbering. Transferred content from the original "Footnotes" to "Notes." Deleted "original "Footnotes" section.

This page titled 4.7.2.3: Elements of a Trade Secret is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

4.7.2.4: The Secrecy Requirement

Learning Objectives

After completing this section, you will be able to

- Explain the steps that businesses can take in order to protect trade secrets.
- Describe how trade secrets can be accessed through passive and active disclosure.

The second element necessary to gain trade secret protection is evidence that the trade secret's owner made a reasonable effort to maintain secrecy. This flexible standard set out by the UTSA is often a point of contention, which results in litigation. To determine whether a piece of information rises to the level of a trade secret, courts often engage in a highly factual inquiry into the security measures taken by the owner to protect a trade secret. For as the Supreme Court has noted, if an owner has not taken steps to protect the trade secret, their proprietary interests may be diminished.[xi]

Under Company Lock and Key

Because the UTSA may be varied by any state, the level of secrecy required varies from state to state. For example, a version of the UTSA adopted by some states contains a provision that reduces the burden of secrecy placed on the owner of a trade secret. Specifically, this version of the UTSA provides that the existence of a trade secret is not negated merely because an employee or other person has acquired the trade secret without express or specific notice that it is a trade secret if, under all the circumstances, the employee or other person knows or has reason to know that the owner intends or expects the secrecy of the type of information comprising the trade secret to be maintained. Although this provision tips the scales in favor of trade secret protection, it should not be viewed as removing the affirmative obligation the UTSA places on one seeking protection of trade secrets. In both states that have and have not modified the UTSA in this manner, the measures taken to secure the information are key to finding whether the information will be protected as a trade secret.

Because trade secret protection is only given to information that is subject to reasonable efforts to maintain secrecy, every person or company that desires to protect its trade secrets under the UTSA should adopt and implement a trade secret protection plan. The specific elements of such a plan will vary depending on the nature of the intellectual property to be protected, as well as the circumstances surrounding its development and use. However, in taking reasonable efforts to protect the secrecy of the information, many different factors are considered. Because no rigid formula exists to determine if reasonable efforts to maintain secrecy are in place, courts look to many of the best practices, discussed below, to determine whether an owner of information made reasonable efforts to protect its secrecy.

Protection Plans

Memorializing a trade secret protection plan in writing helps to ensure that the business or person is indeed taking reasonable measures to protect the trade secret. The person or company seeking to protect its intellectual property through the use of trade secret law should maintain a written statement of its trade secret security policy. This statement can be written in the form of a policy manual and helps prevent situations in which "unwritten rules" are followed or rules are enforced in a lax manner and then eventually ignored. Moreover, should litigation become necessary, it is easier to persuade a court that the security procedures were indeed a primary concern if they are enumerated in writing. However, if security measures are merely written, but not followed, it may be difficult to prove reasonable efforts to maintain secrecy.[xii]

A trade secret protection plan should include a program for ensuring that employees are made aware of the trade secret status of information and that the employees are periodically reminded of their confidentiality obligations. Discussions between the employer and the employee may be appropriate. Writing an annual conflict of interest letter to remind management-level employees of their responsibilities in keeping information confidential may also suffice. An essential part of this program is making certain that employees are aware of and familiar with the written trade secret protection plan discussed above.

Employees should also be required to sign confidentiality agreements acknowledging that they have access to trade secret

information, that they will not disclose or misappropriate that information, and that they will report all disclosures. Finally, “exit interviews” should be conducted for employees leaving the company. This interview provides an opportunity to remind employees of their written obligation not to disclose or misuse any trade secret information, and to extend an open door to discuss any issues or concerns that may arise in connection with their future employment.[xiii]

In addition to educating employees, reasonable efforts to protect information often include physically restricting access to trade secrets. Trade secret information should be treated differently from nonproprietary information and should, if possible, be physically segregated so that it is not readily accessible. For example, it is not uncommon to restrict visitor access to sales and administrative headquarters. In addition, access to the intellectual property should be restricted to those with a need to know or access the information. Depending on the nature of the intellectual property to be protected, reasonably restricting access to trade secrets may be as simple as segregating the information in separate file drawers, or may require measures as extreme as designating an entirely separate facility to house the intellectual property.[xiv]

If possible under the circumstances, physical security measures should be instituted, including the use of locks on doors, gates, and cabinets containing trade secrets. These physical security steps should be obvious to both employees and third parties who may have occasion to visit the area in which trade secrets are stored.[xv] These types of security measures communicate that an effort is being made to protect the information.

Strategies for Protecting Trade Secrets

Given the ease with which documents can be reproduced, trade secrets that are reduced to writing should be treated with special care. Documents containing trade secrets should be clearly labeled with shorthand instructions that are consistent with the written trade secret protection plan. For example, documents may be marked “secret” or “confidential.”[xvi] Access to these documents should be monitored and all persons allowed access to the documents should be informed of the meaning of the trade secret designation on the document and the relevant provisions of the written trade secret protection plan. This attention to detail should also be followed in arranging for proper destruction of the documents.[xvii]

Like trade secrets that are reduced to writing, trade secrets stored on a computer system are particularly susceptible to theft. Access to a computer system containing trade secrets should be restricted in accordance with the written trade secret protection plan and all physical computer- related items such as discs and printouts should contain the same trade secret legend used to identify documents containing intellectual property.

When allowing third parties, be they members of the public or business associates, access to a facility in which trade secrets are stored or used, care must be taken to ensure that trade secrets are not inadvertently disclosed. Disclosure can occur either passively or actively. **Passive disclosure** might occur by carelessly leaving documents containing trade secrets in open view when business associates from another company visit your facility. **Active disclosure** might occur during a guided tour of the facility when a tour guide reveals the nature of a project or process the company is working on or has recently perfected.[xviii]

The type of active trade secret disclosure that can occur outside of the facility where trade secrets are housed is through the dissemination of confidential information at trade shows, in magazine articles, in publications, in press releases, or during public speeches. Such disclosures often occur unwittingly because they are made by marketing executives or mid-level managers who are not themselves instrumental in the development of the trade secret and who are often charged with a goal somewhat at odds with maintaining the secrecy of intellectual property—that is, their duty often is to publicly disclose the company’s achievements. Similar difficulties might arise when engineers or those responsible for the creation of trade secrets exchange ideas with their colleagues from other companies at professional seminars.[xix] One method of curtailing such disclosures is to insist that all publications, articles, and speeches be reviewed and approved by a “filter” committee prior to dissemination.[xx]

Although the very nature of many businesses requires that a company disclose its trade secrets to prospective buyers, licensees, joint venturers, or other outsiders, this does not mean that a company must surrender all rights to its trade secrets. Rather, when the company enters into a transaction with a third party, it should cautiously monitor the flow of information to the third party and document the nature of the trade secrets exposed and the specific limited use to which they may be put.[xxi] You might consider putting outsiders on notice regarding the need to keep information confidential, and even require the signing of confidentiality agreements. The standard of what constitutes reasonable efforts to maintain secrecy varies substantially depending upon the circumstances and the information that is being protected. However, the different methods of maintaining secrecy discussed above all often help to show that the information in question qualifies as a trade secret, and deserves trade secret protection.

Notes

xi. See *Ruckelshaus v. Monsanto Co.*, 467 U.S. 986, 1002 (1984).
xii. See *Lexis-Nexis v. Beer*, 41 F. Supp.2d 950 (D. Minn. 1999).
xiii. Compare *Electro-Craft Corp. v. Controlled Motion, Inc.*, 332 N.W.2d 890 (Minn. 1983) (confidentiality agreement signed by employees failed to identify with sufficient particularity what were and were not trade secrets; plaintiff began conducting "exit interviews" ten days after litigation commenced) and *Gordon Employment, Inc. v. Jewel*, 356 N.W.2d 738 (Minn. Ct. App. 1984) (confidentiality never discussed with employees).
xiv. Id.
xv. Id.
xvi. See *Gemisys Corp. v. Phoenix American, Inc.*, 186 F.R.D. 551, 558-62 (N.D. Cal. 1999).
xvii. *Electro-Craft Corp. v. Controlled Motion, Inc.*, 332 N.W.2d 890 (Minn. 1983) (no technical documents were marked "confidential").
xviii. Id. (confidentiality procedures inadequate where, among other things, plaintiff allowed public tour of its plants).
xix. See e.g. *Mobile Med. Int'l Corp. v. United States*, 95 Fed. Cl. 706, 739 (Fed. Cl. 2010) (finding that although plaintiff may have taken some measures to maintain the secrecy of its trade information, plaintiff's assertions of confidentiality and the steps taken to protect its information were wholly undermined by its publicly displaying a model unit at a trade show).
xx. *Gallowhur Chem. Corp. v. Schwerdle*, 37 N.J. Super. 385, 117 A.2d 416, 423, 108 U.S.P.Q. 260 (1955) (information communicated in a public lecture revealed trade secret); *Fishing Concepts, Inc. v. Ross*, 1985 WL 1549 (D. Minn. 1985) (information claimed secret was revealed in article published in advertising magazine).
xxi. Id.

Note

Changes to the original OER material:

- Accessibility audit called out link issues with roman numeral footnotes. Forked the page in an attempt to correct the issues.
- Created a numbered list using lower case roman numerals to match the needed numbering. Transferred content from the original "Footnotes" to "Notes." Deleted "original "Footnotes" section.

This page titled 4.7.2.4: The Secrecy Requirement is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

4.7.2.5: Misappropriation of Trade Secrets

Learning Objectives

After completing this section, you will be able to

- Describe the differences between the misappropriation of a trade secret through improper means and disclosure without consent.
- Explain how trade secret owners use direct evidence and circumstantial evidence to prove that a misappropriation of trade secrets has occurred.

The enforcement provisions of the UTSA allow companies and individuals to protect information that qualifies as a trade secret. The protections allowed for by the UTSA are triggered by the actual or threatened "misappropriation" of trade secrets. "Misappropriation" is defined in two ways:

1. acquisition of a trade secret through improper means, or
2. disclosure or use of a trade secret without consent.

Misappropriation through Improper Means

The first type of misappropriation under the UTSA is acquisition of a trade secret through improper means. Specifically, the UTSA prohibits the "acquisition of a trade secret of another, by a person who knows or has reason to know the trade secret was acquired by improper means."[xxii] "Improper means" includes theft, bribery, misrepresentation, breach or inducement of a breach of a duty to maintain secrecy, or espionage through electronic or other means.[xxiii] In short, the first type of misappropriation is the mere possession of a trade secret that has been acquired through theft or the breach of a duty to maintain the secrecy of the trade secret.

Misappropriation through Disclosure

The second type of misappropriation under the UTSA is the disclosure or use of someone else's trade secret without their express or implied consent. Although the statutory definition is somewhat complicated, the UTSA essentially prohibits the disclosure or use of a trade secret that is possessed as a result of theft, breach of a duty to maintain its secrecy, or an accident or mistake on the part of the trade secret owner.[xxiv]

In short, the UTSA prohibits the disclosure or use of a trade secret that is possessed as a result of "improper means" as defined in the preceding paragraph. The rule of thumb is that if you possess, disclose, or use a trade secret that you know or have reason to know is not intended to be revealed to you, beware.

UTSA's Definition of Misappropriation

The UTSA's definition of misappropriation is rather broad. The protection granted to trade secret owners under the UTSA is further expanded by prohibiting not only the actual, but also the threatened, misappropriation of trade secrets. The distinction between these two types of misappropriation is basically one of timing. The UTSA prohibits the actual disclosure of trade secrets and provides remedies for those whose trade secrets have already been misappropriated. A more difficult question arises, however, with respect to a trade secret owner's rights prior to the actual misappropriation of trade secrets. To address a trade secret owner's legitimate concerns that misappropriation may occur in the future, the UTSA prohibits "threatened" misappropriation.

A trade secret owner may assert legal rights subsequent to any actual unauthorized disclosure or use of trade secrets. In such situations, the wrong has already occurred and the trade secret owner may seek judicial intervention concerning any legal rights that were violated. This is true regardless of whether the misappropriator obtained the secret without the owner's consent, or

originally with consent but later used the secret for the misappropriator's own gain.

The burden of proof in an actual misappropriation case is still squarely on the shoulders of the trade secret owner who is alleging that trade secrets have been misappropriated. This burden of proof requires the trade secret owner to demonstrate that he or she was in possession of knowledge or information that was not generally known and was valuable precisely because it was not generally known, and that the knowledge or information was subject to efforts that were reasonable under the circumstances to maintain its secrecy. Depending on the facts and circumstances of a given case, together with the nature of the intellectual property to be protected, proving the elements necessary to make out a trade secret case can be a challenging undertaking. The best way to establish a foundation for successful prosecution of trade secret misappropriation is to follow closely the provisions of a written trade secret protection plan, and take measures to ensure the security of the information as discussed above.

Perhaps the most interesting trade secret cases involve the alleged threatened misappropriation of trade secrets. In these cases, the trade secret owner does not contend that their trade secrets have already been compromised, but rather that unless some action is taken to halt the current progress of events, their trade secrets will be misappropriated at some time in the future. Threatened misappropriation cases often arise when a highly skilled or trained employee who has been exposed to a company's trade secrets leaves the company's employ and goes to work for a competitor. On the one hand, the former employer may well have legitimate concerns about possible disclosure of its trade secrets to a competitor; on the other hand, the employee has a right to choose an employer. Consequently, courts are required to balance these competing interests. The result often depends on the legitimacy of the trade secret owner's concerns of misappropriation and the nature of the evidence that is available to substantiate those concerns.

When faced with a situation where an employee who has been exposed to trade secrets leaves to work for a new employer, courts typically consider three factors:

1. whether the former employer and the new employer are competitors,
2. whether the employee's new position is comparable to the former position, and
3. the efficacy of steps taken by the new employer to prevent the misappropriation of trade secrets.[xxv]

Mere suspicion that another party may have wrongfully disclosed or used one's trade secrets, in and of itself, does not enable the trade secret owner to assert these rights successfully. Courts will not prohibit the wrongful use or disclosure of trade secrets based upon the mere uneasiness or suspicions of the trade secret owner that some misappropriation may occur. Rather, the trade secret owner is required to demonstrate that the concerns of misappropriation are legitimate. In most situations, this means that the trade secret owner will be required to present evidence, either direct or circumstantial, that misappropriation is likely to occur.

A trade secret owner may present direct or circumstantial evidence to demonstrate that there is an imminent threatened misappropriation of trade secrets. Direct evidence exists when the trade secret owner can demonstrate that action has already been taken, or words have already been spoken, that indicate that misappropriation of trade secrets is going to occur. Common examples of direct evidence supporting a trade secret owner's legitimate concerns of threatened misappropriation are statements made by the former employee or the new employer. Former employees might make statements about the relative lack of weight they give to their former employer's confidentiality policies. Similarly, the new employer might make statements about the purpose of hiring the new employee, why other potential employees were rejected, or the nature of the duties the new employee will be required to perform. Based on evidence of such actions or statements, a court is likely to protect trade secrets by enjoining prospectively their wrongful use or disclosure.

In the Courtroom

However, in most situations, direct evidence concerning the likelihood of misappropriation does not exist. In such cases, a trade secret owner can present circumstantial evidence to demonstrate that the unauthorized use or disclosure of trade secrets is imminently threatened. Circumstantial evidence demonstrating that imminent misappropriation is likely to occur consists of two types: (1) evidence demonstrating the intent of a party to elicit the wrongful disclosure of the trade secret, and (2) evidence demonstrating that the wrongful disclosure or use of a trade secret is inevitable regardless of any party's intentions.

Evidence demonstrating the intent of the new employer to elicit trade secrets from its new employee takes many forms. Often, a trade secret owner can show that the new employer desires the trade secret and has unsuccessfully attempted to acquire the trade secret information by proper means, either from the trade secret owner, other parties, or independent derivation or reverse engineering.[xxvi] The new employer's intent to misappropriate trade secrets can also be demonstrated by showing that the new employer solicited the trade secret owner's employees and has not made any efforts to solicit employees from the industry in general. Other facts, such as access to a common supplier, access to a former consultant, or failed prior attempts to duplicate

technology can also contribute to a circumstantial showing of intent.

Second, in addition to circumstantial evidence demonstrating an intention to elicit improper disclosure of trade secrets, a trade secret owner can show that such disclosure is inevitable by virtue of the particular circumstances of the situation. This showing is most frequently demonstrated by evidence that the work to be performed by the former employee is very similar to the work previously performed for the trade secret owner, and that the former employee will be unable to perform the new job conscientiously without revealing trade secrets previously learned.

Although circumstantial evidence is generally not considered as strong as direct evidence, it is clear that courts will issue injunctions prohibiting the misappropriation of trade secrets based solely on circumstantial evidence that indicates such activity is imminently likely to occur. For example, in *PepsiCo, Inc. v. Redmond* , 54 F.3d 1262, the defendant moved from a senior position at PepsiCo to a position at rival Quaker Oats Company (then owner of the Gatorade brand, which is now owned by Pepsi). The court held that it was proper for the district court to conclude that based on the defendant's access to Pepsi's trade secrets, it was inevitable as part of the defendant's new employment that the trade secrets would be relied on. The district court concluded that unless the defendant possessed "an uncanny ability to compartmentalize information," he would necessarily make decisions for the new employer by relying on the trade secrets. This was true even though there was no direct evidence that Redmond disclosed any of Pepsi's trade secrets.

Act Fast

A trade secret owner who suspects that either actual or threatened misappropriation has occurred or is likely to occur must act promptly to enforce their legal rights. Although the statute of limitations under the UTSA is three years, courts have the power to deny the relief requested by trade secret owners even when an action has been initiated within the statute of limitations period, if the action has not been initiated reasonably promptly. The UTSA assumes that a trade secret owner who suspects misappropriation will move with reasonable alacrity to protect intellectual property, and recognizes that if such action is not taken, then third parties may materially change their positions in a good-faith belief that no trade secret rights have been violated. Under such circumstances, a court may find that the trade secret owner has been estopped from exercising, or waived their right to seek protection under the UTSA. These principles essentially allow the court to find that even though the information that a party seeks to protect meets the requirements of a trade secret, equity demands that the information not be protected as a trade secret because protecting it would harm a third party.[xxvii]

Notes

xii. See Unif. Trade Secrets Act § 1(2)(i), U.L.A (1985).
iii. See Unif. Trade Secrets Act § 1(1), U.L.A (1985).
iv. See Unif. Trade Secrets Act §§ 2(2)(ii)(A)-(C), U.L.A (1985).
xv. *Surgidev Corp. v. Eye Technology Inc.*, 648 F. Supp. 661, 695 (D. Minn. 1986), aff'd, 828 F.2d 452 (8th Cir. 1987).
vi. For example, in *Convolve, Inc. v. Compaq Computer Corp.*, 2006 U.S. Dist. LEXIS 13848, the plaintiffs asserted that the defendants were misappropriating trade secrets in violation of a nondisclosure agreement. As evidence of this violation, the plaintiffs showed that the defendants tried unsuccessfully to reverse engineer the trade secret-protected software. That evidence, interalia, was sufficient to withstand a motion for summary judgment.
vii. See *Watermark Solid Surface, Inc. v. Sta-Care, Inc.*, 2009 WL 723199 (W.D. Wis. 2009) (discussing waiver of trade secret protection); Christianson v. Colt Industries Operating Corp., 766 F.Supp. 670 (C.D. Ill. 1991) (outlining common equitable defenses in trade secrets cases).

Note

Changes to the original OER material:

- Accessibility audit called out link issues with roman numeral footnotes. Forked the page in an attempt to correct the issues.
- Created a numbered list using lower case roman numerals to match the needed numbering. Transferred content from the original "Footnotes" to "Notes." Deleted "original "Footnotes" section.
- Converted an item in the first paragraph to a numbered list. Converted another to a numbered list.

This page titled 4.7.2.5: Misappropriation of Trade Secrets is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

4.7.2.6: Remedies Available for the Misappropriation of Trade Secrets

Learning Objectives

After completing this section, you will be able to

- Describe how injunctions are used to enjoin actual and threatened misappropriations of trade secrets.
- Explain how remedies are awarded under the Uniform Trade Secrets Act.

The UTSA provides a full array of remedies, including **statutory damages**, **injunctive relief**, **attorneys' fees**, and **exemplary damages**.[xxviii] The remedy or remedies will depend on the extent of trade secrets involved, the manner of misappropriation or conversion, and the type of damage sustained by the owner.

Injunctive Relief

As discussed above, under the UTSA, injunctive relief is available to enjoin actual, as well as threatened, misappropriation. Injunctive relief allows the owner of a trade secret to obtain a court order prohibiting the actual or threatened misappropriation of a trade secret. The ability to enjoin threatened misappropriation is very important because it enables the owner of the trade secret to prevent public disclosure of the trade secret.[xxix] Injunctive relief may take many forms. The scope of an injunction may vary from a broad prohibition on any use of material found to be a trade secret, to a more limited prohibition on the scope of an employee's work with a new employer.[xxx] Both the scope and subject matter of an injunction vary depending upon the specific facts of the case and the nature of the trade secret in question. The UTSA provides that an injunction shall terminate once the trade secret ceases to exist, but may be continued for a reasonable period of time to extinguish any commercial advantage that otherwise would be derived from the misappropriation. This is especially important where, at the time of the action, the information or knowledge is no longer a trade secret. In such a case, the court may still award the owner injunctive relief that is narrowly crafted to deprive the defendant of the benefit of time saved by improper use of the owner's trade secret.[xxxi]

Table 4.7.2.6.1*: Injunctive Relief*

Definition	Injunctive relief allows the owner of a trade secret to legally prohibit a former employee, for example, from misappropriating, publicly disclosing, or otherwise profiting from knowledge of the trade secret.
Scope	The scope of an injunction varies based on the on the nature of the potential misappropriator's knowledge of the trade secret, and on the potential competitive advantage that knowledge might grant.
Duration and Termination	The duration of an injunction is determined based on the amount of time it would take a competitor to reveal the trade secret by proper means. An injunction can be terminated when the trade secret ceases to exist.
Permanent Injunctions	To determine the appropriate length of time, courts may rely on the opinions/experience of experts and competitors. They may impose permanent injunctions or rely on other considerations seen as important by the court.

Where the information has not been publicly disclosed, the issue becomes more difficult. The duration of the injunction is often calculated according to the time that would be required for independent development of the information or knowledge that was misappropriated. These "head start" injunctions assume that the information could be properly obtained through reverse engineering or independent development.[xxxii]

To determine what length of time is appropriate, courts may rely on the opinions of experts and the experience of other competitors.[xxxiii] However, a court may also opt to impose a permanent injunction that can later be modified, or in some circumstance rely on other considerations deemed important by the court.[xxxiv]

Courts also exhibit broad flexibility in the nature and duration of relief in cases dealing with former employees. When an injunction is issued in these cases, it usually will not be an outright ban on employment. Rather, it will be a more limited injunction for a more limited duration.[xxxv] Most courts are willing to allow continued employment, but will restrict the employee from working with a particular product line or area of the new employer's business.[xxxvi] However, if the role the employee will assume would inevitably lead to the disclosure of trade secrets, the employment may be prohibited altogether.[xxxvii]

Damages Awarded

Damages are often awarded in addition to an injunction and are not an exclusive remedy.[xxxviii] For example, in *Seagate Tech., LLC v. Western Digital Corp.* , 834 N.W.2d 555 (Minn. App. 2013), the Minnesota Court of Appeals affirmed a judgment of $630 million in favor of the plaintiffs against the defendants stemming from a trade secret misappropriation dispute. Under the UTSA, damages are measured by the actual loss caused by the misappropriation as well as any resulting unjust enrichment of the misappropriator. Where the amount of actual loss or unjust enrichment cannot be accurately calculated, the UTSA provides that damages may also be measured by the imposition of a reasonable royalty for the misappropriator's unauthorized disclosure or use of the trade secret. Exemplary or punitive damages may be awarded where the court finds that the misappropriation was willful and malicious.[xxxix] For example, in *Mattel, Inc. v. MGA Entm't, Inc.*, 801 F. Supp. 2d 950 (C.D. Cal 2011), Mattel was found to have engaged in a prolonged practice of encouraging employees to use false pretenses to acquire competitive information from the defendant. As a result of this conduct, the jury awarded, and the court affirmed a verdict of $85 million dollars in exemplary damages. A punitive or exemplary award may not exceed twice the amount of the damage award.[xl] However, some states have modified this cap on punitive damages, and allow for punitive damages greater than two times the amount of actual damages.[xli]

Table 4.7.2.6.1*: Case Studies*

Case	Description of Damages
Seagate Tech., LLC v. Western Digital Corp.	In this case, the Minnesota Court of Appeals affirmed a judgment of $630 million in favor of the plaintiff stemming from a trade secret misappropriation. Under the UTSA, damages are measured by the actual loss caused by the misappropriation and any resulting unjust enrichment of the misappropriator.
Mattel, Inc. v. MGA Entm't. Inc.	In this case, Mattel was found to have encouraged employees to use false pretenses to acquire competitive information from the defendant. As a result of this conduct, $85 million dollars in exemplary damages was awarded. A punitive or exemplary award may not exceed twice the amount of the damage award.[xlii] However, some states have modified this cap.[xliii]

In exceptional circumstances, an injunction may entitle the owner of the trade secret to a royalty for the period of time during which the use of the trade secret was prohibited. For example, the Supreme Court of Georgia has affirmed a trial court's reliance on a seldom-used provision of the UTSA to impose a "royalty injunction" instead of an injunction prohibiting the use of the plaintiff's trade secret.[xliv]

A court may also award attorneys' fees to a prevailing party in exceptional circumstances. A bad faith claim of misappropriation, a bad faith resistance to a motion to terminate an injunction, or a finding of willful and malicious appropriation are exceptional circumstances in which an award of attorneys' fees may be appropriate. However, courts are generally disinclined to award attorneys' fees.[xlv] The attorney fee provision is included to dissuade parties from filing malicious misappropriation actions in bad faith, and considered along with the need for exemplary damages.[xlvi]

Notes

iii. See Unif. Trade Secrets Act §§ 2-5, U.L.A (1985).
ix. *Standard Brands, Inc. v. Zumpe*, 264 F. Supp. 254, 269 (E.D. La. 1967).
xx. See *3M v. Pribyl*, 259 F.3d 587, 598 (7th Cir 2001) (upholding broad injunction on use of material in employee handbook, as overly specific injunctions may invite evasion).
xi. See *Cherne Indus., Inc. v. Grounds & Associates*, 278 N.W.2d 81 (Minn. 1979)
xii. See, e.g., *Surgidev Corp. v. Eye Technology, Inc.*, 828 F.2d 452, 456-457 (8th Cir. 1987) (affirming 15-month injunction based

on the period of time that would be required for independent development of the protected information).

iii. See, Restatement (Third) of Unfair Competition Section 44, comment f.

iv. See e.g., *General Electric Co. v. Sung*, 843 F. Supp. 776, 780-781 (D. Mass. 1994) (seven-year product injunction, taking into account the willfulness of the misappropriation).

xv. See, e.g., *Doubleclick v. Henderson*, No. 116914/97 (N.Y. Sup. Ct. Nov. 5, 1997) (granting an injunction restraining two former employees from competing with their former employer for a period of six months).

vi. See, *Merck & Co. Inc v. Lyon*, 941 F. Supp. 1443 (M.D.N.C. 1996). See also Allis Chalmers Manf. Co. v. Continental Aviation and Eng. Corp., 255 F. Supp. 645 (E.D. Mich. 1966) (injunction limited to research as to a particular type of pump); but see *Union Carbide Corp. v. UGI Corp.*, 731 F.2d 1186 (5th Cir. 1984) (affirming injunction against use of information in 14 broad areas).

vii. *PepsiCo, Inc. v. Redmond*, 54 F.3d 1262, 1272 (7th Cir 1995).

iii. See e.g., *Mikes Train House, Inc. v. Lionel, L.L.C.*, 472 F.3d 398, 413-15 (6th Cir. 2006).

ix. See, e.g., *Aries Information Sys., Inc. v. Pacific Management Sys. Corp.*, 366 N.W.2d 366 (Minn. Ct. App. 1985).

xl. See Unif. Trade Secrets Act § 3(b), U.L.A (1985).

xli. *Reingold v. Swiftships, Inc.*, 210 F.3d 320, 323 (5th Cir. 2000).

lii. *Reingold v. Swiftships*, Inc., 210 F.3d 320, 323 (5th Cir. 2000).

liii. See Unif. Trade Secrets Act § 3(b), U.L.A (1985).

liv. See *Electronic Data Systems Corp. v. Heinemann*, 493 S.E.2d 132 (Ga. Sup. Ct. 1997).

lv. See, e.g., *Gordon Employment, Inc. v. Jewe1*, 356 N.W.2d 738 (Minn. Ct. App. 1984).

lvi. See Unif. Trade Secrets Act § 4 (comment), U.L.A(1985).

Note

Changes to the original OER material:

- Changed Table numbering so that it corresponds to section numbering in this textbook using Libretexts constructs.
- Added Table header designation and re-labeled Table.
- Accessibility audit called out link issues with roman numeral footnotes. Forked the page in an attempt to correct the issues.
- Created a numbered list using lower case roman numerals to match the needed numbering. Transferred content from the original "Footnotes" to "Notes." Deleted "original "Footnotes" section.

This page titled 4.7.2.6: Remedies Available for the Misappropriation of Trade Secrets is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

4.7.2.7: Assessment Questions

1.

Unlike patents and trademarks, the law requires that intellectual property consisting of trade secrets not be publicly disclosed. True or False?

A. True.
B. False.

Answer available at the end of this section.

2.

Which of the following best describes a trade secret?

A. Information that many people do not know about how a product like a smartphone is made.
B. Virtually any information that is of value as a result of not being generally known.
C. Any information relating to the finances of a publicly held company.

Answer available at the end of this section.

3.

Trade secret theft or misappropriation costs the world's richest 40 nations approximately how much of their gross domestic product, or GDP, each year?

A. 1 to 3 percent.
B. 4 to 6 percent.
C. 7 to 10 percent.

Answer available at the end of this section.

4.

What are the two main requirements for information to be protectable as a trade secret?

A. It must be novel and non-obvious.
B. It must be original and expressed in a tangible form that can be seen or copied.
C. It must be not generally known by others, and it must have been subject to reasonable efforts to avoid public disclosure.

Answer available at the end of this section.

5.

What is one key advantage of trade secret protection versus patent protection?

A. It protects the information to be kept secret indefinitely.
B. It prevents others from ever independently deriving the same information.
C. Both of the above are correct.

Answer available at the end of this section.

6.

If you invent faster-than-light travel but believe someone will be able to reverse engineer it by examining your starship, should you patent it or keep it as a trade secret?

A. Keep it as a trade secret.
B. Patent it.

Answer available at the end of this section.

7.

What did the Supreme Court note in the case of *Bonito Boats v. Thunder Craft Boats*?

A. One cannot legally misappropriate trade secrets.
B. Nobody is above the law, not even a trade secret owner.
C. The holder of a trade secret does not have protection against reverse engineering.

Answer available at the end of this section.

8.

If information is protected as a valid trade secret, a competitor can be stopped from using it no matter who he obtained it from. True or False?

A. True.
B. False.

Answer available at the end of this section.

9.

Trade secret protection is guaranteed by the federal government. True or False?

A. True.
B. False.

Answer available at the end of this section.

10.

In which country did trade secret law first originate in the early 1800s?

A. France.
B. The United States.
C. England.

Answer available at the end of this section.

11.

What impact did the Uniform Trade Secrets Act (UTSA) of 1979 have on trade secret protection in the United States?

A. It was the first trade secret law binding on every state.
B. It served as a model that was adopted by 48 of the 50 U.S. states.
C. It served as a model that was adopted by every state.

Answer available at the end of this section.

12.

Which of the following best describes the problem with the current lack of uniform federal protection of trade secrets?

A. There is no real protection against misappropriation of trade secrets.
B. The protections against trade secret theft are very uneven.
C. A federal trade secret law would merely replicate state protections.

Answer available at the end of this section.

13.

Which of the following could be considered a trade secret so long as reasonable steps had been taken to prevent its disclosure?

A. Sales information.
B. Customer lists.
C. Manufacturing techniques.
D. All of the above.

Answer available at the end of this section.

14.

Is information not known to the public considered to be a trade secret?

A. Yes.
B. No.
C. It depends.

Answer available at the end of this section.

15.

Would a customer list always be considered a trade secret?

A. It depends on whether the customers want to be identified or not.
B. To be considered a trade secret, the names of customers must not be available or discernable through public sources.

Answer available at the end of this section.

16.

Can a mere idea be considered a trade secret?

A. No. Just as you cannot patent or copyright a mere idea, you cannot protect a mere idea with a trade secret.
B. Yes. Under the Uniform Trade Secrets Act, even an idea can sometimes be considered a trade secret.

Answer available at the end of this section.

17.

To gain trade secret protection, what must the owner do with confidential information?

A. Publicize it as widely as possible.
B. Not disclose it.
C. Take active steps to keep the information secret.

Answer available at the end of this section.

18.

What are the benefits of developing a written trade secret plan?

A. It serves as a policy manual that prevents situations in which unwritten rules are followed and security procedures are ignored.
B. In the event of litigation, it can serve as proof that you took active steps to keep the information secret.
C. Both of these.

Answer available at the end of this section.

19.

How should companies handle written trade secrets?

A. Shred them.
B. Mark them “secret” or “confidential,” among other measures taken.
C. Never store them on a computer.

Answer available at the end of this section.

20.

Which of these is an example of “passive disclosure” of a trade secret?

A. Leaving confidential documents lying around for anyone to see.
B. Inadvertently disseminating information at trade shows or conferences.
C. Disclosing confidential information in press releases, newspaper articles, or marketing collateral.

Answer available at the end of this section.

21.

What is one way to guard against the active disclosure of trade secret information?

A. Prevent employee attendance at trade shows and seminars.
B. Appoint a trade secret committee to approve publications, speeches, and marketing collateral prior to disclosure.
C. Avoid meeting with prospective buyers, customers, or licensees.

Answer available at the end of this section.

22.

Which of the following is NOT an example of misappropriation of a trade secret?

A. Acquisition of a trade secret through improper means.
B. Deliberate disclosure of a trade secret by the trade secret owner.
C. Disclosure of a trade secret without consent.

Answer available at the end of this section.

23.

The UTSA not only prohibits the actual disclosure or use of a trade secret, it also protects against the "threatened" disclosure of trade secrets. When might a court intervene to stop a "threatened" disclosure before the actual disclosure takes place?

A. An employee goes to work for another company in a different industry.
B. An employee caught embezzling is fired and the employer believes the same lack of morals will lead to the fired employee disclosing trade secrets.
C. An employee goes to work for an arch competitor in a comparable job position.

Answer available at the end of this section.

24.

The burden of proof in a threatened trade secret misappropriation case lies with whom?

A. The trade secret owner.
B. The former employee who leaves to go work for a competitor.
C. The competitor, who must prove he did no wrong by hiring the ex-employee.

Answer available at the end of this section.

25.

What remedies does the UTSA provide for the misappropriation of trade secrets? (Choose all that apply.)

A. Criminal penalties of up to five years for a first offense.
B. Statutory damages.
C. Injunctive relief.
D. Exemplary damages.
E. Attorneys' fees.

Answer available at the end of this section.

26.

What remedies were imposed in the case of *Mattel, Inc. v. MGA Entm't, Inc.* in 2013?

A. An injunction barring a new employer from hiring the former employee of the trade secret owner.
B. $85 million in exemplary (punitive) damages awarded to the trade secret owner.

Answer available at the end of this section.

Answers

1.
A. True

2.
B. Virtually any information that is of value as a result of not being generally known.

3.
A. 1 to 3 percent.

4.
C. It must be not generally known by others, and it must have been subject to reasonable efforts to avoid public disclosure.

5.
A. It protects the information to be kept secret indefinitely.

6.
B. Patent it.

7.
C. The holder of a trade secret does not have protection against reverse engineering.

8.
B. False

9.
B. False

10.
C. England

11.
B. It served as a model that was adopted by 48 of the 50 U.S. states.

12.
B. The protections against trade secret theft are very uneven.

13.
D. All of the above.

14.
C. It depends.

15.
B. To be considered a trade secret, the names of customers must not be available or discernable through public sources.

16.
A. No. Just as you cannot patent or copyright a mere idea, you cannot protect a mere idea with a trade secret.

17.
C. Take active steps to keep the information secret.

18.
C. Both of these.

19.
B. Mark them "secret" or "confidential," among other measures taken.

20.
A. Leaving confidential documents lying around for anyone to see.

21.
B. Appoint a trade secret committee to approve publications, speeches, and marketing collateral prior to disclosure.

22.

B. Deliberate disclosure of a trade secret by the trade secret owner.

23.

C. An employee goes to work for an arch competitor in a comparable job position.

24.

A. The trade secret owner.

25.

B. Statutory damages.

26.

B. $85 million in exemplary (punitive) damages awarded to the trade secret owner.

This page titled 4.7.2.7: Assessment Questions is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

4.7.3: Trademark Basics

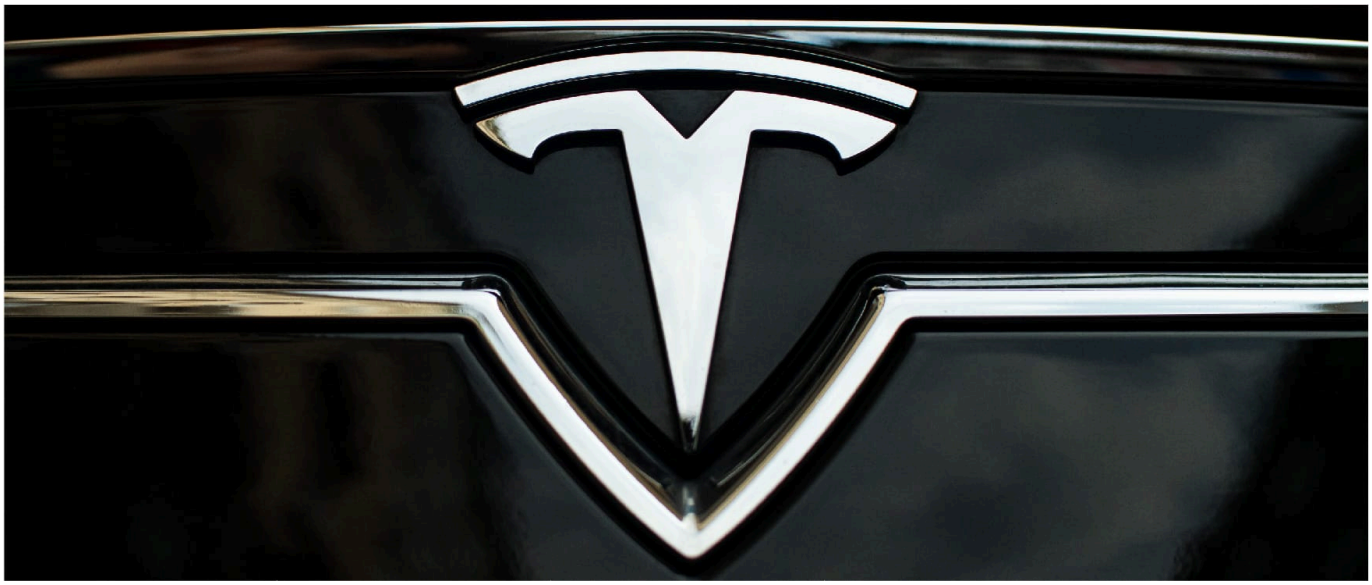


Figure 4.7.3.1: Close-up of a chrome Tesla logo on the front of a Tesla car. (credit: modification of work “Close-up of TESLA trademark sign on a car” by Ivan Radic/flikr.com, CC BY 2.0)

Note

Changes made:

- Converted Figure to use Libretexts template to get the figure number to be correct relative to this textbook's section numbering.
- Updated to reflect recent landmark Supreme Court cases.

This page titled 4.7.3: Trademark Basics is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

4.7.3.1: Core Concepts

Learning Objectives

After completing this section, you will be able to:

- Understand trademarks and their properties.
- Identify differences between trademarks and other intellectual property rights.

A **trademark** is an intellectual property right granted by a government to an individual, business, or legal entity that creates and uses a distinctive word, name, symbol, or device to distinguish its products or services from those from any other entity in the marketplace.

The original purpose of a trademark was to indicate the origin of goods and services. Trademarks thus protected the public by preventing mistakes, confusion, or deception by those who would "palm off" their goods as those of another. But it also served to protect the market and reputation (or goodwill) of the producers of goods. As modern markets evolved, trademarks also developed into guarantees of quality as well as potent marketing and advertising devices.

Trademarks as Branding

Consider, for example, the role that Nike's "Swoosh" logo plays in its $106 billion shoe, equipment, and apparel business. In 1971, Nike founder Phil Knight paid graphic design student Carolyn Davidson a mere $35 to design the "Swoosh" logo for the fledgling new company. According to the Portland *Oregonian* newspaper, when Knight saw her design, he reportedly told her, "I don't love it, but maybe it will grow on me." Nike attorneys nonetheless registered the logo with the U.S. Patent and Trademark Office (USPTO) on June 18, 1971.[i]

Today, this one logo is estimated to be worth as much as $20 billion, and is recognized around the world as a symbol of Nike's quality workmanship and design. Indeed, its vital role in protecting Nike's market share and reputation explains why the company so strenuously protects its trademark rights from being infringed by counterfeiters. As for design student Carolyn Davidson, Phil Knight gave her Nike stock in 1983 that is today worth more than $850,000.

Trademarks vs. Other Intellectual Property Rights

Trademarks share with other intellectual property rights the power to encourage and reward creative enterprise. Trademarks also share with patent rights and copyrights the public policy goal of marshaling the benefits of creative endeavor—in this case, the distinctive branding of one's products and services from those of others—to the public good. They do this by protecting the consumer from deception and encouraging sellers to provide quality products.

But trademarks are different from other intellectual property rights in three key respects. In the first place, the legal foundation for U.S. trademark law comes not from rights expressly enumerated in the Constitution, as is the case with patent rights and copyrights. Rather, it lies in the Commerce Clause of the Constitution, which gives Congress the authority to regulate interstate commerce and enact whatever necessary and proper legislation is required to do that.

Trademarks are also different from other intellectual property rights in that they are not limited in duration. Patents and copyright are granted only for limited periods of time because society benefits by putting an invention or literary work into the public domain once the inventor or artistic creator has recouped the costs of innovation and been rewarded for the pioneering endeavor. Trademarks, however, never hinder the sales of other products or services, so they are granted in perpetuity so long as they are not abandoned by the trademark owner.

Trademarks Mean Business

Finally, trademarks exist only in conjunction with commercial activity. An inventor may receive a patent for a new invention and

never employ or “practice” that invention in a business or research endeavor. Similarly, an author can receive a copyright for an original literary or artistic work and yet never publish, display, or sell it. A trademark, however, cannot exist by itself, apart from commercial activity.

Thus a trademark cannot be obtained by mere adoption. It can only be acquired through commercial use or in anticipation thereof—i.e., through the sale of goods and services.

Footnotes

i. Allen Brettman, “*Creator of Nike’s Famed Swoosh Remembers Its Conception 40 Years Later*,” the Oregonian, June 15, 2011.

Note

Changes made:

- Corrected typo (inserted space): "*Oregonian* newspaper"
- Added colon in Learning Objectives
- Changed Footnotes to a numbered list using lower Roman numerals.

This page titled 4.7.3.1: Core Concepts is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

4.7.3.2: Early Trademark Systems

Learning Objectives

After completing this section, you will be able to

- Discuss the origins of trademarks.
- Analyze examples of early forms of trademarks.

Bronze Age Origins

Scholars of antiquity give Early Bronze Age potters the credit for creating the world's first trademarks by imprinting their works with distinctive markings. We assume that these marks were meant to indicate the origin of a particular work or the identity of its craftsman, but the historical record is not conclusive on this point. But an examination of the potters' seals found on Corinthian artifacts dating to 2000 BCE suggests this is a reasonable supposition. Trade and commerce between the tribes and early civilizations of that era were expanding rapidly, and in order to be mutually beneficial, trade requires a certain amount of trust in the provenance and quality of goods. These early potters' trademarks seemed to have served that purpose by distinguishing the goods of quality craftsmen from those of unknown or uncertain sources.

Trademarks have always been inextricably bound up with commerce. And as commerce grew and developed over the centuries, so did the use of trademarks. Marks have been found, for example, on works ranging from Egyptian pots to the swords of Roman blacksmiths. But it was in the medieval period, with the emergence of powerful craft guilds, that trademark usage really expanded. Their marks identified a work as being made by a particular guild or member of that guild, and therefore continued the long tradition of identifying the origin of goods. But medieval trademarks also served other functions. They became a means by which guilds could control the quality of work of fellow guild members and, because of a trademark's association with quality, ultimately, a source of competitive advantage in the market.[ii] Trademarks thus began to acquire something akin to the existential "moral rights" found in later copyright statutes—i.e., the right of a creator to defend the quality, originality, and "personality" of their work.

Trademark use in medieval times also acquired a public interest function. As Arthur R. Miller and Michael H. Davis note in *Intellectual Property: Patents, Trademarks, and Copyright in a Nutshell*, "Statutes dating back as early as the thirteenth century show that [trademark] was eventually recognized as having social consequence. These statutes were meant to protect the public by preventing the sale of unidentified goods whose quality could not be ascertained."

One notable trademark from the time was that of Löwenbräu Brewery, which claims to have used its lion mark since its founding in 1383.[iii]

Early Trademark Cases

Scholars had until only recently attributed the first reported trademark case in Anglo-American law to be *Southern v. How*, which was decided in 1618. This despite the fact that the case involved not a trademark but the sale of counterfeit jewels.[iv] Its connection to trademark came from a reference by the presiding judge to an earlier, unnamed, and unreported case in 1584 involving a suit brought by a cloth maker against another cloth maker who had used his mark. That earlier clothier case, only recently discovered and now known as *Sanforth's Case*, is now held to be the earliest reported trademark case in Anglo- American law. It establishes beyond a doubt that even 250 years before the Industrial Revolution, trademark infringement was viewed as a tort of deceit and a violation of the laws against unfair competition.[v]

To quote from a contemporary 1656 legal report:

> "The action upon the case was brought in the [Court of] Common Pleas by a clothier, that whereupon he had gained great reputation for his making of his cloth ... to his benefit and profit, and that he used to set his mark on his cloth whereby it should be known to be his cloth; and another clothier, observing it, used the same mark to his ill-made cloth [in order] to deceive him."

In the more than four centuries since then, as commerce and industry have evolved into an $80 *trillion* a year global marketplace unimaginable to the people of *Sanforth's Case* time, its verdict that trademark infringement is competition most unfair and demands redress still serves as the foundation and wellspring of worldwide trademark policies and statutes today.

Footnotes

ii. Arthur R. Miller and Michael H. Davis,*Intellectual Property: Patents, Trademarks, and Copyright in a Nutshell*. (5th ed., p. 25). St. Paul MN: West Publishing Co., 2007.
iii. Gary Richardson, *Brand Names Before the Industrial Revolution*, National Bureau of Economic Research, NBER Working Paper No. 13930, April,2008.
iv. McKenna, M. *The Normative Foundations of Trademark Law*(December 30, 2010). Notre Dame Law Review, Vol. 82, No. 5, p. 1839, 2007. Retrieved from SSRN: SSRN link.
v. Keith M. Stolte, *How Early Did Anglo-American Trademark Law Begin? An Answer to Schechter's Conundrum*, Fordham Intellectual Property, Media and Entertainment Law Journal, Volume 8, Issue 2, 1997.

Note

Changes made to original OER material:

- Accessibility changes to Footnotes.

This page titled 4.7.3.2: Early Trademark Systems is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

4.7.3.3: U.S. Trademark Law

Learning Objectives

After completing this section, you will be able to

- Grasp the basics of trademark law.
- Analyze notable trademark cases.

As the young United States rapidly developed its commerce and industry in the late eighteenth and early nineteenth centuries, propelled in no small part by the nation's unusually democratic and effective patent laws, the states developed increasingly sophisticated and complex trademark laws. Eventually, however, the transformation of America's many local and regional markets into a single, unified, national economy propelled Congress to try to federalize trademark law despite the lack of any express constitutional authority to do so. So Congress passed the first national trademark laws in 1870 and 1876. But in a set of three court challenges consolidated into a single appeal before the U.S. Supreme Court in 1879, Justice Samuel Freeman Miller ruled for the majority that the Patent and Copyright Clause of the Constitution gave Congress no explicit authority to regulate trademarks, and declared the 1870 and 1876 trademark laws unconstitutional.[vi]

The Commerce Clause and the Lanham Act of 1946

Two years later, Congress instead acted under the Commerce Clause of the Constitution to pass the Trade Mark Act of 1881. This first national trademark law, however, only regulated trademarks used in commerce with other nations and with Indian tribes. It wasn't until 1905 that Congress passed another trademark statute regulating the use of marks within the United States.

Those two trademark laws were subsequently overhauled by the Lanham Act of 1946, which remains the principal law of the land on trademarks in the United States to this day.[vii] It broadened national registration of trademarks, and also gave owners of unregistered marks access for the first time to the federal courts. The act also established remedies such as bars on the sale or importation of infringing products, and required the renewal of trademarks every ten years to weed out the registry trademarks that are no longer in commercial use (known as the "deadwood" prevision).[viii]

After 119 years in business, for example, the world-famous retailer Woolworth's went bankrupt in 1997 and its trademark was no longer protected. The 83-year-old Hostess Twinkies brand, however, was acquired by two private equity firms for $410 million because they see future profit in the iconic brand—and the customers who love it.

Requirements for a U.S. Trademark

The requirement that a trademark must be actively used or intended to be used in commercial activity in order to be protected by law has been a fundamental feature of U.S. trademark law since its inception. The Lanham Act specifically requires applicants to submit "a verified statement that the mark is in use in commerce, specifying the date of the applicant's first use of the mark in commerce and those goods or services specified in the notice of allowance on or in connection with which the mark is used in commerce."[ix] A trademark is presumed to have been abandoned when its owner ceases to use it for three or more years, after which anyone else can register and use the mark.

Footnotes

vi. See *United States v. Steffens* 100 U.S. 82 (1879); *United States v. Wittemann*100 U.S. 82 (1879); *United States v. Johnson* 100 U.S. 82 (1879) Retrieved From Caselaw link.
vii. Harvard University. (2012). *Overview of Trademark Law*. Retrieved from Harvard cyber law.
viii. Op. cit., Miller and Davis.
ix. 15 U.S.C. §1051. Retrieved from Cornell Law School's Legal Information Institute.

Note

Changes to original OER material:

- Accessibility fixes to Footnotes.
- Removed reference to OER Chapter 1, “Section 1.5: What the U.S. Patent System Wrought,” as this is not used here and would be confusing with Chapter 1 of this book.

This page titled 4.7.3.3: U.S. Trademark Law is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

4.7.3.4: Four Types of Trademarks

Learning Objectives

After completing this section, you will be able to

- Identify the four types of trademarks.
- Compare differences between types of trademarks.

Can I Trademark That?

The term "trademark" is usually used to describe any of the four types of marks that can be registered with the U.S. Patent and Trademark Office. The two primary types of marks are **trademarks** and **service marks**. The two other marks—**certification marks** and **collective marks**—occur much less frequently and must meet different requirements for registration.[x]

Trademarks

Trademarks identify *products*—i.e., physical goods and commodities—that are either manufactured, produced, grown, or that exist naturally. A trademark is a word, name, symbol, or device—or combination of these—used to identify and distinguish the source of that product. Examples of trademarks include the Nike "Swoosh" symbol, the arched "M" for McDonalds, and the apple symbol with a small curved bite taken out of it for Apple Computer.

Service Marks

Service marks are exactly the same in principle as trademarks except that these words, names, symbols, or devices identify and distinguish the source of a *service*. Examples of service marks include the sleek silver greyhound dog on Greyhound buses, and United Parcel Service's brown shield emblazoned with the bold yellow letters "UPS."

Certification Marks

A certification mark is any word, phrase, symbol, or design—or a combination of any of these—owned by one party that certifies the goods and services of others when they meet certain standards or requirements.[xi] A certification mark identifies either the nature of a product or service—for example, that it meets the quality standards needed to receive the "Good Housekeeping Seal of Approval"—or the **origin** of products or services, as in the certification mark "Washington State" given to apples grown in that state.

Collective Marks

Collective marks come in one of two varieties: collective trademarks and service marks, and collective membership marks. A collective trade or service mark is any word, phrase, symbol, or design that is owned by a cooperative, association, collective group, or organization and is used by its members to indicate the source of goods or services.[xii] An example of a collective trademark is the "Girl Scouts" mark seen on cookies every February, or the designation "CPA" to identify the services provided by a Certified Public Accountant.

A collective membership mark, by contrast, is used to indicate that a person is a member of some organization, such as a trade union or an association like the Rotary Club, but is not used to identify the source of goods and services.

Footnotes

x. United States Patent and Trademark Office. (2012, May 24). Trademarks. Retrieved from USPTO Trademarks.
xi. United States Patent and Trademark Office. (2012, March 09). *What Is a Collective Membership Mark*?. Retrieved from USPTO Trademarks FAQ.
xii. United States Patent and Trademark Office. (2012, March 09). Frequently Asked Questions About Trademarks. Retrieved from USPTO Trademarks FAQ.

Note

Changes made:

- Removed text about watching video.
- Inserted space in "and**collective".**
- Inserted space in "the**origin".**

This page titled 4.7.3.4: Four Types of Trademarks is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

4.7.3.5: Subject Matter of Trademarks

Learning Objectives

After completing this section, you will be able to:

- Learn the subject matter of trademarks.
- Analyze characteristics of various forms of trademarks.

As discussed, there are four types of trademarks: trademarks, service marks, certification marks, and collective marks. But what does it mean to say that a trademark may be a "word [or phrase], name, symbol [or design], or device." The answers may in some cases surprise you.

Words, Phrases, and Names

There is little confusion about what is meant by a trademarkable *word*, *phrase*, or *name*. The only point to remember here is that words, names, and phrases that are simply descriptive of the goods or services with which they are associated do not necessarily qualify for registration as a trademark.[xiii] They must be distinctive and indicate the origin of a product or service. What is meant by distinctive—and how distinctive they must be—will be discussed in the next section.

Symbols and Devices

Symbols and **devices** are where things start to get interesting in trademark law. The language of the Lanham Act does not specify what is meant by a "symbol" or a "device" that can be trademarked. The U.S. Supreme Court took careful note of that fact, ruling in 1995 that, "Since human beings might use as a 'symbol' or 'device' almost anything at all that is capable of carrying meaning, this language, read literally, is not restrictive."[xiv]

In the United States, therefore, trademarks can include almost anything that carries distinctive meaning and identifies the origin of products and services. This includes slogans, letters, numbers, logos, three-dimensional designs—even colors, scents, and sounds that indicate the source of a good or service to consumers.[xv]

A trademarkable symbol can be a *number*. The number 5, for example, is a trademarked symbol of Chanel No. 5 perfume (No. 73788555). So is the number 31, which is the trademarked symbol of Baskin-Robbins 31 Flavors (No. 72172718).

No trademark can block the use of the numbers 5 or 31 in math. Nor have the courts allowed the trademarking of mere part numbers,[xvi] model numbers,[xvii] or grades,[xviii] because these are not distinctive enough and do not indicate the origin of the goods in question. Regular gas, also known as "87 octane," also cannot be trademarked because it doesn't tell you whether that gas comes from Exxon, Shell, or BP.

Yet despite the above, one of the most iconic trademarked numbers is 501. Originally, the number 501 was simply the lot number Levi's assigned to the famous copper-riveted waist overalls. But over time, 501 acquired meaning in the minds of consumers and came to be very strongly associated with *Levi's* brand jeans. How otherwise un-trademarkable symbols (or words and designs) can acquire secondary (trademark) meaning will be discussed in the next section.

As for *scents*, the fundamental case for trademarking these came about when the USPTO denied Celia Clarke's request in 1990 for a trademark for scented yarns and threads with a "fresh, floral fragrance reminiscent of Plumeria blossoms." She appealed, and the Trademark Trial and Appeal Board overruled the examiner, granting her a trademark for the scent.

"[F]ragrance is not an inherent attribute or natural characteristic of applicant's goods but is rather a feature supplied by applicant," the board noted. "Moreover, applicant has emphasized this characteristic of her goods in advertising, promoting the scented feature of her goods. Applicant has demonstrated that customers, dealers and distributors of her scented yarns and threads have come to recognize applicant as the source of these goods. . . . In her advertisements and at craft fairs, applicant has promoted

her products as having a scented nature. We believe that applicant has presented a prima facie case of distinctiveness of her fragrance mark."[xix]

On the other hand, a trademark cannot be granted for any scent that serves a function other than identifying the product's source. The scent of perfumes and air fresheners, therefore, cannot be registered, nor can the sulfurous smell that serves as a warning for natural gas leaks.

Still, there remains a good deal of uncertainty about when a scent is trademarkable and when it is not. Consider the world-famous Cinnabon smell—that sweet, sugary, cinnamon-infused scent that is absolutely unmistakable (if not also irresistible). But it is not trademarked because even though it is certainly distinctive, it cannot be easily distinguished from the smell of cinnamon buns made by any number of other bakeries.

Sounds can also be trademarked provided they indicate the source of the product or service with which they are associated, and indeed, there are approximately 700 trademarked sounds registered at the USPTO.[xx] These include Tarzan's Yell, the THX theme heard at the beginning of a movie, MGM's roaring lion, America Online's "You've got mail" announcement, the sequence of chimes heard with the display of the NBC logo on TV, and, naturally, the sound of a duck quacking "Aflac!" for the American Family Life Assurance Company.

Designs and Trade Dress

A *design*, like a logo, can also be trademarked, provided it distinguishes the origin of the product or service from any other source or producer. In these trademarks, the logo must be unique and consist of more than simple stylization.

In *Star Industries v. Bacardi*, Star sued Bacardi for using a similar stylized "O" on its label for orange-flavored rum as Star used on its orange-flavored Georgi vodka. The court ruled against Star, however, because its design consisted merely of a stylized symbol that had not acquired a distinguishing (or secondary) meaning in the minds of consumers that enabled them to identify the producer.[xxi]

Examples of pure trademarked designs, without any associated words, include Nike's "Swoosh" and Apple's famous logo of an apple with a bite taken out of it.

This is why design trademarks often take a "design plus words" approach, as in the trademarked Lacoste logo featuring the word "Lacoste" above the famous green alligator.

Trade dress, on the other hand, refers to the **overall appearance** of a product or service that indicates its source. Trade dress can include features such as size, shape, color or color combinations, texture, graphics, or even particular sales techniques. The key to claiming trade dress protection is that the attributes must be distinctive. For example, Taco Cabana has trademarked its distinctive and "festive eating atmosphere having interior dining and patio areas decorated with artifacts, bright colors, paintings and murals," and won an infringement case against a competitor restaurant called Two Pesos. By creating a theme that was similar to Taco Cabana, Two Pesos created confusion in the minds of consumers.[xxii]

A design trademark may be one part of a product's trade dress, but it is distinct from trade dress because it only covers the stylization of words, letters, numbers, or of a specific design like the Lacoste green alligator.

Imagine that an Apple competitor opened a retail store with the same colors, lighting, atmosphere, and overall look and feel as the Apple store but sold their own products. The goal would not be to stop the sale of the goods, which are not necessarily infringing any of Apple's design trademarks or patents, but to stop the competitor from infringing the look and feel of the Apple store. In this case, Apple would file a trade dress suit.

Both design trademarks and trade dress are different from design patents. Although all three cover only **nonfunctional** designs and appearances, design patents strictly protect only the new and original ornamental design of an article of manufacture, and the actual drawing of a design patent limits what is protected. A design trademark, on the other hand, protects a particular word, name, symbol, or design used in commerce to distinguish a product's source. And trade dress protects the overall appearance of the product and can include anything that gives a product or service meaning and distinguishes it from those of any other producer.

A design patent might protect the new and original ornamental design of a lamp, for example, so long as that appearance does

not affect the lamp's function. A design trademark would protect the words or symbols used on the lamp that identify it as coming from a particular producer. And trade dress protects the overall "look and feel" of the lamp.

Sometimes both forms of protection can be obtained, providing an extra advantage to the owner of these rights. Examples of products with both design trademarks and design patents include the *Dustbuster* vacuum cleaner, the *Pepsi* bottle, and the *Honeywell* round thermostat.

Footnotes

iii. United States Patent and Trademark Office. (2012, August 09). *Trademark Manual of Examining Procedure* § 1202. Retrieved from USPTO.
iv. *Qualitex Co. v. Jacobson Products Co.*, 514 U.S. 159, 7 (1995) Retrieved from: Today's Engineer.
xv. Henderson, C. *International Standards for The Protection of Intellectual Property Rights: Trademarks and Geographical Indications*. Retrieved from USPTO website: USPTO.
vi. See *Southco, Inc. v. Kanebridge Corp.*, 258 F.3d 148, 149 n.2 (3d Cir. 2001).
vii. See *See In re Dana Corp.*, 12 USPQ2d 1748 (TTAB 1989).
iii. See *See In re Union Oil Co.*, 33 USPQ 43 (C.C.P.A.1937).
ix. In re Clarke, 17 U.S.P.Q.2d 1238 (T.T.A.B. 1990) Retrieved from IPMALL.
xx. See Constine, J. (2011, December 20).*Don't Let Your Company's Sound Be Stolen, Trademark it With Trademarkia*. Retrieved from TECHCHRUNCH.
xi. *Star Industries, Inc. v. Bacardi & Co., LTD.*, 412 F.3d 373 (2d Cir. 2005) Retrieved from Justia.
xii. 932 F.2d 1113, 1117 (5th Cir. 1991), *aff'd by* 505 U.S. 763 (1992), derived from Justia.

Note

Changes made:

- Inserted a space in "the**overall**".
- Added colon in Learning Objectives.
- Accessibility fixes to Footnotes.

This page titled 4.7.3.5: Subject Matter of Trademarks is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

4.7.3.6: The Spectrum of Distinctiveness

Learning Objectives

After completing this section, you will be able to:

- Discover the elements of distinctiveness within trademarks.
- Identify the various kinds of marks used.

The fundamental and overriding requirement for a trademark is **distinctiveness**. The requirement for distinctiveness is analogous to the requirement for **novelty** in patent rights and **originality** in copyright. Without distinctiveness, there can be no trademark.

As Miller and Davis note, "Naturally, a trademark must be distinctive if it is to serve the function of identifying the origin of goods and thereby avoid confusion, deception, or mistake. If a trademark is to protect purchasers from confusion over what they are purchasing, then the trademark somehow must be recognizable, identifiable, and different from other marks."[xxiii]

Some trade and service marks achieve distinction via their inherent nature—Nike's curved "Swoosh" check mark suggesting speed and agility, for example.

But other marks gain distinctiveness as a result of marketing, eventually forming a powerful association over time in the minds of consumers. As an example, the term "Raisin Bran" is merely descriptive of the foods used in the cereal and would not be eligible for trademark registration had Kellogg Company not demonstrated through evidence of use (including sales, advertising expenditures, and consumer surveys) that the buying public had come to distinctively associate Raisin Bran with this particular Kellogg's cereal.

The public's distinct association of a trademark to a company is known as secondary meaning, and that is something that can only be established over time. An understanding of secondary meaning is crucial when discussing the **spectrum of distinctiveness.**

In *Abercrombie & Fitch Co. v. Hunting World, Inc.* in 1976, Judge Henry Jacob Friendly established five basic categories of marks along a spectrum of distinctiveness, ranging from "fanciful" marks that are inherently distinctive on one end and on the other end "generic" marks that are never distinctive and thus not eligible for trademark registration.[xxiv]

Fanciful Marks

Fanciful marks are invented words, symbols, or devices that have no relation to the good or service being sold and have no meaning other than to distinctly identify the product or service and distinguish it in the minds of consumers from those of any other vendor. "Xerox" is a good example of a fanciful mark, as are "Google" and "Kodak." Fanciful marks are considered the strongest type of mark and are prima facie registrable with the USPTO.

Arbitrary Marks

Arbitrary marks are real words in common usage that have no descriptive relationship to the product or service being sold. Examples of arbitrary marks include "Apple" (for the computer company), "Oracle" (for the software company), and "Galaxy" (for the mobile phone). Arbitrary marks are also unreservedly eligible for trademark registration.

Suggestive Marks

Suggestive marks are marks that suggest or imply a quality or characteristic of the goods and services being sold. They require imagination, insight, or perception on the part of the consumer as to the nature of the article.[xxv] Examples include "iPad" for the tablet computer and "Coppertone" for the sunscreen lotion.

A suggestive mark is the minimum required for a mark to be unconditionally registrable absent a secondary meaning. Or in the

words of Judge Friendly, “The validity of the mark ends where *suggestion* ends and *description* begins.”[xxvi] From this point on along the spectrum of distinctiveness, things get a bit tricky.

Descriptive Marks

A **descriptive mark** explicitly describes the purpose, nature, or an attribute of a product or service and is, therefore, not eligible for trademark registration unless a secondary meaning or association has been developed in the public’s mind through usage.[xxvii] Ineligible descriptive marks include “Lightweight” and “Faster.” However, if a descriptive mark takes on a secondary meaning in the public mind—“Sharp” televisions, for example, or “Windows” for windowing software—then it can be eligible for trademark registration.

The bar against descriptive marks is designed to protect for public use common concepts and language that everyone needs to use. Imagine having to pay the coiner of the term “reverse mortgage” every time you wanted to describe such a financial arrangement. It would create a nightmare of social costs, which can be prevented simply by allowing the descriptive term to remain in the public domain. The bar against descriptive marks spotlights the overall social purpose of trademarks, which is to prevent fraud and confusion in the market while not conferring a monopoly on anyone.

Generic Marks

The last and least distinctive category of marks are **generic marks** that are simply the common name for the goods and services being sold. Examples of generic marks might be “Aspirin” for the analgesic, or the use of the word “Baskets” for a basket store without any accompanying logo or other design element to distinguish from a merely generic description of the items sold. A generic mark is never eligible for registration no matter how much evidence a mark owner offers that it has acquired a secondary meaning through advertising and marketing.

Ironically, a product or service can become so successful in the marketplace that the public begins to associate its trademark with an entire category of similar products and services. When this happens, a valid trademark can become a generic name and the trademark will be lost, as happened with both “Cellophane” and “Aspirin.” To prevent this from happening, companies like “Kleenex,” “Xerox,” and more recently “Google” have gone to great legal and advertising expense to prevent the generic misuse of their trademarks so these remain valid.

There are many companies that sell aspirin, but aspirin itself can no longer be trademarked.

Adobe, for example, sent emails to many web authors advising them to discontinue using the term “photoshopped” and instead say that their photos were “modified by Adobe Photoshop software.” Similarly, Xerox spent heavily on advertisements warning that “you cannot Xerox a document, but you can copy it on a Xerox brand copying machine.”

Footnotes

iii. Op. cit., Miller and Davis.
iv. *Abercrombie & Fitch Co. v. Hunting World, Inc.*, 537 F.2d 4, Court of Appeals. Retrieved from Google Scholar.
xv. Derived from: Bitlaw degrees.
xvi. Op. cit., *Abercrombie & Fitch Co. v. Hunting World, Inc.*
vii. United States Patent and Trademark Office. (2011, January). *Trademark Manual of Examining Procedure (TMEP) - 8th Edition*. Retrieved from USPTO.

Note

Changes made:

- Inserted a space in "and**originality**".
- Inserted a space in "where*suggestion*".
- Added a colon in Learning Objectives.
- Accessibility fixes for Footnotes.

This page titled 4.7.3.6: The Spectrum of Distinctiveness is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

4.7.3.7: Bars to Trademark

Learning Objectives

After completing this section, you will be able to:

- Identify the bars to trademark.
- Dissect criteria for trademark registration.

The Prior Use Bar

No one can register a trademark if it has been used by other parties prior to its use by the applicant. Indeed, if the applicant uses such a mark, he or she may be liable for trademark infringement.[xxviii] The only exception to the prior use bar is when the mark has been used in an entirely different market or field of business, as in the trademark "Progressive" granted to two companies, one a backpack manufacturer and the other an auto insurance company.

The Functionality Bar

To be trademarked, no word, symbol, design, device, scent, or sound may purely or primarily serve a functional purpose. Nor can it be trademarked if its absence in or on the product inflates its cost or reduces its quality.[xxix]

Thus as noted earlier, the scent of a perfume cannot be trademarked because the core requirement of a perfume is that it have a pleasant and distinctive smell.

The design of a chair presents an interesting dilemma, however. Does the design enable a person to sit—i.e., does it have a functional purpose? Or does the design distinctively mark an otherwise perfectly functional chair as coming from an identifiable producer?

This "chicken and egg" question has confronted many applicants, trademark examiners, judges, and juries over the years. The issue is usually decided by asking whether the appearance of a product or its components is inspired more by design or by function?

The issue is most easily decided when the functionality of an item resides in a specific part of the product, and the design elements are found in other parts of the product. A great example of this is the "A Christmas Story Leg Lamp" based on the classic 1983 film comedy. The functional part of the lamp is its electronics and light fixture sitting atop the base. The base, however, is in the shape of a woman's stocking-clad leg, and is thus clearly a design feature that distinguishes the lamp from any other on the market, and it is trademarked (No. 77105065).

Subject Matter Bars

In addition to prior use and functionality, certain kinds of subject matter are generally not eligible for trademark registration. These are surnames, geographic marks, and ornamental marks (immoral or offensive marks were once also ineligible, but recent case law has changed that).[xxx]

Surnames or family names such as "Smith" and "Johnson" cannot be trademarked except under certain conditions. The USPTO assesses five factors in determining whether a surname is eligible for a trademark:

- Whether the surname is rare.
- Whether the term is the surname of anyone connected with the applicant.
- Whether the term has any recognized meaning other than as a surname.
- Whether it has the "look and feel" of a surname.

- Whether the stylization of lettering is distinctive enough to create a separate commercial impression.[xxxi]

"If the mark's character as a surname predominates in terms of its 'primary significance to the purchasing public,' then it is prohibited," write Miller and Davis.[xxxii] But over time, if the surname becomes associated in the public mind with the distinctive origin of products or services, then it may be eligible for a trademark, as is the case with "Sears" and "Macy's."

Geographic marks must also be evaluated for eligibility on the basis of several factors. If a geographic mark is simply *descriptive*, the following tests apply:

- The primary significance of the mark is a generally known geographic location.
- The goods or services originate in the place identified in the mark.
- Purchasers would be likely to believe that the goods or services originate in the geographic place identified in the mark.[xxxiii]

The test above determines whether there is a literal association to the name or not. For example, if an entrepreneur launches a company called "Los Angeles Shoes" to produce shoes in Los Angeles, the name would not pass the test for trademark eligibility because the primary significance of the mark is that it is a generally known location, the product is produced in the location stated in the mark, and consumers would believe the shoes were made in Los Angeles.

The goal of the bar against geographically **descriptive** marks, then, is to protect one party from seizing ownership of a geographic term that ought to remain in the public domain.

If a geographic mark is deceptively **misdescriptive**, however, the test is as follows:

- The primary significance of the mark is a generally known geographic location.
- The goods or services do not originate in the place identified in the mark.
- Purchasers would be likely to believe that the goods or services originate in the geographic place identified in the mark.
- The misrepresentation is a material factor in a significant portion of the relevant consumer's decision to buy the goods or use the services.[xxxiv]

The goal of this test for geographically **misdescriptive** marks is to prevent the confusion and deception of the consumer. For example, if an entrepreneur launches a fashionable shoe business called "New York Shoes" but actually manufactures the shoes in China, he will not receive a trademark for the name because it will likely confuse or deceive consumers.

Sometimes, however, a geographic name may be trademarked, as was the case with the online retailing giant's trademark "Amazon." In this case, the geographic name "Amazon" is not indicative of the source of goods, nor are consumers under the impression that the goods sold by the retailer come from the Amazonian region of South America. In addition, the geographic nature of the name "Amazon" is not a significant factor in whether or not consumers decide to purchase books, clothing, electronics, or anything else from the company.

Nor can *ornamental* marks be trademarked. Ornamental marks such as a "smiley face" logo are merely decorative and are too vague to distinguish the origin of a product or service.

Footnotes

viii. 15 U.S.C. §1115.
ix. *Inwood Laboratories Inc. v. Ives Laboratories, Inc.*, 456 U.S. 844 (1982) Retrieved from supreme.justia.
xx. United States Patent and Trademark Office.(2012, March 09). *Frequently Asked Questions About Trademarks*. Retrieved from USPTO Trademarks FAQ. But see *Matal v. Tam*, 582 U.S. 218 (2017) (ruling the Lanham Act's disparagement clause unconstitutional as a violation of the First Amendment) and *Iancu v. Brunetti*, 588 U.S. 388 (2019) (ruling the Lanhan Act's prohibitions on offensive or scandalous material a viewpoint discrimination violation of the First Amendment). Retrieved from Justia.
xi. United States Patent and Trademark Office. (2011, January). *TMEP 1211.01 "Primarily Merely A Surname"*. Retrieved from USPTO *TMEP 1211.01*.
xii. Op. cit., Miller and Davis.
xiii. United States Patent and Trademark Office. (2011, January). *TMEP 1211.01(a) "Primarily Merely A Surname."* Retrieved from USPTO Trademark *TMEP 1211.01(a)*.
xiv. United States Patent and Trademark Office. (2011, January). *TMEP 1211.01(b) "Primarily Merely A Surname."* Retrieved from

USPTO Trademark *TMEP 1211.01(b)*.

Note

Changes made:

- Shortened the description for "Immoral or offensive marks".
- Added a colon in Learning Objectives.
- Added a qualifier to a footnote to include references to the *Matal v. Tam* and *Iancu v. Brunetti* cases and edited the main text to reflect that the law now permits registration of offensive marks.

This page titled 4.7.3.7: Bars to Trademark is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

4.7.3.8: Establishing Trademark Protection

Learning Objectives

After completing this section, you will be able to:

- Identify the reasons for establishing trademark registration.
- Understand the process of trademark protection.

Establishing Your Trademark Protection

As in the case with copyrights, trademarks do not necessarily have to be registered with the federal government, but doing so has benefits. The owners of unregistered trademarks have the exclusive right to use the mark concerning similar goods and services in the geographic area of actual use of the mark, the right to bring civil action against infringers, and protection against false advertising. Unregistered trademarks are usually identified by the "TM" symbol displayed on the product, package, or advertising used to promote them. Trademarks formally registered with the USPTO, however, receive additional benefits:

- A legal presumption of ownership of the mark and exclusive right to use the mark nationwide on or in connection with the goods/services listed in the registration.
- Public notice of claim of ownership of the mark.
- The ability to bring an action concerning the mark in federal court.
- The use of the U.S. registration as a basis to obtain registration in foreign countries.
- The ability to record the U.S. registration with the U.S. Customs and Border Protection (CBP) Service to prevent importation of infringing foreign goods.
- The right to use the federal registration symbol ®
- Listing in the United States Patent and Trademark Office's online databases.[xxxv]

Because registration costs money, many small businesses choose to rely upon the common law protections afforded trademarks, especially when just starting up. A case in point was Google, which used an unregistered trademark for its first six years in operation. Only in 2006 did the company formally apply for a registered trademark for "Google." It was registered in 2012.

But trademark protection can be vital to a small business, as one recent case amply demonstrates. Payam "Peter" Tabibian first registered the 1950s-style red, white, and yellow logo for his start-up Z-Burger chain in 2007, a few months before he opened his first restaurant the next year in a hip Washington neighborhood. As the *New York Times* reported, "his first restaurant was an immediate hit, attracting students from nearby American University. The place [later] drew public praise when it gave free food to federal workers who were furloughed from their jobs during government budget cutbacks."

But Mr. Tabibian eventually fell out with his partners, who then tried to stop him from continuing to use his Z-Burger brand with his expanding chain of restaurants; they also demanded he transfer the trademarks to them.

But because Mr. Tabibian had done his homework and taken the time to properly register his trademarks, a federal judge ruled in August of 2015 that he was the legitimate owner of the trademarks and could continue to use them.

As James Gibson, law professor at the University of Richmond School of Law, told the *New York Times*, the Z-Burger case shows why "it is very important early on to pick a trademark that's going to work, and make sure ownership is sewn on. That's particularly true if nobody knows who you are and you're operating in a limited area. The federal trademark gives you a lot of nationwide rights."

In order for your trademark to be eligible for registration with the USPTO, however, you must actually use it in conjunction with commercial activity—i.e., with the marketing and sale of products and services—or be about to use it within six months. Absent the selling of athletic shoes and other sports attire, Nike's $20 billion "Swoosh" logo would be nothing more than a worthless doodle.

When applying for trademark registration, the applicant must file an affidavit attesting to the date the mark was first used in

commerce as well as a specimen showing how the mark was first used, whether in advertising or on the product in a store setting. It is imperative that the applicant identify the earliest provable date of first use in commerce, to lessen the likelihood that another party can claim prior use of that mark and invalidate the trademark.

The USPTO also offers an "intent-to-use" application, which allows an entity to apply for a trademark that is not yet used in commerce.

"If you have not yet used the mark but plan to do so in the future, you may file based on a good faith or bona fide intent to use the mark in commerce," notes a USPTO publication. "A bona fide intent to use the mark is more than an idea, [but] less than market ready. For example, having a business plan, creating sample products, or performing other initial business activities may reflect a bona fide intent to use the mark."[xxxvi]

The mark must then be put into use within six months of receiving a "notice of allowance" from the USPTO or the applicant risks having to pay extension fees or losing the mark entirely.

If the trademark is still in commercial use five years after registration, the owner is eligible to receive an additional protection known as "incontestability," which immunizes the mark from many challenges, including challenges to validity, ownership, registration, and descriptiveness. (Incontestable marks can still be challenged on some grounds, such as genericism, functionality, and abandonment.) To receive incontestability status, a § 15 Declaration must be submitted to the USPTO.

Once incontestability is achieved, no one can contest the validity, ownership, or registration of the mark, nor can anyone contest the owner's exclusive right to use the mark. An individual or a corporate entity can register a mark.

A trademark can be assigned to others in the event of purchase or acquisition by another entity.

The §15 Declaration must include:

A. The registration number and the date of registration.
B. The fee for each class of goods/services in the registration to which the Declaration pertains.
C. A statement that declares:
    - *The mark has been in continuous use in commerce for a period of five years and is still in use in commerce*
    - *No final decision exists adverse to the owner's claim of ownership of the mark for the goods/services, or to the owner's right to register the mark or to keep the same on the register*
    - *No pending proceeding exists involving the claimed rights in the USPTO or in the courts*
D. A signed and dated affidavit or declaration under 37 C.F.R. §2.20[xxxvii]

There are other reasons why an owner might assign trademarks to another party, including a business name change, a security agreement, a license, a lien, as collateral for a loan, or as a result of a bankruptcy procedure. Ford used its trademarks as collateral for a $23.5 billion loan. Ford was still able to use its trademarks, but if the company had defaulted on the loan, the trademarks would have been assigned to the lender.[xxxviii]

Just like tangible assets, trademarks and other intellectual property such as patents and copyrights often have enormous value. In fact, some sources estimate that in contrast to 40 years ago, when plant, equipment, and other tangible assets comprised 80 percent of the market value of most public companies, today it is intangible assets like intellectual property, business methods, and know-how that make up roughly 80 percent of the value of public companies.[xxxix]

Trademarks can also be licensed to others. In 2012, licensed trademarks generated $5.45 billion for North American owners, reported the Licensing Industry Merchandisers' Association. That same year, the top 150 global brand licensors earned $230 billion in sales of trademarked products, with the Walt Disney Company alone accounting for nearly $40 billion in sales.

Abandonment of a mark results in loss of ownership and the associated rights of the mark. For a mark to be considered abandoned, it must be discontinued from active use in commerce with no intent to resume the commercial use of the mark. Evidence of nonuse for three consecutive years is usually considered to be sufficient proof of abandonment. Upon abandonment, the trademark owner loses the right to prevent others from infringing the mark.

Footnotes

xv. Derived from: USPTO Trademarks.

(vi. *Protecting Your Trademark: Basic Facts About Trademarks*, U.S. Patent and Trademark Office, derived from: USPTO Trademark Basics.
vii. United States Patent and Trademark Office. (21 August, 2012).*Maintain/Renew a Registration: How to Keep a Registration Alive*. Retrieved from USPTO Registration.
viii. CBS News. (23 May, 2012). *Ford Regains Assets, Including Blue Logo, After Credit Upgrade*. Retrieved from CBS News.
(ix. "Intangible Asset Market Value," Ocean Tomo, Retrieved from: Intangible Asset Market Value.

Note

Changes made:

- Removed text about watching a video at the beginning of the section.
- Changed Item C in the §15 Declaration text to use a bulleted/dashed list for accessibility.
- Added a colon in Learning Objectives.
- Added a space between the and New York Times

This page titled 4.7.3.8: Establishing Trademark Protection is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

4.7.3.9: Trademark Infringement

Learning Objectives

After completing this section, you will be able to:

- Identify the factors involved in trademark infringement.
- Understand the basis for determining similarity.

Trademark Infringement and Remedies

The purpose of trademarks is to enable consumers to distinguish the source of products and services as well as help businesses defend the quality and reputation of those goods. Trademarks benefit society by protecting consumers and businesses alike, and by preventing confusion in the marketplace. This principle lies at the heart of the trademark system.

It is also the basis on which the courts rule on claims of trademark infringement. The most common way that trademark infringement causes confusion is through **similarity**. In the words of the Lanham Act, no trademark may be registered "which so resembles" another name or mark that it will be likely "to cause confusion, or to cause mistake, or to deceive."[xl]

To determine similarity, the mark must be evaluated as a whole rather than through its individual components. This is important because there are only a limited number of parts—words, names, shapes, symbols, or devices—that can be used separately or together to create a unique and distinctive trademark that indicates the origin of goods.

The courts generally look at five key factors to determine if a defendant's use of your mark is likely to cause confusion among consumers.

The two most important are:

- The similarity in the overall impression created by the two marks (including the marks' look, phonetic similarities, and underlying meanings)
- The similarities of the goods and services involved.

But the court also looks at three other factors:

- The strength of your mark.
- Any evidence of actual confusion by consumers.
- The intent of the defendant in adopting its mark.

The basis for determining similarity, again, is not an assessment of the similarity between each component of two trademarks but of their totality as seen and experienced by the consumer. And the key test for similarity is **resemblance**, not **sameness**. In short, two marks do not have to be exact mirror images of each other to be legally similar. According to a 2003 decision by the U.S. Court of Appeals for the Federal Circuit, "Similarity is not a binary factor but is a matter of degree."[xli] Each and every attribute of a mark need not be replicated exactly by those of another mark to be deemed similar. But the more one mark's attributes are similar to those of another, the greater the likelihood that there is legal similarity between the marks as a whole.

The social networking site LinkedIn, for example, sports a logo made up of the joined words "Linked" and "in" displayed in a white font with the "in" component surrounded by a blue colored square with rounded edges. If a new networking site for professionals was launched that featured a logo with the same words except that the "in" component was surrounded by a red colored square, there would likely be enough similarity between the two logos to justify a case for infringement. The colors of one component of the logo may be different, but when the concept, wording, font style, and design are considered as a whole, any judge or jury would probably find the similarity between the marks so substantial as to constitute infringement.

But similarity is not the only basis for assessing infringement. The question is also whether any similarity between marks is likely to cause confusion in the minds of consumers. This takes into account such factors as the similarity of the goods and services

on which the two marks are being used. If a company that made metal chains called itself "LinkedIn," that would less likely cause consumers to believe that the chains came from the same source as the social networking site, because the manufacture of chains is a very different line of business than running a social networking site.

An entity can be guilty of **contributory infringement** if it intentionally encourages the infringement of a valid mark by a third party. One can also be liable for contributory infringement if one continues to produce or distribute a product with knowledge that the beneficiary is infringing a mark. As the Supreme Court held in the 1982 case *Inwood Labs Inc. v. Ives Labs, Inc.*, "If a manufacturer or distributor intentionally induces another to infringe a trademark, or if it continues to supply its product to one whom it knows or has reason to know is engaging in trademark infringement, the manufacturer or distributor is contributorily responsible for any harm done as a result of the deceit."[xlii]

> "*If a manufacturer or distributor intentionally induces another to infringe a trademark, or if it continues to supply its product to one whom it knows or has reason to know is engaging in trademark infringement, the manufacturer or distributor is contributorily responsible for any harm done as a result of the deceit*."[xliii]
>
> -*Inwood Labs., Inc. v. Ives Labs., Inc.*

Another type of infringement is **vicarious infringement**. The definition of vicarious infringement is somewhat convoluted, but Cornell University's Legal Information Institute describes it thusly:

> "A person may be held liable for the infringing acts committed by another if he or she had the right and ability to control the infringing activities and had a direct financial interest in such activities. The existence of direct infringement is required to establish a claim of vicarious infringement; however, it is not necessary for the alleged infringer to have intent or knowledge of the infringement."[xliv]

Vicarious infringement most often occurs on Internet sites that allow outside parties to post or upload files. One such example was *Sega Enters., Ltd. v. MAPHIA*, in which third parties uploaded unauthorized games that displayed the Sega mark to the Maphia site with the knowledge and consent of—and to the profit of—Maphia owners.

Finally, there is a form of harm called **trademark dilution**, which is described by the International Trademark Association as "the weakening of a famous mark's ability to identify and distinguish goods or services, regardless of competition in the marketplace or the likelihood of confusion. Dilution typically occurs as the result of blurring or tarnishment of the famous mark. . . . The concept of dilution developed from the idea that because some marks are so well-known and famous, they deserve an extra level of protection beyond the likelihood-of-confusion analysis. Dilution theory seeks to prevent the coexistence of a mark that is sufficiently similar to a famous mark, regardless of the goods and/or services associated with the allegedly diluting mark."

The Federal Trademark Dilution Act of 1995 and the Federal Dilution Revision Act of 2006 created a federal cause of action for trademark dilution that is separate from infringement. Generally speaking, an owner of a trademark eligible for dilution protection can prevent another company from using that mark no matter how dissimilar the goods and services of the two companies are. That makes dilution protection potentially broader in scope than infringement protection. After the 2006 revision, however, only marks that are "widely recognized by the general consuming public of the United States" are eligible for dilution protection, which means that only a very small percentage of trademarks used in the United States—undoubtedly less than 1 percent—qualify for that protection. Google has protection against dilution. Your local dry cleaner does not.

Footnotes

xl. See 15 U.S.C. § 1052 Retrieved from Cornell Law.
xli. In re Coors Brewing Co., Fed. Cir. 2003 Retrieved from: Ipo.org.
xlii. *Inwood Labs., Inc. v. Ives Labs., Inc.*, 456 U.S. 844, 854 (1982) Retrieved from Google Scholar.
xliii. *Inwood Labs., Inc. v. Ives Labs., Inc.*, 456 U.S. 844, 854 (1982) Retrieved from Google Scholar.
xliv. Derived from: Cornell vicarious infringement.

Note

Changes made:

- Removed the text about watching a video at the beginning of the section.
- Inserted a space in "is**resemblance**".
- Inserted a space in "called**trademark**".
- Added a colon in Learning Objectives.
- Edited footnotes for accessibility.
- Removed stray period by *Sega v. MAPHIA*.

This page titled 4.7.3.9: Trademark Infringement is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

4.7.3.10: Trademark Remedies

Learning Objectives

After completing this section, you will be able to

- Explain remedies for trademark infringement.
- Discuss effects of infringement.

Remedies for trademark infringement are similar to those available to owners of other intellectual property, including the possibility of receiving monetary damages for lost profits and injunctive relief. But even more so than is the case with patent and copyright owners, trademark owners almost invariably insist on injunctive (e.g. prohibiting continued sales) to cure the infringement. That's because when a trademark is infringed, the biggest loss to the owner is not necessarily revenue but the potential damage to the reputation of the brand in the minds of consumers.

Imagine, for example, that a manufacturer in Bangladesh started shipping substandard athletic shoes with the Nike "Swoosh" on them into the U.S. market. The loss of revenue resulting from U.S. customers mistakenly buying the Bangladeshi manufacturer's counterfeit shoes might not amount even to a rounding error on Nike's books. But if customers began to lose trust in the "Swoosh" symbol as a reliable indicator of Nike's quality design and manufacturing, the damage to Nike's brand and future profits could be significant. The company would want to secure an immediate injunction against the importation of the offending shoes and thereby stop the ongoing damage to its trademark and brand.

Incidentally, in cases where the products being imported are not patented or copyrighted, the courts may only require the importer to stop use of the mark, but allow for the continued importation of the product.[xlv]

In addition to injunctive relief, monetary damages may also be awarded in trademark infringement cases. Monetary damages are primarily aimed at compensating trademark owners for lost profits due to the infringing activity, and may be trebled if the infringement is deemed intentional. The courts may base an award on any or all of the following measures:

- The defendant's profits, either as an indicator of the plaintiff's loss or under an unjust enrichment theory.
- Actual business damages and losses suffered by the mark owner as a result of the infringement, including lost profits and the cost of notifying customers.
- Punitive damages to punish the infringer (available in state courts only)
- Reasonable attorneys' fees incurred by the mark owner to prosecute the case.

Footnotes

xlv. See *William R. Warner & Co. v. Eli Lilly & Co.*, 265 U.S. 526 (1924).

This page titled 4.7.3.10: Trademark Remedies is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

4.7.3.11: Fair Use of Trademarks

Learning Objectives

After completing this section, you will be able to:

- Discuss fair use of trademarks.
- Identify examples of fair use cases.

Not every use of a trademark is infringement.

Individuals and entities may be permitted the fair use of another's trademark under two conditions. These are known as **nominative fair use** and **classic fair use**.

Nominative fair use covers many occasions on which a party other than the trademark owner is using the mark of another party to refer to genuine goods or services. An example of nominative fair use is when an auto repair shop advertises that it repairs BMWs and Hondas. It is using the mark "Honda" to refer to cars genuinely produced by the Honda Motor Company, and simply asserting that it is able to conduct repairs on those cars. Another example of nominative fair use is use of a competitor's trademark in comparative advertising. When Verizon advertises that it has better coverage than AT&T, it is making nominative fair use of AT&T's mark because it is actually referring to AT&T's services.

Under nominative fair use, the entity using the trademark is permitted to use only so much of the mark as is necessary to identify the good or service to which it is referring, and cannot imply any sponsorship or endorsement by the mark holder (unless, of course, the repair shop really is a BMW- or Honda- approved repair facility). Another example is the use of another's trademark for the purpose of a product review.

Classic fair use occurs when a trademark is used in good faith for its primary meaning, and no consumer confusion is likely to occur. Classic fair uses typically involve marks that are descriptive. For example, even though Kellogg's has gained secondary meaning for the mark "All Bran," another cereal manufacturer might be able to describe its cereal as "all bran," if indeed it consists entirely of the hard outer layers of cereal grains.

Although fair use of trademarks is in many ways similar to the fair use of copyrights, there are differences. Under copyright law, you can use a copyright for the purpose of parody. You can also use trademarks for the purpose of parody, but whoever does so must be very careful to avoid confusion in the marketplace and diluting or tarnishing the trademark in question.

Note

Changes made to original OER material:

- Changed "using the mark another party" to "using the mark of another party".
- Added a colon in Learning Objectives.
- Accessibility fixes.

This page titled 4.7.3.11: Fair Use of Trademarks is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by OpenStax via source content that was edited to the style and standards of the LibreTexts platform.

4.7.3.12: Assessment Questions

1.

What is the purpose of the intellectual property right we call a trademark?

A. To protect the inventions of businesses.
B. To indicate the origin of goods or services.
C. To reward businesses for their creative works.

Answer available at the end of this section.

2.

Which of the following is NOT true of trademarks?

A. They protect the public by preventing confusion or deception about the source of goods and services.
B. They protect the market reputation and goodwill of the producers of goods.
C. They are granted at the moment of creation.
D. They are excellent marketing and advertising tools.

Answer available at the end of this section.

3.

How much did Nike pay a graphic designer in 1971 to produce its swoosh logo now estimated to be worth as much as $20 billion?

A. $850,000.
B. $25,000.
C. $35.

Answer available at the end of this section.

4.

What value do trademarks provide society?

A. They encourage and reward creative enterprise.
B. They marshal the benefits of this creativity to the public good.
C. They protect the consumer from deception.
D. All of the above.

Answer available at the end of this section.

5.

What is the legal foundation for federal trademark law?

A. The Commerce Clause of the U.S. Constitution.

B. Article 1, Section 8 of the U.S. Constitution.
C. State laws.

Answer available at the end of this section.

6.

How are trademarks different from patents and copyrights?

A. They are not limited in duration.
B. They do not hinder the sale of similar products and services by others.
C. They cannot be obtained by mere adoption, but only through commercial use.
D. All of the above are true statements.

Answer available at the end of this section.

7.

Why did trademarks not emerge until the start of widespread trade in the Bronze Age?

A. There was no government agency to register trademarks before that time.
B. That was when written records of trademarks could be printed.
C. Widespread trade requires a certain level of public trust in the provenance and quality of goods, which trademarks helped to build.

Answer available at the end of this section.

8.

Besides indicating the source of a good or service, how did trademarks change with the emergence of guilds in the Middle Ages?

A. The guilds issued trademarks only to their own members.
B. They became a means by which guilds could ensure the quality of work done by guild members.
C. The guilds started charging money to issue a trademark.

Answer available at the end of this section.

9.

What is one of the oldest trademarks still in existence?

A. The Royal Dutch Tulip Company.
B. Wedgewood China.
C. Löwenbräu Brewery.

Answer available at the end of this section.

10.

What is the significance of an old lawsuit known as *Sanforth's Case*?

A. It proves that 250 years before the Industrial Revolution, trademark infringement was viewed as deceit and a violation of laws against unfair competition.
B. It is the first trademark infringement case known to be recorded.
C. It serves as the foundation for modern trademark laws and policies.
D. All of the above.

Answer available at the end of this section.

11.

What caused U.S. trademark laws to evolve from a patchwork of state laws into a unitary federal trademark system?

A. The Industrial Revolution transformed America's many local and regional markets into a unified national economy.
B. Producers of goods and services began to grow into national entities.
C. State laws often conflicted with each other.
D. All of the above.

Answer available at the end of this section.

12.

Why were the first two national trademark laws ruled unconstitutional by the Supreme Court?

A. The court ruled that they infringed on non-trademark owners' first amendment rights to free speech.
B. The court ruled in 1879 that the patent and copyright clauses of the Constitution gave Congress no authority to regulate trademarks.
C. The court ruled that they were too vaguely written.

Answer available at the end of this section.

13.

What was the chief weakness of the Trade Mark Act of 1881?

A. It only regulated the trademarks used in commerce with other nations and with Indian tribes.
B. It only allowed for the registration but not the enforcement of trademarks.
C. It derived its authority from the Commerce Clause of the U.S. Constitution.

Answer available at the end of this section.

14.

The Lanham Act is the principal federal law on trademarks in the United States. Which of the following was NOT a feature of trademark law enacted by the act?

A. It broadened the national system of trademark registration.
B. It set the trademark renewal period.
C. It required registration of trademarks before owners could gain access to the federal courts for redress.
D. It established remedies in the case of infringement.

Answer available at the end of this section.

15.

When does federal law create a presumption of trademark abandonment?

A. When its owner ceases to use it in commerce.
B. When its owner ceases to employ it throughout the national territory of the United States.
C. When its owner ceases to use the trademark in commerce for three or more years.

Answer available at the end of this section.

16.

Of the four individual types of marks, what does a trademark identify?

A. Services provided to either consumers or other businesses.
B. Physical goods or commodities that are manufactured, produced, grown, or that exist naturally.
C. Both of the above.

Answer available at the end of this section.

17.

A collective mark identifies a trademark owned by a member of a certain organization. True or False?

A. True.
B. False.

Answer available at the end of this section.

18.

A certification mark indicates what?

A. The certified or verified origin of a product or service.
B. A product or service that meets certain standards or requirements.

Answer available at the end of this section.

19.

A trademark may be a word (or phrase), name, symbol, or device. Which of these do the following trademarks represent?

A. Chanel handbags.
B. The MGM lion's roar.
C. Tiffany's blue color.
D. The power-on chime in Apple's Mac computers.

Answer available at the end of this section.

20.

Why has no company been able to gain trademark protection for “87 Octane” as a brand of gasoline?

A. It’s a number.
B. It’s not trade dress.
C. It doesn’t indicate the origin of the gas—whether Exxon, Shell, or another company.

Answer available at the end of this section.

21.

What’s the difference between a design protected as a trademark and trade dress?

A. A trademark-protected design consists of a discrete symbol or logo on the product or service, whereas trade dress is its overall “look and feel.”
B. A trademarked trade dress covers a product or service’s overall features like its size, shape, and color combinations rather than a particular symbol or design.
C. Both of these.

Answer available at the end of this section.

22.

In what way are trademarked designs and trade dress similar to design patents?

A. All three protect the unique look of a company’s products from being infringed.
B. All three protect only the nonfunctional elements of the product’s appearance.
C. Both of these.

Answer available at the end of this section.

23.

What is the trademark world’s corollary to novelty in patents?

A. Non-obviousness.
B. Distinctiveness.
C. Originality.

Answer available at the end of this section.

24.

Of the following five categories of marks—fanciful, arbitrary, suggestive, descriptive, and generic—which one of them can never be protected as a trademark?

A. Descriptive.
B. Generic.
C. Fanciful.

Answer available at the end of this section.

25.

Which kind of mark is owned by the following companies?

A. Apple.
B. Xerox.
C. Coppertone.
D. Windows.

Answer available at the end of this section.

26.

Why was the store Toys “R” Us able to gain trademark protection for its name, even though “toys” is a generic word used by thousands of toy stores nationwide?

A. Because Toys “R” Us was the first to use the name “toys” and thus has priority.
B. Because Toys “R” Us was the first to use the name nationwide.
C. Because of the fanciful spelling of its name.

Answer available at the end of this section.

27.

Why did “Cellophane” and “Aspirin” lose their trademarks

A. The original companies making these products went out of business.
B. The public came to associate the name with an entire category of similar products.

Answer available at the end of this section.

28.

What are some of the bars to getting a trademark?

A. The prior use bar.
B. The functionality bar.
C. Subject matter bars.
D. All of the above.

Answer available at the end of this section.

29.

If geographic marks that are deceptively misdescriptive are ineligible for trademark protection, then why was Amazon able to gain trademark protection for its name, which refers to a geographic area of South America?

A. The name “Amazon” is not indicative of the geographic source of the retailer’s goods, nor do consumers think those goods come from the Amazonian region.
B. The name “Amazon” is not a factor in whether or not consumers decide to purchase books, clothing, electronics, or anything else from the company.
C. Amazon’s brand stands for low prices and great customer service, not for anything of a geographic nature.

D. All of the above.

Answer available at the end of this section.

30.

As with copyrights, trademarks do not have to be registered with the federal government, but doing so gives their owners all of the following additional rights EXCEPT for which one?

A. The exclusive right to use the mark in their geographic area of business.
B. The ability to seek assistance from the U.S. Customs and Border Protection Service in preventing the importation of infringing foreign goods.
C. The exclusive right to use the mark nationwide.
D. The right to use the federal registration symbol ® instead of the ™ symbol.

Answer available at the end of this section.

31.

You can't file a trademark registration application unless you are currently engaged in business.

A. True.
B. False.

Answer available at the end of this section.

32.

Which of the following explains why a trademark owner may assign it to another?

A. A business name change.
B. A bankruptcy proceeding.
C. As collateral for a loan.
D. All of the above.

Answer available at the end of this section.

33.

How does trademark infringement harm consumers?

A. By confusing consumers about the source of goods or services.
B. By allowing others to "palm off" lower-quality knockoffs in place of the high-quality goods they are seeking to buy.
C. By damaging trust between consumers and providers of goods and services.
D. All of the above.

Answer available at the end of this section.

34.

How is the "similarity" between two marks in a trademark infringement case determined?

A. By examining each individual component of a mark and seeing if another mark uses individual components that are similar.
B. By examining the totality of the two marks as seen and experienced by the consumer in the context of the goods or services on which they are used and seeing if they are similar enough to cause confusion or deception.

Answer available at the end of this section.

35.

What is trademark dilution as opposed to trademark infringement?

A. Dilution occurs when a mark famous to the general American public is blurred or tarnished by a similar mark, regardless of whether that mark is used on similar goods or services.
B. Dilution occurs when a trademark grows weaker in the minds of consumers as a result of its owner not advertising or featuring it prominently enough on products.

Answer available at the end of this section.

36.

Which of the following is true about trademark remedies?

A. Damages can be significantly increased if trademark infringement is deemed willful.
B. Courts frequently use injunctions based on trademark law to stop the importation of products that once carried infringing marks even after those marks have been removed.
C. Federal trademark law also provides for punitive damages for trademark infringement.

Answer available at the end of this section.

37.

A party may be permitted the fair use of another's trademark under two conditions, known as **nominative fair use** and **classic fair use**. Which of the following is an example of nominative fair use?

A. When a business compares its products with those of a competitor, with no intent to confuse customers as to the source of the competitor's goods.
B. When a trademark of another is used to refer to that other's goods or services in legitimate connection with one's own, such as a notice on the packaging of a small company's graphic design program that it can run on the Microsoft Windows operating system.
C. Both of the above are correct.

Answer available at the end of this section.

Answers

1.
B. To indicate the origin of goods or services.

2.
C. They are granted at the moment of creation.

3.
C. $35.

4.
D. All of the above.

5.

A. The Commerce Clause of the U.S. Constitution.

6.

D. All of the above are true statements.

7.

C. Widespread trade requires a certain level of public trust in the provenance and quality of goods, which trademarks helped to build.

8.

B. They became a means by which guilds could ensure the quality of work done by guild members.

9.

C. Löwenbräu Brewery.

10.

D. All of the above.

11.

D. All of the above.

12.

B. The court ruled in 1879 that the patent and copyright clauses of the Constitution gave Congress no authority to regulate trademarks.

13.

A. It only regulated the trademarks used in commerce with other nations and with Indian tribes.

14.

C. It required registration of trademarks before owners could gain access to the federal courts for redress.

15.

C. When its owner ceases to use the trademark in commerce for three or more years.

16.

B. Physical goods or commodities that are manufactured, produced, grown, or that exist naturally.

17.

A. True.

18.

A. The certified or verified origin of a product or service.

19.

A. Name (Chanel)

B. Device (a sound)

C. Device(the distinctive blue color)

D. Device (a sound)

20.

C. It doesn't indicate the origin of the gas—whether Exxon, Shell, or another company.

21.

C. Both of these.

22.

C. Both of these.

23.

B. Distinctiveness.

24.

B. Generic.

25.

A. Arbitrary. "Apple" is an example of a real word in common usage that has absolutely no relationship to the product or service being sold.

B. Fanciful. "Xerox" is a completely invented word that has no meaning except to indicate the source of a copier manufacturer. Fanciful marks are the strongest type of mark you can have.

C. Suggestive. The word "iPad" suggests or implies a quality or characteristic of the good being sold and is distinctive to its manufacturer (Apple).

D. Suggestive. Again, "Coppertone" suggests the quality or color of the tan you'll get, while indicating which company is selling it.

E. Descriptive. "Windows" is one of the rare marks that describe the nature of a distinctive product (computer software) from a unique manufacturer.

26.

C. Because of the fanciful spelling of its name.

27.

B. The public came to associate the name with an entire category of similar products.

28.

D. All of the above.

29.

D. All of the above.

30.

A. The exclusive right to use the mark in their geographic area of business.

31.

B. False.

32.

D. All of the above.

33.

D. All of the above.

34.

B. By examining the totality of the two marks as seen and experienced by the consumer in the context of the goods or services on which they are used and seeing if they are similar enough to cause confusion or deception.

35.

A. Dilution occurs when a mark famous to the general American public is blurred or tarnished by a similar mark, regardless of whether that mark is used on similar goods or services.

36.

A. Damages can be significantly increased if trademark infringement is deemed willful.

37.

B. When a trademark of another is used to refer to that other's goods or services in legitimate connection with one's own, such as a notice on the packaging of a small company's graphic design program that it can run on the Microsoft Windows operating system.

This page titled 4.7.3.12: Assessment Questions is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

Terms

- **ASIC:** Application Specific Integrated Circuit
- **ELF:** Early Life Failure
- **EMC:** Electromagnetic Compatibility emission testing measures the electromagnetic emissions produced by electronic devices. This testing ensures that devices do not interfere with other electronic equipment and comply with established standards. The focus is on radiated emissions and conducted emissions. Electromagnetic Compatibility (EMC) testing is a critical process to ensure that electronic devices and systems operate without causing or being affected by electromagnetic interference (EMI).
- **EMI:** Electromagnetic interference. EMI testing measures emissions of a device and the device's immunity to interference for regulatory compliance.
- **FOSS:** Free and Open Source Software
- **GPL:** General Public License
- **HALT:** Highly Accelerated Life Testing. It is performed during the system design phase.
- **HASS:** Highly Accelerated System Screening is performed after products have been produced by manufacturing as a production quality assurance method used to identify latent manufacturing defects in newly built products. The goal is to find potential early-life failures that might be caused by manufacturing assembly issues, component soldering defects, or part misalignments that may have escaped normal quality control processes.
- **OSS:** Open Source Software
- **PCB:** Printed Circuit Board
- **PCBA:** Printed Circuit Board Assembly
- **SDK:** Software Development Kit, supporting software supplied by an external network ASIC vendor for integrating ASICs into network applications. It may also refere to Switch Development Kit.

This page titled 4.8: Glossary is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

5: EU ROHS COMPLIANCE

The RoHS thumbnail image was taken from File:RoHS.svg - Wikimedia Commons. This file is made available under the Creative Commons CC0 1.0 Universal Public Domain Dedication. The person who associated a work with this deed has dedicated the work to the public domain by waiving all of their rights to the work worldwide under copyright law, including all related and neighboring rights, to the extent allowed by law. You can copy, modify, distribute and perform the work, even for commercial purposes, all without asking permission.

5.1: Introduction

5.2: The Directive

5.3: Implementing the Directive

5.4: Project Management Considerations

5.5: Legal Considerations

5.6: Glossary

This page titled 5: EU RoHS Compliance is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

5.1: Introduction

This case study concentrates on the initial version of the Restriction of Hazardous Substances that went into effect on July 1, 2006 restricting the presence of six substances in electrical and electronic equipment.

Subsequent versions of the RoHS Directive went into effect on January 3, 2013 and on July 22, 2019 (covering 4 more substances).

RoHS Directive Evolution

RoHS versions

Version	Directive	Effective Date	Coverage
RoHS 1	2002/95/EC	7/1/2006	Restricts the use of six hazardous materials found in electrical and electronic products: • Lead • Mercury • Cadmium • Hexavalent Chromium • Polybrominated biphenyls (PBB) • Polybrominated diphenyl ethers (PBDE)
RoHS 2	2011/65/EU	1/3/2013; Exemptions expire 7/21/2016	RoHS compliance is required for “CE” marking of products. RoHS 2 added Categories 8 and 9, and has additional compliance record-keeping requirements.
RoHS 3	2015/863	7/22/2019	Added 4 more substances (phthalates): • Bis(2-Ethylhexyl) phthalate (DEHP): < 1000 ppm • Benzyl butyl phthalate (BBP): < 1000 ppm • Dibutyl phthalate (DBP): < 1000 ppm • Diisobutyl phthalate (DIBP): < 1000 ppm

Although the RoHS directive applied to EU member states, it also affected companies outside of the EU that were going to "put on the market" electronic equipment after July 1, 2006. Technology companies in the Unites States were affected by this because many companies sold a considerable portion of their products for revenue in the EU.

This page titled 5.1: Introduction is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

5.2: The Directive

Prohibited substances

Article 4 Prevention paragraph 1 states: "Member States shall ensure that, from 1 July 2006, new electrical and electronic equipment put on the market does not contain **lead, mercury, cadmium, hexavalent chromium, polybrominated biphenyls (PBB) or polybrominated diphenyl ethers (PBDE)**. National measures restricting or prohibiting the use of these substances in electrical and electronic equipment which were adopted in line with Community legislation before the adoption of this Directive may be maintained until 1 July 2006."[1]

Operationally, this did not mean a total eradication of the substances, but rather fitting under a certain threshold maximum concentration. These concentrations were later published in the RoHS Directive that followed the initial one.[2]

Table 5.2.1: RoHS Substance Concentration Limits

Substance	Maximum Concentration (percent)	Comments
Lead (element Pb)	0.1% (1,000 parts per million PPM)	Used in solders, components, printed circuit boards
Hexavalent Chromium (Cr6)	0.1% (1,000 parts per million PPM)	Used in dyes, plating solutions, alloys
Mercury (element Hg)	0.1% (1,000 parts per million PPM)	Used in some components, batteries, pigments
Cadmium (element Cd)	0.01% (100 parts per million PPM)	Used in batteries, platings
Polybrominated Biphenyls (PBB)	0.1% (1,000 parts per million PPM)	Used as a flame retardant in plastics, insulation
Polybrominated Diphenyl Ethers (PBDE)	0.1% (1,000 parts per million PPM)	Used as a flame retardant in plastics, insulation

Initial exemptions

Article 4 Prevention paragraph 2 states: "Paragraph 1 shall not apply to the applications listed in the Annex."[1]

The Annex states:

Applications of lead, mercury, cadmium and hexavalent chromium, which are exempted from the requirements of Article 4(1)

1. Mercury in compact fluorescent lamps not exceeding 5 mg per lamp.
2. Mercury in straight fluorescent lamps for general purposes not exceeding:

    - halophosphate 10 mg
    - triphosphate with normal lifetime 5 mg
    - triphosphate with long lifetime 8 mg.

1. Mercury in straight fluorescent lamps for special purposes.
2. Mercury in other lamps not specifically mentioned in this Annex.
3. Lead in glass of cathode ray tubes, electronic components and fluorescent tubes.
4. Lead as an alloying element in steel containing up to 0,35 % lead by weight, aluminium containing up to 0,4 % lead by weight and as a copper alloy containing up to 4 % lead by weight. [sic, using European comma instead of a decimal point]
5. [sic, no text before the sub-bullets]

    - Lead in high melting temperature type solders (i.e. tin-lead solder alloys containing more than 85 % lead),
    - lead in solders for servers, storage and storage array systems (exemption granted until 2010),
    - lead in solders for network infrastructure equipment for switching, signalling, transmission as well as network management for telecommunication,

- lead in electronic ceramic parts (e.g. piezoelectronic devices).

8. Cadmium plating except for applications banned under Directive 91/338/EEC (1) amending Directive 76/769/EEC (2) relating to restrictions on the marketing and use of certain dangerous substances and preparations.
9. Hexavalent chromium as an anti-corrosion of the carbon steel cooling system in absorption refrigerators.
10. Within the procedure referred to in Article 7(2), the Commission shall evaluate the applications for:

    - Deca BDE, — mercury in straight fluorescent lamps for special purposes,
    - lead in solders for servers, storage and storage array systems, network infrastructure equipment for switching, signalling, transmission as well as network management for telecommunications (with a view to setting a specific time limit for this exemption), and
    - light bulbs, as a matter of priority in order to establish as soon as possible whether these items are to be amended accordingly[1]

Commission Decision 2005/747/EC amended the original RoHS Directive on 21 October 2005:

- 7...lead in solders for servers, storage and storage array systems, network infrastructure equipment for switching, signalling, transmission as well as network management for telecommunications...
- 15. Lead in solders to complete a viable electrical connection between semiconductor die and carrier within integrated circuit Flip Chip packages.

References

1. DIRECTIVE 2002/95/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 January 2003 on the restriction of the use of certain hazardous substances in electrical and electronic equipment. OJ L 37/19, 13.2.2003.
2. DIRECTIVE 2011/65/EU OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 8 June 2011 on the restriction of the use of certain hazardous substances in electrical and electronic equipment (recast). OJ L 174/88, 1.7.2011.

This page titled 5.2: The Directive is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

5.3: Implementing the Directive

Let's take a look at how technology companies approached attaining compliance with the EU RoHS Directive.

Level of Compliance

For their initial efforts at RoHS compliance, technology companies primarily made use of two exemptions mentioned earlier:

- lead in solders for servers, storage and storage array systems (exemption granted until 2010),
- lead in solders for network infrastructure equipment for switching, signalling, transmission as well as network management for telecommunication.

The terms "RoHS-6" and "RoHS-5" came into use with the following meanings:

- RoHS-6 is required (being under the threshold concentrations for all of the six substances), unless an **exemption** applies
- RoHS-5 involves invoking the Directive's legal exemption for lead in solder, where compliance is being under the threshold concentrations for all of the six substances except for lead.

Applications for exemptions are reviewed by the EU Technical Adaptation Committee that reports to the EU Commission. Requested and granted exemptions are published. The duration of a granted exemption is approximately 4 years. It is also driven by what happens in the market (what is demonstrated to be technically feasible).

The technical literature at the time was filled with analysis of replacements for lead solder and the risks involved. As a result, most initial compliance was at the RoHS-5 level.

Self-declaration

Compliance with the EU RoHS Directive is a “self-declaration” process, meaning that the company putting the product on the market in the EU is responsible for making the declaration of compliance.

Determine Scope

At the time of the RoHS Directive, a tech company likely had many products for sale in the EU. A key activity driven by Product Management and Sales organizations was to determine which of the organization’s products would be certified as RoHS-compliant. This required examination of product sales history, inspecting contracts with customers, reacting to technical feasibility assessments by Engineering, among other things.

Some products might require certifications-only work to demonstrate compliance. Other products might require engineering changes need to be made, necessitating engineering development work and re-validation.

At the end of this analysis work, some products would be **non-compliant**. The disposition of non-compliant products could be:

- Non-compliant products could still be sold in regions other than the EU
- Non-compliant products are potential candidates for product retirement.

Create a RoHS-compliance Process

To attain RoHS compliance, the following were elements of the process:

- Request and collect certifications from suppliers
- Ensure the validity of certifications, possibly testing composition of some parts via an outside lab, at some expense.
- Collect certifications via the organization's official Engineering change control management process using Engineering Change Orders (ECO) and Manufacturing Change Orders (MCO), attaching proof of certification to parts.

- Monitor the compliance level of product Stock Keeping Units (SKUs) through Bill of Material (BOM) roll-up of compliance information associated will all parts.
- Do not throw the “OK to ship” switch until 100% covered.

Regarding that last item, the RoHS Directive imposes a very high standard to achieve compliance. Essentially every single part in a product must be compliant. Thinking of this as receiving a grade in a course would yield this:

- Grade for being 99.99% compliant or less: **F** (“Fail”). Often a grade of 99.99% would earn a student an A+, but not so here!
- Grade for being 100.00% compliant: **A** (actually, more like a “Pass”)

Certifications

The types of evidence of compliance from part suppliers at the time varied greatly. In addition to end-product companies starting late with their compliance efforts, such was often the case with electronic component suppliers. The types of information available included:

- Certificates of Compliance/Conformance
- A binary “yes/no” compliance statement on the supplier's company letterhead, signed by an appropriate company representative.

These may or may not have detailed data on restricted substance concentrations.

As time went on, more formal methods became available such as the following IPC forms:

- IPC-1751 Generic Requirements for Declaration Process Management
- IPC-1752 Materials Declaration Management
- IPC-1752-1 and 1752-2 PDF forms, covering six classes of declaration
- IPC-1752-3 Users Guide for Material Declaration Forms

IPC-1752A is an XML based standardized framework for materials declaration management for exchanging material composition information in the supply chain. It was standardized in 2010 and subsequently amended, including in 2014.[6, 7]

Determine compliance classifications

To get started, an organization needed to define compliance classification tags in the parts database. These might include classifications such as:

- Compliant (perhaps further refined as RoHS-5 or RoHS-6)
- Non-compliant
- Not Applicable or “Waived”
- Unknown or “Missing Information” (used for parts initially before certificates of compliance became available)
- Exempt

Bill of Materials Roll-ups

A major objective was determining whether the product being sold or to be sold was RoHS-compliant or not. To do this, it was necessary to inspect the Bill of Materials (BOM) for the product and look at the compliance status of the constituent parts, being aware of the following:

- Overall Stock Keeping Unit (SKU) status is a roll-up of lower-level child parts
- The status of a parent part is determined by the status of the Manufacturer parts reporting to it
    - Replace non-compliant parts with compliant ones, possibly with a new part number (P/N)
    - If the status of manufacturer parts is mixed (e.g., some are compliant, some are non-compliant), the Approved Manufacturers List (AML) or Approved Vendors List (AVL) needs to be checked
        - Remove non-compliant parts
        - Continue to pursue certifications

AML/AVL Clean-up

As mentioned above, in certain circumstances the Approved Manufacturers List (AML) or Approved Vendors List (AVL) needed to be cleaned up to render the overall BOM to be compliant.

EXAMPLE 1				
12345-00	SCREW		RoHS-5 Compliant	**Actions:**
	Supplier 1	S1-12345	RoHS-5 Compliant	None.
	Supplier 2	S2-12345	RoHS-5 Compliant	
	Supplier 3	S3-12345	RoHS-5 Compliant	
	Supplier 4	S4-12345	RoHS-5 Compliant	
EXAMPLE 2				
12345-00	SCREW		Missing Info	**Actions:**
	Supplier 1	S1-12345	RoHS-5 Compliant	Keep chasing certs on Suppliers 3 and 4.
	Supplier 2	S2-12345	RoHS-5 Compliant	If Suppliers 3 and 4 are too slow in responding,
	Supplier 3	S3-12345	Missing Info	remove them from the AVL/AML.
	Supplier 4	S4-12345	Missing Info	
EXAMPLE 3				
12345-00	SCREW		Non-Compliant	**Actions:**
	Supplier 1	S1-12345	RoHS-5 Compliant	Keep chasing certs on Supplier 3.
	Supplier 2	S2-12345	RoHS-5 Compliant	Remove Supplier 4 from the AVL/AML.
	Supplier 3	S3-12345	Missing Info	
	Supplier 4	S4-12345	Non-Compliant	
EXAMPLE 4				
12345-00	SCREW		Non-Compliant	**Actions:**
	Supplier 1	S1-12345	Non-Compliant	Pull new P/N 67890-00, update design file/BOM.
	Supplier 2	S2-12345	Non-Compliant	Add 12345-00 and 67890 as a pair in FROM-TO list.
	Supplier 3	S3-12345	Non-Compliant	Add compliant suppliers to AVL/AML for 67890-00.
	Supplier 4	S4-12345	Non-Compliant	Alternatively (more dangerous!), keep P/N 123456-00,
				delete Suppliers 1-4 (appropriate dispositions such as
				Purge), and add new Suppliers.

Figure 5.3.1: AML/AVL Clean-up example. (Robert Slate, CIS 79 course materials)

Let's look in detail at the AML/AVL clean-up examples in Figure 5.3.1.

Example 1: In this example, each of the parts from the four suppliers has been determined to be RoHS-5 Compliant, so the part 12345-00 (Screw) rolls up as RoHS-5 Compliant.

Example 2: In this example, the parts from Suppliers 1 and 2 have been determined to be RoHS-5 Compliant. However, information from Suppliers 3 and 4 is still missing. In this case, the action to take is to continue pressuring Suppliers 3 and 4 to provide the certification of compliance for their parts. At some cut-off date or deadline, if Suppliers 3 and 4 fail to provide the certification of compliance for their parts, then their parts must be removed from the AVL for part 12345-00, leaving only Suppliers 1 and 2.

Example 3: In this example, the parts from Suppliers 1 and 2 have been determined to be RoHS-5 Compliant. Information from Supplier 3 is still missing, and as in Example 2, Supplier 3 needs to be pressured to provide the certification of compliance for their part. For Supplier 4, the information received from them designates their part as being "Non-compliant" and therefore Supplier 4 must be removed from the AVL for part 12345-00.

Example 4: In this example, the information received from all four of the suppliers indicates that all of the suppliers have indicated that their parts are Non-Compliant. In this case, a safe strategy is to remove part 12345-00 from the BOM (and the system) and replace it with a new part number (called "pulling a new P/N") say P/N 67890-00 and update engineering design files and BOMs. Then suppliers for part 67890-00 need to be added that supply compliant parts. Parts 12345-00 and 67890-00 are added to the From-To (Non-compliant to Compliant) list. Alternatively, and more risky, is to keep part 12345-00, delete Suppliers 1-4, with appropriate dispositions such as "Purge" to purge inventory of those parts from manufacturing inventory, and then hope to find suppliers with replacement compliant parts.

Compliance Tool requirements

To do the BOM roll-up checks for compliance, it was necessary to create a tool (as one of this textbook's authors did) and to regularly run the tool to assess BOM compliance (as the other author did). Running the tool was done on a weekly basis and the tool was refined as needed to incorporate learnings. This was a labor-intensive process.

On the road to meeting the RoHS Directive, industry-wide parts management databases had little to no support for RoHS

compliance. That was why we needed to create our own tool. Our Engineering parts management database tool only provided an add-on module one month before the RoHS Directive compliance deadline, which rendered it operationally not usable for the effort.

Role of Component Engineering

There is a Component Engineering role, housed either in Engineering or in Manufacturing, that performed these important responsibilities pertinent to RoHS-compliance:

- Scrutinize component data sheets.
- Verify component is suitable for higher lead-free soldering temperatures (defined in JEDEC 020-C) if going for RoHS-6 [check max temp].
- If going RoHS-5, need to ensure whether “RoHS-compliant” part is backward compatible with RoHS-5 process (e.g., will it be solderable at lower RoHS-5 temperature?).
- Check finishes: watch out for pure tin finishes for “tin whiskering” effect (ensure they pass test criteria in JESD22A121).

If your organization wanted to target target RoHS-5 compliance, there were some important considerations regarding parts or components:

- If the industry demand is weighted towards RoHS-6 parts, RoHS-5 (and non-compliant) parts will go End-of-Life (EOL) over time (but how fast was, in general, unknown).
- One needed to pay particular attention to parts such as screws, washers, stand-offs. It was unknown what the status of such parts might be on the "gray market."
- To avoid production line-down situations, proactively add RoHS-6 backward-compatible parts as alternates on the AVL/AML.
- Determine if the are any adjustments needed in manufacturing profiles (usually done by Manufacturing Engineering)

Manufacturing/Reliability

If going the full route to RoHS-6, it was necessary to be aware of a litany of things:

- Tin whiskers (see JESD22A121 spec). There was a concern that the new lead-free solders could result in tin whiskers that could cause short circuits.
- PCB qualification/PCB testing due to new board finishes. The major change is in the use of laminate materials that can withstand higher processing temperatures and different final finishes (thermal capacities higher than the “typical” 150°C glass transition (Tg) range).[5]
- Process verification
- Line and materials handling
  - Watch cross-contamination of parts if the manufacturing facility (factory) has RoHS and non-RoHS assemblies and rework going on concurrently.
- Marking of lead-free assemblies (IPC-1066)[3]
- Assembly inspection
- Factory qualification for RoHS

Lead-free soldering issues

- “Lead-free solder has generally been found to generate more defects of voiding, tomb-stoning, solder beading, bridging, and misalignment...Process optimization is critical for volume manufacturing with lead-free solder.”[1]
- Found SAC (SnAgCu) solder alloys with Ag (silver) content between 3-4% is acceptable.[1]
- Overall yield is slightly lower for lead-free solder as compared with SnPB (tin lead) solder for reflow.[1]
- AOI (Automated Optical Inspection) re-training needed due to the lead-free solder joints appearing duller and more grainy, less shiny (due to Sn dendrites). [1,4].

Mechanical shock/drop

There were concerns that the use of lead-free solders could result in issues if the product were to be dropped or experience

mechanical shock:

- “SAC (SnAgCu) solder alloys can fail at loads up to 50% lower than SnPb when subjected to shock, drop, or static board bending.”[2]
- Cause: a combination of brittle inter-metallics, board degradation due to higher reflow temperatures, and a greater transfer of stress because SAC is a stiffer material than SnPb (tin lead).[2]

Verification and Testing

Depending on the types of changes that were being made to the product for RoHS compliance, engineering involvement was required to determine whether other work was required. The team made the assessment on a case-by-case basis, based on what changed. The additional work could include some of the following items:

- HALT (Highly Accelerated Life Testing)
- HASS (Highly Accelerated Stress Screening)6
- Vibration testing
- SQA (System or Software Quality Assurance) testing with the Software group to confirm that product software recognizes and operates correctly with the new part numbers (if the part numbers were changed)
- Field reliability data collection and analysis.

References

1. “Optimization of Lead-Free Soldering Processes for Volume Manufacturing”, presented at IPC Printed Circuits Expo, SNEMA Council APEX and Designers Summit 05.
2. “Long-term reliability of Pb-free electronics”, Electronic Products, September, 2005.
3. IPC-1066 “Marking, Symbols and Labels for Identification of Lead-Free and Other Reportable Materials in Lead-Free Assemblies, Components and Devices”, January, 2005. Available from www.ipc.org.
4. IPC-A-610D-Acceptability of Electronic Assemblies, published February, 2005.
5. “Lead Free: Making the Adjustments. Surface Mount Technology (SMT), October 1, 2005.
6. IPC - Association Connecting Electronics Industries, IPC-1752A Materials Declaration Management, IPC 175X Schema Version 2.0. February 2010. IPC-1752A. Accessed July 13, 2026.
7. IPC - Association Connecting Electronics Industries, IPC-1752A with Amendments 1 and 2, Materials Declaration Management. February 2014. IPC-1752A with Amendments. Accessed July 13, 2026.

This page titled 5.3: Implementing the Directive is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

5.4: Project Management Considerations

First, some general observations:

- Many companies start very late on compliance projects.
- Sometimes an outside consultant needs to be the bearer of the news that compliance is not optional but is a necessity.
- Starting late puts enormous pressure on the project team to meet a directive or regulation's deadline.

Assuring BOM Compliance can be extraordinarily labor-intensive:

- Chasing down certificates of compliance of parts from suppliers can consume a lot of resources and take a lot of time.
- Assuring BOM Compliance by checking BOM roll-ups is an iterative process and is necessary to show progress and actionable exceptions.
- Maintaining From-To lists (Non-compliant to Compliant) is desirable to ease the work of design teams.
- There is a lot of paperwork requesting certificates of compliance of parts from suppliers, filing ECOs and MCOs to clean up the BOM, etc.

Changes in part numbering

This is a brief overview of some of the considerations associated with product part numbering and downstream effects on the work required of the project team.

- Marketing/Sales order number (e.g., 12345)
  - Different approaches have been used. Many think it best to change the P/N, e.g., to 12345G (G suffix for “Green”) or 12345PF (as in Pb-Free)
  - However, some contracts with customers specify the part number, so changing the part number could be an issue.
  - Once you decide on how to handle the product part number, it is important to stick to your decision as it affects order management systems, Return Merchandise Authorization (RMA), other IT systems, etc. Changing the decision on this in the middle of the project, or worse toward the end of the project, can have a major impact.
- Manufacturer P/N (e.g., 912345-xx-yy)
  - If xx=“Type” and yy=Rev, then define a new xx field to designate a RoHS assembly.

High-level schedule

This is an example of a high-level schedule for a company getting a late start on RoHS compliance. The full schedule with all relevant items would be considerably larger.

	Task Name	Duration	Start	Finish	Predecessors	Resource Names
1	Determine applicable products	5 days	Mon 8/1/05	Fri 8/5/05		Product Mgmt
2	Determine compliant SKU part numbering	10 days	Mon 8/8/05	Fri 8/19/05	1	Sales,Product Mgnt
3	Gather component certifications from suppliers	150 days	Mon 8/8/05	Fri 3/3/06	1	Operations
4	Add certs to components in Agile data base	140 days	Mon 8/22/05	Fri 3/3/06	3SS+10 days	Component Eng
5	Create SKU BOM compliance tester tool	10 days	Wed 9/7/05	Tue 9/20/05		PM
6	Validate SKU Compliance (compliant BOM)	168 days	Wed 9/21/05	Fri 5/12/06	5	PM,Quality
7	Remove non-compliant components from AVL via MCO	67 days	Wed 12/14/05	Thu 3/16/06	6SS+60 days	Eng PM,Operations
8	File ECOs for proto SKUs	10 days	Fri 3/17/06	Thu 3/30/06	7	Eng PM
9	Assemble proto SKUs with replacement parts	20 days	Fri 3/31/06	Thu 4/27/06	8	Operations
10	Eng test of changed SKUs	15 days	Fri 4/28/06	Thu 5/18/06	9	HW Eng,SQA
11	Release ECOs to Manufacturing for compliant SKUs	10 days	Fri 5/19/06	Thu 6/1/06	10,7,6	Eng PM
12	Manufacture compliant SKUs	20 days	Fri 6/2/06	Thu 6/29/06	11	Operations
13	Archive proof of SKU compliance	5 days	Fri 6/2/06	Thu 6/8/06	11,6	PM
14						

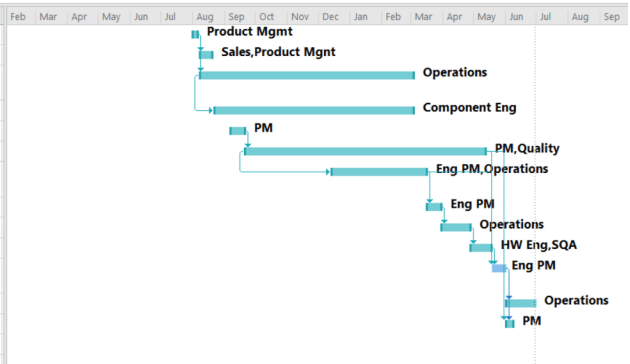


Figure 5.4.1: High Level Schedule. (Copyright; Robert Slate, CIS 79 course materials)

Accessibility Note

1. Task 1: Determine applicable products, 5 days, starting 8/1/05, ending 8/5/05, no predecessors, Resources: Product Management.
2. Task 2: Determine compliant SKU part numbering, 10 days, starting 8/8/05, ending 8/19/05, predecessor Task 1, Resources: Sales, Product Management.
3. Task 3: Gather component certifications from suppliers, 150 days, starting 8/8/05, ending 3/3/06, predecessor Task 1, Resources: Operations.
4. Task 4: Add certs to components in Agile data base, 140 days, starting 8/22/05, ending 3/3/06, predecessor Task 3SS+10 days, Resources: Component Engineering.
5. Task 5: Create SKU BOM compliance tester tool, 10 days, starting 9/7/05, ending 9/20/05, no predecessor, Resources: PM.
6. Task 6: Validate SKU Compliance (compliant BOM), 168 days, starting 9/21/05, ending 5/12/06, predecessor Task 5, Resources: PM, Quality.
7. Task 7: Remove non-compliant components from AVL via MCO, 67 days, starting 12/14/05, ending 3/16/06, predecessor Task 6SS+60 days, Resources: Eng PM, Operations.
8. Task 8: File ECOs for proto SKUs, 10 days, starting 3/17/06, ending 3/30/06, predecessor Task 7, Resources: Eng PM.
9. Task 9: Assemble proto SKUs with replacement parts, 20 days, starting 3/31/06, ending 4/27/06, predecessor Task 8, Resources: Operations.
10. Task 10: Eng test of changes SKUs, 15 days, starting 4/28/06, ending 5/18/06, predecessor Task 9, Resources: HW Eng, SQA.
11. Task 11: Release ECOs to Manufacturing for compliant SKUs, 10 days, starting 5/19/06, ending 6/1/06, predecessor Tasks 10, 7, 6, Resources: Eng PM.
12. Task 12: Manufacture compliant SKUs, 20 days, starting 6/2/06, ending 6/29/06, predecessor Task 11, Resources: Operations.
13. Task 13: Archive proof of SKU compliance, 5 days, starting 6/2/06, ending 6/8/06, predecessor Task 11,6, Resources: PM.

This page titled 5.4: Project Management Considerations is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

5.5: Legal Considerations

Directive versus Regulation

The EU RoHS is a Directive rather than a Regulation (such as GDPR discussed elsewhere in this textbook). Let's look briefly at how they differ.

Definition: Regulation

"A '**regulation**' is a binding legislative act. It must be applied in its entirety across the EU."[2]

Definition: Directive

"A "**directive**" is a legislative act that sets out a goal that EU countries must achieve. However, it is up to the individual countries to devise their own laws on how to reach these goals."[2]

Representations of Compliance

The ability of a company to assert that its product is RoHS-compliant is heavily dependent on the correctness of the assertions of RoHS-compliance by the suppliers of the components or parts. What if you doubt the validity of the assertion of RoHS-compliance of a part by the supplier?

Fundamentally, the contracts that your company has with its suppliers should cover this situation. If the part ultimately is determined to be non-compliant, the contract in place should provide a statement of remedy. Unfortunately, these remedies often only cover the cost of the parts that have been purchased. A much larger issue is recalling the product containing the non-compliant component and replacing it with a compliant one if available. Sometimes replacing a component in a product is not feasible and the whole product would need to be replaced, at considerable cost, and with an impact on the company's reputation.

Detecting Non-compliance

As mentioned earlier, compliance is a self-declaration. However, there were concerns at the time of implementation of the Directive that officials at ports of entry to the EU would be using so called "RoHS guns" to selectively test incoming products for the presence of some of the restricted substances. These "RoHS guns" were X-Ray Fluorescence devices.

X-Ray Fluorescence (XRF) helps detect hazardous substances restricted under the RoHS Directive. It identifies heavy metals like lead, mercury, and cadmium quickly and efficiently. XRF works best for inorganic elements, making it a popular tool for compliance testing.[3]

XRF uses X-rays to excite atoms in a sample. When atoms stabilize, they release energy as photons. The device measures this energy to identify elements in the sample. Each element emits a unique energy, allowing precise detection. XRF creates an energy spectrum showing what elements are present and how much of each exists. This data supports both qualitative and quantitative analysis.[3]

XRF effectively detects key RoHS-restricted elements like lead, mercury, and cadmium. It can also find chromium but cannot tell if it’s hexavalent chromium. Additional tests are necessary for that. However, XRF cannot analyze organic compounds such as phthalates or PBBs. XRF can detect total bromine which works as a pre-screening method. However, these substances require other methods like gas chromatography-mass spectrometry (GC-MS) for precise detection.[3]

XRF offers many advantages. It’s fast and non-destructive, meaning samples remain intact after analysis. It works well for solid materials with minimal preparation, saving time and reducing costs. Handheld XRF devices allow on-site testing, making them useful in industrial and environmental settings. These devices analyze samples in seconds, increasing efficiency.[3]

XRF struggles to detect light elements like carbon or oxygen. It also cannot analyze molecules or organic substances. Despite these limits, XRF remains a top choice for detecting heavy metals and ensuring compliance with environmental regulations like RoHS.[3]

Contractual part number references

A matter of much discussion during initial efforts to make products RoHS-compliant was how to mark compliant products in ways that distinguished them from non-compliant products. Some companies chose to create a new product number, sometime affixing an "E" suffix to the previous part number to designate it as environmentally compliant. Other companies chose to retain existing part numbers and just cut into production the compliant products. In this situation, RoHS-compliance could be inferred by the revision number of a printed circuit board or by the manufacturing cut-in date.

From a legal perspective, the use of new product numbers to show RoHS-compliance was potentially problematic. If the company's contract with a customer stated that a product with part number 123456789 needed to be supplied at a specified price over a period of years, attempting to deliver part number 123456789-E that was RoHS-compliant could be rejected by the customer's receiving department as an incorrect part number per the purchase order and contract of record. The customer could also state that in order for the new part number 123456789-E to be received, the customer would have to put the contract out for re-bid, with consequent downside risks.

Penalties (fines and imprisonment)

Article 8 Penalties states: "Member States shall determine penalties applicable to breaches of the national provisions adopted pursuant to this Directive. The penalties thus provided for shall be effective, proportionate and dissuasive."[1]

As a result, the penalties set by the various EU member states are different.

Some examples of penalties that have been imposed are:

- Germany: fines up to 30,000 euros, and a maximum of 1 year imprisonment to the importer of non-RoHS compliant products.
- Denmark: no maximum fine. Imprisonment can be up to 2 years.

Note: Article 23 of the subsequent RoHS Directive 2011/65/EU also allows member states to set the rules on penalties.

Risks

Non-compliance with RoHS carries risk across several areas of the business, from lost revenue and regulatory penalties to damaged partner relationships.

Revenue

- If a product is identified as non-compliant, that product is quarantined. Quarantine is for an indefinite period of time.
- Quarantine applies to ALL countries with RoHS-like legislation worldwide after the investigation process.
- Damage to company brand name.

Penalties

- Depends on the severity of the non-compliance (event vs. systemic; deliberate versus unintentional).

Criminal

- Willful actions regarding non-compliance are subject to criminal penalties.

Recall, Repair, or Replacement Costs

- Producer must recall non-compliant product (in Ireland—by placing ads in three major national newspapers, with some embarrassment!)
- The cost of repair or replacement depends on the magnitude of the issue.

Partner or Customer Flight

- Expect compliance as a condition of business.
- Expect vendor to assume RoHS risk contractually.
- Financial risks due to indemnity clauses with distributors, partners and customers.

1. DIRECTIVE 2002/95/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 January 2003 on the restriction of the use of certain hazardous substances in electrical and electronic equipment. OJ L 37/19, 13.2.2003.
2. EU Types of Legislation. An official web site of the European Union. Unless otherwise indicated (e.g. in individual copyright notices), content owned by the EU on this website is licensed under the Creative Commons Attribution 4.0 International (CC BY 4.0) licence. This means that reuse is allowed, provided appropriate credit is given and changes are indicated.
3. Wikipedia - RoHS. Text is available under the Creative Commons Attribution-ShareAlike 4.0 License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.

This page titled 5.5: Legal Considerations is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

Terms

- **AML:** Approved Manufacturers List. A list of vetted manufacturers with whom the organization is allowed to place purchase orders.
- **AVL:** Approved Vendors List. Same as AML, but uses the word "Vendor" instead of "Manufacturer."
- **BOM:** Bill of Materials. An organized list itemizing all the parts and assemblies required to manufacture a product. If it includes the cost of each part, it is called Costed BOM.
- **ECO:** Engineering Change Order.
- **EOL:** End of Life. A part of a product life cycle for retiring a product.
- **From-To list:** A list, often maintained in spreadsheet form, that maps previously used non-compliant parts (by part number) to replacement parts that are compliant.
- **IPC:** A trade association that standardizes assembly of electronic equipment, previously known as "Institute of Printed Circuits."
- **MCO:** Manufacturing Change Order.
- **PCB:** Printed Circuit Board. A circuit board that houses components and contains routing traces to connect the components to perform its functions.
- **PPM:** Parts per million.
- **RoHS:** Restriction of Hazardous Substances.
- **SKU:** Stock Keeping Unit.
- **SQA:** System or Software Quality Assurance.
- **XRF:** X-Ray Fluorescence, a method used to detect some of the elements that are part of the list of restricted substances.

This page titled 5.6: Glossary is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

6: BOEING 737 MAX QUALITY ISSUE RECOVERY 2018

6.1: Introduction

6.2: Project Scope- Steps to Recertification

6.3: Project Management Considerations

6.4: Managing Quality

6.4.1: Introduction to Managing Quality

6.4.2: Gurus of Quality

6.4.2.1: Walter Shewhart (1891-1967)

6.4.2.2: Edwards Deming (1900-1993)

6.4.2.3: Joseph Juran (1904-2008)

6.4.2.4: Philip Crosby (1926-2001) and Armand Fiegenbaum (1920-2014)

6.4.3: Cost of Quality

6.4.4: Quality Systems

6.4.4.1: Total Quality Management (TQM)

6.4.4.2: ISO

6.4.4.3: Six Sigma

6.4.5: Tools for Quality Improvement

6.4.5.1: Check Sheets

6.4.5.2: Histograms

6.4.5.3: Control Charts

6.4.5.4: Pareto Charts

6.4.5.5: Scatter Diagrams

6.4.5.6: Cause and Effect Diagrams

6.4.6: References

6.5: Legal Considerations

6.6: Glossary

This page titled 6: Boeing 737 Max Quality Issue Recovery 2018 is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

6.1: Introduction

This case study covers two crashes of Boeing 737 MAX 8 aircraft in late 2018 and early 2019, the subsequent grounding of all Boeing 737 MAX 8 aircraft, the analysis that was done to determine the root causes of the crashes, and recommendations for changes in order to re-certify the aircraft.

The case study will handle this as a quality issue with the project being the work required to attain FAA recertification.

This chapter will:

- Describe Boeing 737 model history
- Describe the circumstances of the two crashes
- Examine the National Transportation Safety Board (NTSB) recommendation to the the FAA
- Examine Federal Aviation Administration (FAA) actions
- Explore Congressional hearing material
- Explore project management considerations, especially quality analysis techniques
- Touch on legal considerations

Short History of the Boeing 737

Boeing's 737 aircraft had been a dependable and successful aircraft since its development 50 years earlier. At the time of Boeing's announcement to develop the new 737 MAX, it had delivered more than **9,000** of the 737 aircraft, in their various derivative models, since the first 737 had come off Boeing's assembly line in 1967.[1, p40]

Producing a brand new aircraft would have been technically exciting and challenging, but the costs and time involved in doing so made Boeing's executives anxious that they would lose out in competition to Airbus. After the decision was made to update the 737 and replace it with the new 737 MAX aircraft, an aggressive production schedule and cost-conscious measures targeting competing with Airbus played a major role in the program.

On March 8, 2017, the FAA granted an amended type certificate to Boeing for the 737-8 aircraft, the first of the 737 MAX family. The MAX is the 4th generation 737 model airplane and is the successor to the company's 737 Next Generation (NG) family of aircraft. The 737 MAX was the 12th derivative model of the 737 aircraft, which was first certified half a century earlier in 1967.[1, p5]

In May 2017, the 737 MAX first entered revenue passenger service with Malindo Air, a Malaysian air carrier, two months after its FAA certification.[1, p5] Seventeen months later on October 29, 2018 the 737 MAX suffered its first fatal crash.[1,p5] The second crash occurred on March 10, 2019.

On March 12, 2019 international agencies began grounding the 737 MAX, followed by grounding by US authorities on March 13, 2019.

Flight Data Recorder

A **flight recorder** is an electronic recording device placed in an aircraft for the purpose of facilitating the investigation of aviation accidents and incidents. The device may be referred to colloquially as a "**black box**", an outdated name which has become a misnomer because they are required to be painted bright orange, to aid in their recovery after accidents.[7]

There are two types of flight recording devices: the **flight data recorder** (**FDR**) preserves the recent history of the flight by recording dozens of parameters collected several times per second; the **cockpit voice recorder** (**CVR**) preserves the recent history of the sounds in the cockpit, including the conversation of the pilots. The two devices may be combined into a single unit. Together, the FDR and CVR document the aircraft's flight history, which may assist in any later investigation.[7]



Figure 6.1.1: Cockpit voice recorder and flight data recorder, each with an underwater locator beacon on the front. (Public domain; By Dennis R. Grossi[7])

Two orange rectangular boxes each with an underwater locator beacon on the front.

Circumstances of the aircraft accidents

Note

This section pulls material from NTSB Safety Recommendation Report ASR-19-01[3].

Lion Air Flight 610

On October 29, 2018, Lion Air flight 610, a Boeing 737 MAX 8, PK-LQP, crashed in the Java Sea shortly after takeoff from Soekarno-Hatta International Airport, Jakarta, Indonesia. The flight crew had communicated with air traffic control and indicated that they were having flight control and altitude issues before the airplane disappeared from radar. The flight was a scheduled domestic flight from Jakarta to Depati Amir Airport, Pangkal Pinang City, Bangka Belitung Islands Province, Indonesia. All 189 passengers and crew on board died, and the airplane was destroyed. The National Transportation Safety Committee of Indonesia took the lead on the investigation.

The airplane's digital flight data recorder (DFDR) recorded a difference between the left and right angle of attack (AOA) sensors that was present during the entire accident flight; the left AOA sensor was indicating about 20° higher than the right AOA sensor. During rotation, the left (captain's) stick shaker activated, and DFDR data showed that the left airspeed and altitude values disagreed with, and were lower than, the corresponding values from the right. The first officer asked a controller to confirm the altitude of the airplane and later also asked the speed as shown on the controller radar display. After the flaps were fully retracted, a 10-second automatic aircraft nose-down (AND) stabilizer trim input occurred. After the automatic AND stabilizer trim input, the flight crew used the stabilizer trim switches (located on the outboard side of each control wheel) and applied aircraft nose-up (ANU) electric trim. According to DFDR data, about 5 seconds after the completion of the pilot trim input, another automatic AND stabilizer trim input occurred. The crew applied ANU electric trim again. DFDR data then showed that the flaps were extended for almost 2 minutes. However, the flaps were then fully retracted, and automatic AND stabilizer trim inputs occurred more than 20 times over the next 6 minutes; the crew countered each input during this time using ANU electric trim. The last few automatic AND stabilizer trim inputs were not fully countered by the crew.

During the preceding Lion Air flight on the accident airplane with a different flight crew, the DFDR recorded the same difference between left and right AOA of about 20° that continued until the end of the recording. During rotation, the left control column stick shaker activated and continued for the entire flight, and DFDR data showed that the left airspeed and altitude values disagreed with, and were lower than, the corresponding values from the right. After the flaps were fully retracted, a 10-second automatic AND stabilizer trim input occurred, and the crew countered the input with an ANU electric trim input. After several automatic AND stabilizer trim inputs that were countered by pilot-commanded ANU electric trim inputs, the crew noticed that the airplane was automatically trimming AND. The captain moved the stabilizer trim cutout (STAB TRIM CUTOUT) switches to CUTOUT. He then moved them back to NORMAL, and the problem almost immediately reappeared. He moved the switches back to CUTOUT. He stated that the crew performed three non-normal checklists: Airspeed Unreliable, ALT DISAGREE (altitude disagree), and Runaway Stabilizer. The pilots continued the flight using manual trim until the end of the flight. Upon landing, the captain informed an engineer of IAS DISAGREE (indicated airspeed disagree) and ALT DISAGREE alerts, in addition to FEEL DIFF PRESS (feel

differential pressure) light problems on the airplane.

On November 5, 2018, after evidence emerged of a potential contributor to the accident, the FAA conducted a preliminary risk assessment using the continued operation safety process established in FAA Order 8110.107A, Monitor Safety/Analyze Data (MSAD). Based on this risk assessment, the FAA determined urgent mandatory action was needed.[6,p6]

On November 6, 2018, the FAA issued a Continued Airworthiness Notification to the International Community (CANIC) (see Appendix D) to notify CAAs around the world that the FAA intended to issue an Emergency Airworthiness Directive (AD) to correct an unsafe condition on all 737 MAX aircraft. On November 7, 2018, the FAA issued Emergency AD 2018-23-51 (see Appendix C) requiring flight crews to use a revised runaway stabilizer operational procedure if they encountered certain conditions. The Emergency AD was an interim action, and further action was planned based on what the FAA and Boeing learned from investigating the JT610 accident.[6,p6]

Ethiopian Air Flight 302

On March 10, 2019, Ethiopian Airlines flight 302, a Boeing 737 MAX 8, Ethiopian registration ET-AVJ, crashed near Ejere, Ethiopia, shortly after takeoff from Addis Ababa Bole International Airport, Ethiopia. The flight was a scheduled international passenger flight from Addis Ababa to Jomo Kenyatta International Airport, Nairobi, Kenya. All 157 passengers and crew on board died, and the airplane was destroyed. The investigation is being led by the Ethiopia Accident Investigation Bureau.

The airplane's DFDR data indicated that shortly after liftoff, the left (captain's) AOA sensor data increased rapidly to 74.5° and was 59.2° higher than the right AOA sensor; the captain's stick shaker activated. Concurrently, the airspeed and altitude values on the left side disagreed with, and were lower than, the corresponding values on the right side; in addition, DFDR data indicated a Master Caution alert. Similar to the Lion Air accident flight, a 9-second automatic AND stabilizer trim input occurred after flaps were retracted and while in manual flight (no autopilot). About 3 seconds after the AND stabilizer motion ended, using the stabilizer trim switches, the captain, who was the pilot flying, partially countered the AND stabilizer input by applying ANU electric trim. About 5 seconds after the completion of pilot trim input, another automatic AND stabilizer trim input occurred. The captain applied ANU electric trim and fully countered the second automatic AND stabilizer input; however, the airplane was not returned to a fully trimmed condition. Cockpit voice recorder data indicated that the flight crew then discussed the STAB TRIM CUTOUT switches, and shortly thereafter DFDR data were consistent with the STAB TRIM CUTOUT switches being moved to CUTOUT.

However, because the airplane remained in a nose-down out-of-trim condition, the crew was required to continue applying nose-up force to the control column to maintain level flight. About 32 seconds before impact, two momentary pilot-commanded electric ANU trim inputs and corresponding stabilizer movement were recorded, consistent with the STAB TRIM CUTOUT switches no longer being in CUTOUT. Five seconds after these short electric trim inputs, another automatic AND stabilizer trim input occurred, and the airplane began pitching nose down.

Partial data (e.g., altitude, airspeed and vertical speed) was available from a land-based Automatic Dependent Surveillance—Broadcast (ADS-B) system. On March 11, 2019, the FAA issued another CANIC (see Appendix D) indicating the FAA was actively involved in supporting the investigations. On March 11, the FAA received additional ADS-B data. The data was analyzed and reviewed on March 12 and through the early hours of March 13. Also on March 13, physical evidence from the crash site indicated that the aircraft was in a configuration that would have armed MCAS. The satellite ADS-B data from ET302, when compared with the Lion Air Flight Data Recorder (FDR) and physical data, indicated a potential causal link between the accidents. On March 13, with the potential relationship established between the two accidents, the FAA issued an Emergency Order of Prohibition (see Appendix A) grounding 737 MAX aircraft.[6,p6-7]

A common component in both of the airplane accidents was the new flight control feature: MCAS.[1,p11]

Definition: MCAS

Maneuvering Characteristics Augmentation System

Boeing developed MCAS to address **stability issues** in certain flight conditions induced by the plane's new, larger engines, and their relative placement on the 737 MAX aircraft compared to the position of the engines on the 737 NG.[1, p12]

The specifics of this are that the CFM International LEAP engine on the 737 MAX was larger and mounted further forward from the wing and higher off the ground than on previous generations of the 737. Boeing discovered that the aircraft engine had a

tendency to push the aircraft's nose up when operating under certain conditions (flaps up, high angle of attack, manual flight). MCAS was intended to mimic the flight behavior of the previous Boeing 737 Next Generation (NG). Boeing indicated that this change eliminated the need for pilots to have simulator training on the 737 Max 8.[5]

MCAS was implemented to compensate for an excessive (high) angle of attack measured by an Angle of Attack (AOA) sensor by adjusting the elevators on the horizontal stabilizer near the airplane's tail before the aircraft would potentially stall. [5] The adjustment to the elevators should lift the tail and point the nose of the plane down.

Original Design Certification of the 737 Max 8

Note

This section pulls material from NTSB Safety Recommendation Report ASR-19-01[3].

The 737 MAX 8 is a derivative of the 737-800 Next Generation (NG) model and is part of the 737 MAX family (737 MAX 7, 8, and 9).5 The 737 MAX incorporated the CFM LEAP-1B engine, which has a larger fan diameter and redesigned engine nacelle compared to engines installed on the 737 NG family. During the preliminary design stage of the 737 MAX, Boeing's testing and analysis revealed that the addition of the LEAP-1B engine and associated nacelle changes produced an ANU pitching moment when the airplane was operating at high AOA and mid Mach numbers. After studying various options for addressing this issue, Boeing implemented aerodynamic changes as well as a stability augmentation function, MCAS, as an extension of the existing speed trim system to improve aircraft handling characteristics and decrease pitch-up tendency at elevated AOA. As the development of the 737 MAX progressed, the MCAS function was expanded to low Mach numbers.[3, pp3-4]

Definition: Nacelle

A nacelle is an enclosed structure on an aircraft for an engine.

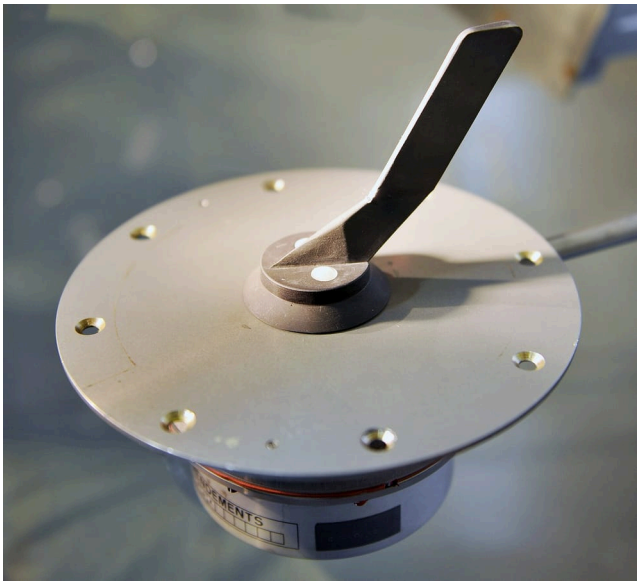


Figure 6.1.2: Angle of Attack Sensor. (Hannes Grobe - Own work, CC BY-SA 3.0[8])

An Angle of Attack (AOA) sensor adjusts itself to align with the direction of airflow sensed at the nose of the airplane. It reflects the angle of the airplane's wing relative to that airflow.

Both the 737 NG and the 737 MAX include a Speed Trim System that automatically moves the horizontal stabilizer in response to changes from trimmed airspeed while in manual flight for speed stability. The 737 MAX includes an additional pitch augmentation flight control law called MCAS, which affects pitch characteristics in manual flight, with flaps up and at elevated angles of attack,

to compensate for the aerodynamic pitch effects associated with the 737 MAX's larger, more forward located engines. MCAS enables the 737 MAX to comply with FAA regulatory requirements for certain handling qualities and cannot be removed unless replaced with a comparable system. Boeing has not presented a comparable system so the need for MCAS remains.[6]

The MCAS flight-control law is part of the Speed Trim System and is commanded by the FCC using data from sensors and other airplane systems. As a stability-augmentation system, MCAS automatically makes commands without pilot action. Either pilot can override MCAS commands to control the stabilizer by using the electric stabilizer trim switches on the control wheel. Additionally, flightcrews can set the STAB TRIM CUTOUT switches on the center control stand to CUTOUT to stop and disable MCAS.[6]

The MCAS flight-control law becomes active when the airplane exceeds an AOA threshold that varies depending on Mach (airspeed). If activated by a high AOA, MCAS moves the horizontal stabilizer at a rate of 0.27 degrees per second, which is the same trim rate as Speed Trim with flaps down. The magnitude of the MCAS command is a function of Mach and angle of attack. At higher airspeeds, flight-control surfaces are more effective than at lower airspeeds. Therefore, a smaller MCAS command at higher airspeeds has the same effect as a larger MCAS command at lower speeds. At low Mach, when the stabilizer has lower effectiveness, the MCAS command can be up to the maximum of 2.5 degrees of incremental stabilizer movement. The MCAS flight control law is reset after the AOA falls sufficiently below the AOA threshold.[6]

As originally delivered, the MCAS became active during manual flight (autopilot not engaged) when the flaps were fully retracted and the airplane's AOA value (as measured by either AOA sensor) exceeded a threshold based on Mach number. When activated, the MCAS provided automatic trim commands to move the stabilizer AND. Once the AOA fell below the threshold, the MCAS would move the stabilizer ANU to the original position. At any time, the stabilizer inputs could be stopped or reversed by the pilots using their stabilizer trim switches. If the stabilizer trim switches were used by the pilots and the elevated AOA condition persisted, the MCAS would command another stabilizer AND trim input after 5 seconds.[3,p4]

The NTSB reviewed sections of Boeing's system safety analysis for stabilizer trim control that pertained to MCAS on the 737 MAX. Boeing's analysis included a summary of the functional hazard assessment findings for the 737 MAX stabilizer trim control system. For the normal flight envelope, Boeing identified and classified two hazards associated with "uncommanded MCAS" activation as "major." One of these hazards, applicable to the MCAS function seen in these accidents, included uncommanded MCAS operation to maximum authority.10 Boeing indicated that, as part of the functional hazard assessment development, pilot assessments of MCAS-related hazards were conducted in an engineering flight simulator, including the uncommanded MCAS operation (stabilizer runaway) to the MCAS maximum authority.[3,p5]

To perform these simulator tests, Boeing induced a stabilizer trim input that would simulate the stabilizer moving at a rate and duration consistent with the MCAS function. Using this method to induce the hazard resulted in the following: motion of the stabilizer trim wheel, increased column forces, and indication that the airplane was moving nose down. Boeing indicated to the NTSB that this evaluation was focused on the pilot response to uncommanded MCAS operation, regardless of underlying cause. Thus, the specific failure modes that could lead to uncommanded MCAS activation (such as an erroneous high AOA input to the MCAS) were not simulated as part of these functional hazard assessment validation tests. As a result, additional flight deck effects (such as IAS DISAGREE and ALT DISAGREE alerts and stick shaker activation) resulting from the same underlying failure (for example, erroneous AOA) were not simulated and were not in the stabilizer trim safety assessment report reviewed by the NTSB.[3,p5]

MCAS Training and Documentation

"Boeing ultimately won the FAA's approval to give pilots just an **hour of training through an iPad** about the differences between the MAX and the previous 737 generation. MCAS was not mentioned."[2]

"The FAA, after internal deliberations, also agreed to **keep MCAS out of the manual**, reasoning that MCAS was a software code that operates in the background as part of the flight-control system..."[2]

NTSB Analysis: Assumptions about Pilot Recognition and Response in the Safety Assessment

On the Lion Air flight immediately before the accident flight and the Lion Air and Ethiopian Airlines accident flights, the DFDR recorded higher AOA sensor data on the left side than on the right (about 20° higher on the previous Lion Air flight and the Lion Air accident flight and about 59° higher on the Ethiopian Airlines accident flight). As previously stated, the MCAS becomes active when the airplane's AOA exceeds a certain threshold. Thus, these erroneous AOA sensor inputs resulted in the MCAS activating on

the accident flights and providing the automatic AND stabilizer trim inputs. The erroneous high AOA sensor input that caused the MCAS activation also caused several other alerts and indications for the flight crews. The stick shaker activated on both accident flights and the previous Lion Air flight. In addition, IAS DISAGREE and ALT DISAGREE alerts occurred on all three flights. Also, the Ethiopian Airlines flight crew received a Master Caution alert. Further, after the flaps were fully retracted, the unintended AND stabilizer inputs required the pilots to apply additional force to the columns to maintain the airplane's climb attitude.[3,p7]

Definition: AND

AND = Automatic Aircraft Nose Down

Multiple alerts and indications can increase pilots' workload, and the combination of the alerts and indications did not trigger the accident pilots to immediately perform the runaway stabilizer procedure during the initial automatic AND stabilizer trim input. In all three flights, the pilot responses differed and did not match the assumptions of pilot responses to unintended MCAS operation on which Boeing based its hazard classifications within the safety assessment and that the FAA approved and used to ensure the design safely accommodates failures. Although a number of factors, including system design, training, operation, and the pilots' previous experiences, can affect a human's ability to recognize and take immediate, appropriate corrective actions for failure conditions, industry experts generally recognize that an aircraft system should be designed such that the consequences of any human error are limited.[3,p7]

Neither Boeing's system safety assessment nor its simulator tests evaluated how the combined effect of alerts and indications might impact pilots' recognition of which procedure(s) to prioritize in responding to an unintended MCAS operation caused by an erroneous AOA input.[3,p8]

Thus, the NTSB concluded that the assumptions that Boeing used in its functional hazard assessment of uncommanded MCAS function for the 737 MAX did not adequately consider and account for the impact that multiple flight deck alerts and indications could have on pilots' responses to the hazard. Therefore, the NTSB recommended that the FAA require that Boeing:

1. ensure that system safety assessments for the 737 MAX in which it assumed immediate and appropriate pilot corrective actions in response to uncommanded flight control inputs, from systems such as MCAS, consider the effect of all possible flight deck alerts and indications on pilot recognition and response; and
2. incorporate design enhancements (including flight deck alerts and indications), pilot procedures, and/or training requirements, where needed, to minimize the potential for and safety impact of pilot actions that are inconsistent with manufacturer assumptions.[3,p8]

NTSB Analysis: System Diagnostic Tools

Multiple alerts and indications in the cockpit can increase pilots' workload and can also make it more difficult to identify which procedure the pilots should conduct.[3, p10]

Human factors research has identified that, for non-normal conditions, such as those involving a system failure with multiple alerts, where there may be multiple flight crew actions required, providing pilots with understanding as to which actions must take priority is a critical need. This is particularly true in the case of functions implemented across multiple airplane systems because a failure in one system within highly integrated system architectures can present multiple alerts and indications to the flight crew as each interfacing system registers the failure. For example, the erroneous AOA output experienced during the two accident flights resulted in multiple alerts and indications to the flight crews, yet the crews lacked tools to identify the most effective response. Thus, it is important that system interactions and the flight deck interface be designed to help direct pilots to the highest priority action(s).[3, pp10-11]

Research demonstrates that emergency situations increase workload and require additional effort to manage effectively because of the stress involved and the lack of opportunity for pilots to practice these skills compared to those used in normal operations. In addition, research into pilot responses to multiple/simultaneous anomalous situations, along with data from accidents, indicates that multiple competing alerts may exceed available mental resources and narrow attentional focus leading to delayed or inadequately prioritized responses. According to FAA research, "in some airplanes, the complexity and variety of ancillary warnings and alerts associated with major system failures can make it difficult for the flightcrew to discern the primary failure." The researchers noted that better system failure diagnostic tools are needed to resolve this issue.[3, p11]

Thus, the NTSB concludes that aircraft systems that can more clearly and concisely inform pilots of the highest priority actions when multiple flight deck alerts and indications are present would minimize confusion and help pilots respond most effectively. Therefore, the NTSB recommends that the FAA develop design standards, with the input of industry and human factors experts, for aircraft system diagnostic tools that improve the prioritization and clarity of failure indications (direct and indirect) presented to pilots to improve the timeliness and effectiveness of their response. The NTSB further recommends that once the design standards have been developed as recommended in Recommendation A-19-15, the FAA require implementation of system diagnostic tools on transport-category aircraft to improve the timeliness and effectiveness of pilots' response when multiple flight deck alerts and indications are present.[3, p11]

NTSB Recommendations to the Federal Aviation Administration (FAA)[3, pp12-13]

Require that Boeing (1) ensure that system safety assessments for the 737 MAX in which it assumed immediate and appropriate pilot corrective actions in response to uncommanded flight control inputs, from systems such as the Maneuvering Characteristics Augmentation System, consider the effect of all possible flight deck alerts and indications on pilot recognition and response; and (2) incorporate design enhancements (including flight deck alerts and indications), pilot procedures, and/or training requirements, where needed, to minimize the potential for and safety impact of pilot actions that are inconsistent with manufacturer assumptions. (A-19-10)

Require that for all other US type-certificated transport-category airplanes, manufacturers (1) ensure that system safety assessments for which they assumed immediate and appropriate pilot corrective actions in response to uncommanded flight control inputs consider the effect of all possible flight deck alerts and indications on pilot recognition and response; and (2) incorporate design enhancements (including flight deck alerts and indications), pilot procedures, and/or training requirements, where needed, to minimize the potential for and safety impact of pilot actions that are inconsistent with manufacturer assumptions. (A-19-11)

Notify other international regulators that certify transport-category airplane type designs (for example, the European Union Aviation Safety Agency, Transport Canada, the National Civil Aviation Agency-Brazil, the Civil Aviation Administration of China, and the Russian Federal Air Transport Agency) of Recommendation A-19-11 and encourage them to evaluate its relevance to their processes and address any changes, if applicable. (A-19-12)

Develop robust tools and methods, with the input of industry and human factors experts, for use in validating assumptions about pilot recognition and response to safety-significant failure conditions as part of the design certification process. (A-19-13)

Once the tools and methods have been developed as recommended in Recommendation A-19-13, revise existing Federal Aviation Administration (FAA) regulations and guidance to incorporate their use and documentation as part of the design certification process, including re-examining the validity of pilot recognition and response assumptions permitted in existing FAA guidance. (A-19-14)

Develop design standards, with the input of industry and human factors experts, for aircraft system diagnostic tools that improve the prioritization and clarity of failure indications (direct and indirect) presented to pilots to improve the timeliness and effectiveness of their response. (A-19-15)

Once the design standards have been developed as recommended in Recommendation A-19-15, require implementation of system diagnostic tools on transport-category aircraft to improve the timeliness and effectiveness of pilots' response when multiple flight deck alerts and indications are present. (A-19-16)

References

1. Majority Staff of the Committee on Transportation and Infrastructure, Final Committee Report, The Design, Development & Certification of the Boeing 737 Max, September, 2020. Final Committee Report Boeing 737 Max.
2. Dominic Gates and Mike Baker, "The inside story of MCAS: How Boeing's 737 MAX system gained power and lost safeguards", Seattle Times, June 22, 2019 and updated June 24, 2019. Seattle Times MCAS. Accessed March 27, 2026.
3. NTSB Safety Report. National Transportation Safety Board. Safety Recommendation Report ASR-19-01. "Assumptions Used in the Safety Assessment Process and the Effects of Multiple Alerts and Indications on Pilot Performance", September 19, 2019. Accessed March 24, 2026.Official Report of the Special Committee to review the Federal Aviation Administration's Aircraft Certification Process, January 16, 2020.

4. Wikipedia - MCAS. Accessed April 2, 2026. Text is available under the Creative Commons Attribution-ShareAlike 4.0 License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.
5. FAA, Summary of the FAA's Review of the Boeing 737 MAX, Return to Service of the Boeing 737 MAX Aircraft, November 18, 2020. FAA Review of Boeing 737 Max. Accessed March 26, 2026.
6. Wikipedia - Flight Recorder. Accessed April 6, 2026. Text is available under the Creative Commons Attribution-ShareAlike 4.0 License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.
7. Dennis R. Grossi. Public Domain, March 15, 2009. Image of Recorders.
8. Hannes Grobe - Own work, CC BY-SA 3.0, AOA Sensor image.

This page titled 6.1: Introduction is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

6.2: Project Scope- Steps to Recertification

The initial project scope might be based on early accident information as it became available and on the NTSB Safety Recommendations to the FAA.[1]

In order to achieve recertification of the aircraft, it would be necessary to implement the recommendations made by the FAA. Let's take a look at those recommendations by looking at the preliminary report published August 3, 2020[2] and the final report on November 18, 2020.[3]

The FAA took into account the final recommendations of the U.S. National Transportation Safety Board (NTSB) and Indonesia's Komite Nasional Keselamatan Transportasi (KNKT), as well as all available preliminary information from the accident investigations.[3.p5]

The FAA and other U.S. bodies commissioned a number of studies to evaluate the initial certification of the 737 MAX, as well as the certification process itself. The FAA invited nine other CAAs to be a part of the Joint Authorities Technical Review (JATR), which assessed the certification of the 737 MAX flight control system. The FAA took into consideration all relevant findings of the JATR, as well as those of the U.S. Department of Transportation Special Committee, the U.S. Department of Transportation Office of the Inspector General (OIG), and other expert reviews.[3.p5]

Even before the formal FAA recommendations, Boeing was suggesting some MCAS changes. On March 11, 2019, after China had grounded the aircraft, Boeing published some details of new system requirements for the MCAS software and for the cockpit displays, which it began implementing in the wake of the prior accident five months earlier:

- If the two AoA sensors disagree with the flaps retracted, MCAS will not activate and an indicator will alert the pilots.
- If MCAS is activated in non-normal conditions, it will only "provide one input for each elevated AoA event."
- Flight crew will be able to counteract MCAS by pulling back on the column.[4]

Summary of FAA Recommendations

The FAA reviewed the contributing factors and available data from both accidents and reviewed the existing MCAS design to identify issues and determine the actions needed to safely operate the 737 MAX. The FAA determined that the aircraft, flightcrew, and maintenance issues summarized in Figure $6.2.1$, Figure $6.2.2$, and Figure $6.2.3$ needed to be addressed along with a summary of the required corrective actions.[3,p7]

The FAA essentially approved design changes for each MAX aircraft, which would prevent MCAS activation unless both AoA sensors register similar readings, eliminate MCAS's ability to repeatedly activate, and allow pilots to override the system if necessary. The FAA began requiring all MAX pilots to undergo MCAS-related training in flight simulators by 2021.[4]

Identified Issue	FAA Determination of Issues That Must Be Addressed	Corrective Action*
Safety Item #1: USE OF SINGLE ANGLE OF ATTACK (AOA) SENSOR: Erroneous data from a single AOA sensor activated MCAS and subsequently caused airplane nose-down trim of the horizontal stabilizer.	Ensure that an erroneous signal from a failed single AOA sensor meets all FAA requirements, does not prevent continued safe flight and landing, and specifically that it does not generate erroneous MCAS activation.	Boeing updated the Flight Control Computer (FCC) software to eliminate MCAS reliance on a single AOA sensor signal by using both AOA sensor inputs and changing flight control laws to safeguard against MCAS activation due to a failed or erroneous AOA sensor.
Safety Item #2: MCAS RESET GENERATES REPETITIVE MCAS COMMANDS: When a continuous erroneous high AOA sensor value exists, the MCAS control law uses pilot release of the electric trim switch to reset MCAS activation. Once reset, the MCAS system will make another airplane nose-down stabilizer trim command after five seconds. This scenario repeats each time MCAS makes a command and the pilot makes an electric trim command of any duration and releases the trim switch.	Ensure that if MCAS is activated, it will generate a single nose-down command activation.	Boeing changed flight control laws to ensure that MCAS will not command repeated movements of the horizontal stabilizer. The revised flight control laws permit only one activation of MCAS per sensed high-AOA event. Any subsequent activation of MCAS would only be possible after the airplane returns to a low-AOA state.

Figure 6.2.1: Issues to be Addressed and Corrective Actions. (Summary of the FAA's Review of the Boeing 737 MAX, Return to Service of the Boeing 737 MAX Aircraft, November 18, 2020.)[3,p7]

Identified Issue	FAA Determination of Issues That Must Be Addressed	Corrective Action*
Safety Item #3: MCAS TRIM AUTHORITY: All MCAS commands were incremental commands, which moved the horizontal stabilizer a fixed amount, regardless of the current position of the stabilizer. Therefore, multiple MCAS commands resulted in a significant horizontal stabilizer mistrim condition, which the flightcrew could not counter using only elevator control.	Ensure that if MCAS is erroneously activated, the MCAS system preserves the flightcrew's ability, using basic piloting techniques, to control the airplane after the activation.	Boeing changed flight control laws to include a limit for MCAS commands. The MCAS will stop commanding stabilizer movement at a point that preserves enough elevator movement for sufficient pilot control of aircraft pitch attitude for current operating conditions.
Safety Item #4: FLIGHTCREW RECOGNITION AND RESPONSE: FDR data from both accidents show that the flightcrews were unable to effectively manage the stabilizer movement and multiple flight deck effects that resulted from the single AOA sensor failure.	Ensure that after any foreseeable failure of the stabilizer system, safe flight is not dependent on the timeliness of the flightcrew performing a non-normal procedure.	In addition to the software changes noted in Safety Items #1, #2 and #3, Boeing revised or added eight non-normal flightcrew procedures to the Airplane Flight Manual and proposed additional training. The flightcrew procedures and the revised pilot training provide the pilot additional information to recognize and respond to erroneous stabilizer movement and the effects of potential AOA sensor failures.
Safety Item #5: AOA DISAGREE: The AOA DISAGREE alert message on the Primary Flight Display is not functional unless the AOA indicator option was chosen by the airline. This alert message is intended to be standard on all 737 MAX airplanes.	Ensure 737 MAX Display System (MDS) software will alert the flightcrew with the AOA DISAGREE message if there is disagreement between the Left and Right AOA Sensors.	Boeing has revised the AOA DISAGREE alert message implementation to achieve the original design intent to be standard on all 737 MAX aircraft.
Safety Item #6: OTHER POSSIBLE FCC STABILIZER RUNAWAY FAILURES: A comprehensive review of the integrated System Safety Analysis (SSA) of MCAS by Boeing and the FAA identified an extremely remote FCC failure condition that required timely pilot intervention to ensure continued safe flight and landing.	Ensure that after any foreseeable failure of the stabilizer system, continued safe flight and landing is not dependent on the timeliness of the flightcrew performing a non-normal procedure.	Boeing implemented a cross FCC Trim Monitor, which can effectively detect and shut down erroneous stabilizer commands from the FCCs. This makes continued safe flight and landing for this type of failure not dependent on pilot reaction time.

Figure 6.2.2: Issues to be Addressed and Corrective Actions. (Summary of the FAA's Review of the Boeing 737 MAX, Return to Service of the Boeing 737 MAX Aircraft, November 18, 2020.)[3,p8]

Identified Issue	FAA Determination of Issues That Must Be Addressed	Corrective Action*
Safety Item #7: MAINTENANCE PROCEDURES RELATED TO THE MCAS: The KNKT final report for Flight JT610 shows several maintenance actions related to repair and installation of a replacement AOA sensor. "*The replacement AOA sensor that was installed on the accident aircraft had been mis-calibrated during an earlier repair. This mis-calibration was not detected during the repair. The investigation could not determine that the installation test of the AOA sensor was performed properly. The mis-calibration was not detected.*"	Revise the component maintenance instructions for the AOA sensor to ensure proper AOA sensor calibration during repair procedures. As part of the KNKT investigation of Flight JT610, the FAA evaluated existing Boeing Airplane Maintenance Manual (AMM) procedures for AOA sensor installation. The KNKT investigation (Reference section 12.3) verified these existing procedures for AOA sensor installation will correctly identify a mis-calibrated AOA sensor.	The Collins Component Maintenance Manual (CMM) for the AOA sensor was revised to include a final independent check to ensure the repair has not introduced a bias. To ensure that each airplane's two AOA sensors are functioning properly upon return to service, operators must perform AOA Sensor System Tests on each airplane prior to its return to service. This test uses a fixture to position the AOA vane and verify that the reading provided by each AOA sensor is accurate.

Figure 6.2.3: Issues to be Addressed and Corrective Actions. (Summary of the FAA's Review of the Boeing 737 MAX, Return to Service of the Boeing 737 MAX Aircraft, November 18, 2020.)[3,p8]

FAA Summary of Changes to Aircraft Design and Operation[3]

Both the 737 MAX and the 737 NG include an STS that automatically moves the horizontal stabilizer in response to changes from trimmed airspeed while in manual flight. The MCAS function is part of the STS. The STS provides increased aircraft speed stability as required by the FAA's certification regulations — namely 14 CFR 25.173, Static Longitudinal Stability — so the pilot has a positive feel (pull- or push-force) when operating the aircraft faster or slower than where it is trimmed within the normal flight envelope. A key difference between the 737 MAX and its 737 NG predecessor is the 737 MAX's larger and more powerful LEAP-1B engines. Because of their size, the engines are installed higher and farther forward than the engines on the 737 NG.[3,p9]

The 737 MAX was designed to handle and feel the same to the pilot as the 737 NG. Without the MCAS function, in some small areas of the flight envelope — such as approaching a stall and during higher g-force maneuvering — the new engines contribute to the control column feeling lighter in the 737 MAX than the regulations allow. These are not areas of the flight envelope in which the airplane normally operates. However, FAA regulations - specifically 14 CFR 25.143, 25.201, 25.203, 25.251, and 25.255 - still require the control column to have a higher pull-force feel in these flight regimes than would exist on the 737 MAX without the added stability from the STS and MCAS function.[3,p10]

For the FAA to certify the 737 MAX, the original design had to meet FAA requirements for control force feel when maneuvering or deviating significantly from trimmed equilibrium. The aircraft manufacturer is responsible for making design decisions and showing compliance to applicable regulations. In this case, Boeing elected to include an additional flight control law in the STS, which is part of the flight control software that provides required control force feel to the pilot.[3,p10]

This flight control law, or MCAS, enhances the feel of the column forces in manual flight, and is only operative with flaps up. MCAS provides signals to move the horizontal stabilizer at elevated angles of attack to compensate for the aerodynamic effects of the 737 MAX's larger and more forward-located engines, resulting in the required column feel to the pilot.[3,p10]

Duration of the MCAS command is a function of AOA and airspeed. At higher airspeeds, flight control surfaces are more effective than at lower airspeeds. Therefore, a smaller MCAS command at higher airspeed has the same effect as a larger MCAS command at low speeds. Without MCAS, the 737 MAX would not meet FAA's regulatory requirements.[3,p10]

Boeing proposed multiple updates to the MCAS function to address Safety Item #1: USE OF SINGLE ANGLE OF ATTACK (AOA) SENSOR, Safety Item #2: MCAS RESET GENERATES REPETITIVE MCAS COMMAND and Safety item #3: MCAS TRIM AUTHORITY in the previous chart. The MCAS activation software now includes a maximum limit of one nose down stabilizer activation during a single elevated AOA event and cannot be reset by pilot activation of the electric trim switches. An AOA sensor monitor was added to prevent MCAS from using an AOA input if it differs from the other AOA input by more than 5.5 degrees. Boeing incorporated a maximum command limit to disable the MCAS and speed trim operations if the stabilizer position exceeds a reference position. This limit ensures sufficient elevator control is available to provide maneuvering capability using control

column inputs alone.[3,p10]

Additional Changes Related to the Flight Control Software Update[3]

An additional software change provides cross FCC monitoring to provide additional protections against MCAS malfunctions or any other erroneous FCC-generated stabilizer trim command, and to address Safety Item #6: OTHER POSSIBLE Flight Control Computer (FCC) STABILIZER RUNAWAY FAILURES. This monitoring is designed to provide additional protection against erroneous trim commands caused by possible — but unlikely — failures, such as a fault within a single integrated circuit in the FCC.[3.p11]

The software update to both primary flight displays (MAX Display System) includes improved AOA DISAGREE annunciation logic, which addresses Safety Item #5: AOA DISAGREE alert message. The AOA DISAGREE annunciation is now always enabled, regardless of the AOA gauge option. An additional improvement locks the AOA DISAGREE annunciation when the aircraft is below 400 feet in altitude.[3,p11]

The software update to both primary flight displays (MAX Display System) includes improved AOA DISAGREE annunciation logic, which addresses Safety Item #5: AOA DISAGREE alert message. The AOA DISAGREE annunciation is now always enabled, regardless of the AOA gauge option. An additional improvement locks the AOA DISAGREE annunciation when the aircraft is below 400 feet in altitude.[3,p11]

As a result of the design work to change the MCAS behavior and the subsequent review of the Integrated Speed Trim System Safety Analysis, the FAA required an additional evaluation of the related aircraft systems for all flight phases and system modes. The FAA evaluated Boeing's determination of a non-compliance with FAA wire separation requirements. As a result, Boeing developed changes to the system, which include physical separation of existing wires and/or routing of new wires in multiple areas of the Main Electric Trim and Auto Trim system.[3.p11]

Training Enhancements

The 737 MAX Joint Operations Evaluation Board (JOEB) concluded evaluation activities at London's Gatwick Airport on September 22, 2020. The JOEB included civil aviation authorities and airline flight crews from the United States, Canada, Brazil and the European Union. The civil aviation authorities and industry pilots selected for participation completed all scheduled tasks as planned. The JOEB determined that all design changes applicable to the 737 MAX are operationally suitable. The Board's evaluations also included assessments of Boeing's proposed differences training and return to service training. The JOEB, through evaluations of multiple crews from a globally diverse pilot group, determined that the proposed training by the applicant was acceptable. The FAA posted the 737 Flight Standardization Board (FSB) report, revision 17, for public comment on October 6, 2020. The comment period closed on November 2, 2020.After considering the comments received, the FAA published the final version of the 737 FSB report, revision 17, on November 18, 2020. This revision adds training requirements for the MCAS, Autopilot Flight Director System (AFDS) enhancements, and additional Special Emphasis Training. The FAA added Appendix 7 to delineate 737 MAX ground and flight training necessary to operate the 737 MAX modified with the new FCC software.[3,p11]

FAA Conclusion Details[3]

Note

> This section is extracted from Reference 3.

Following a thorough, transparent and inclusive process, the FAA determined that Boeing's changes to the 737 MAX design, flightcrew procedures and maintenance procedures effectively mitigate the airplane-related safety issues that contributed to the Flight 610 and Flight 302 accidents. The FAA further determined that the design change addressed additional safety concerns beyond those identified during the accident investigations. This report does not address other safety issues that might have contributed to the accidents but are not related to airplane design, including maintenance, aircraft operator and air traffic control. The FAA believes recommendations related to these other potential contributing factors should be addressed by the appropriate organizations. Further, the FAA evaluated Boeing's proposed flightcrew training through the Flight Standardization Board process. The FAA issued a final Boeing 737 Flight Standardization Board Report documenting the results of the operational evaluation.[3, p76]

Safety Issue #1: Use of Single Angle of Attack (AOA) Sensor

In the original design, erroneous data from a single AOA sensor activated MCAS and subsequently caused airplane nose-down trim of the horizontal stabilizer. In the new design, Boeing eliminated MCAS reliance on a single AOA sensor signal by using both AOA sensor inputs and through flight-control law changes that include safeguards against failed or erroneous AOA indications. The updated FCC software with revised flight-control laws uses inputs from both AOA sensors to activate MCAS. This is in contrast to the original MCAS design, which relied on data from only one sensor at a time, and allowed repeated MCAS activation as a result of input from a single AOA sensor. The updated FCC software compares the inputs from the two sensors to detect a failed AOA sensor. If the difference between the AOA sensor inputs is above a calculated threshold, the FCC will disable the STS, including its MCAS function, for the remainder of that flight and provide a corresponding indication of such deactivation on the flight deck.[3,p76-77]

Safety Issue #2: MCAS Reset Generates Repetitive MCAS Commands

In the original design, when a continuous erroneous high AOA sensor value existed, the MCAS control law used pilot release of the electric trim switch to reset MCAS activation. Once reset, the MCAS system would make another airplane nose-down stabilizer trim command after five seconds. This scenario would repeat each time the MCAS made a command and the pilot made an electric trim command of any duration and released the trim switch. In the new design, Boeing included flight-control law changes to ensure that MCAS will not command repeated movements of the horizontal stabilizer. The revised flight-control laws allow only one activation of MCAS per sensed high-AOA event. A subsequent activation of MCAS is only possible after the airplane returns to a low-AOA state, below the threshold that would cause MCAS activation.[3,p77]

Safety Issue #3 MCAS Trim Authority

In the original design, all MCAS commands were incremental commands, which moved the horizontal stabilizer a fixed amount regardless of the current position of the stabilizer. Therefore, multiple MCAS commands resulted in a significant horizontal stabilizer mistrim condition, which the flightcrew could not counter using only elevator control. In the new design, Boeing included flight-control law changes that limit the magnitude of any MCAS command to move the horizontal stabilizer, so that the final horizontal stabilizer position (after the MCAS command) preserves the flightcrew's ability to control the airplane pitch by using only the control column.[3,p77]

Safety Issue #4 Flightcrew Recognition and Response

FDR data from both accidents show that the flightcrews were unable to effectively manage the stabilizer movement and multiple flight-deck effects that occurred as a result of the single AOA sensor failure. In the new design, Boeing revised eight non-normal flightcrew procedures and proposed additional training. The revised flightcrew procedures and pilot training provide the pilot with additional information to recognize erroneous stabilizer movement and the effects of AOA sensor failures.[3,p77]

> **Note**
>
> FDR is Flight Data Recorder.

Safety Issue #5 AOA DISAGREE Message

In the originally delivered configuration, the AOA DISAGREE alert message on the Primary Flight Display was not functional unless the airline chose the AOA indicator option. This alert message is intended to be standard on all 737MAX airplanes. In the new design, Boeing revised the AOA DISAGREE implementation to maintain the original design intent to be standard on all 737 MAX aircraft. The FAA is requiring an additional software update that alerts the flightcrew to a disagreement between the two AOA sensors due to a sensor failure or calibration issues. The updated software implements an AOA DISAGREE alert message on all 737 MAX airplanes. While the lack of an AOA DISAGREE alert message is not an unsafe condition itself, the FAA is mandating this software update because the flightcrew procedures now rely on this alert message to guide flightcrew action.[3,p77-78]

Safety Issue #6 Other Possible Stabilizer Runaway Failures

The FAA and Boeing conducted a comprehensive review of the integrated SSA of the MCAS function. This review identified an extremely remote failure condition that required timely pilot intervention to ensure continued safe flight and landing. In the new design, Boeing has implemented a Cross-FCC Trim Monitor, which can effectively detect and shut down erroneous stabilizer commands from the FCCs. This makes continued safe flight and landing for this type of failure not dependent on pilot reaction time.[3,p78]

Safety Issue #7 Maintenance Procedures Related to MCAS

The KNKT final report for Flight JT610 shows several maintenance actions related to repair and installation of a replacement AOA sensor. “The replacement AOA sensor that was installed on the accident aircraft had been mis-calibrated during an earlier repair. This mis-calibration was not detected during the repair. The investigation could not determine that the installation test of the AOA sensor was performed properly. The mis-calibration was not detected.” (Final KNKT.18.10.35.04 Aircraft Accident Investigation Report, page 215 linked here). The Collins Component Maintenance Manual for the AOA sensor has been revised to include a final check intended to prevent a repair shop instrumentation error. The KNKT investigation (Reference section 12.3) verified the existing Boeing AMM procedures for AOA sensor installation will correctly identify a mis-calibrated AOA sensor. To ensure that each airplane's two AOA sensors are functioning properly upon return to service, operators will perform AOA Sensor System Tests on each airplane prior to its return to service. This test uses a fixture to position the AOA vane and verify that the reading provided by each AOA sensor is accurate.[3,p78]

Additional FAA Instructions

As a result of the FAA’s review of the flight-control system of the 737 MAX, the Boeing proposed design changes, the certification documentation (including the system safety assessment and system description document) and in consideration of systems that are indirectly related to MCAS, the FAA instructed Boeing to address the following:

- Mitigate single-point failures of the horizontal stabilizer trim system, 78 Summary of the FAA’s Review of the Boeing 737 MAX.
- Improve failure monitoring of the FCC, flight-control software and FCC signal inputs,
- Improve autopilot disengagement during low-speed and/or high-AOA operation,
- Reassess the manual trim and determine the system meets FAA requirements, and
- Reassess aircraft stall characteristics with STS/MCAS and Elevator Feel Shift inoperative and determine the aircraft meets FAA requirements.

The FAA determined that the 737 MAX design changes address the unsafe condition noted in the accident investigations and enhance the safety of the airplane’s flight-control system architecture, operational procedures and maintenance procedures.[3,p78-79]

References

1. National Transportation Safety Board. Safety Recommendation Report ASR-19-01. "Assumptions Used in the Safety Assessment Process and the Effects of Multiple Alerts and Indications on Pilot Performance", September 19, 2019. Accessed March 24, 2026. NTSB Safety Report ASR-19-01.
2. FAA, Preliminary Summary of the FAA’s Review of the Boeing 737 MAX, Return to Service of the Boeing 737 MAX Aircraft, August 3, 2020. FAA Review of the Boeing 737 MAX. Accessed March 26, 2026.
3. FAA, Summary of the FAA’s Review of the Boeing 737 MAX, Return to Service of the Boeing 737 MAX Aircraft, November 18, 2020. FAA Review of the Boeing 737 MAX. Accessed March 26, 2026.
4. Wikipedia. Maneuvering Characteristics Augmentation System Wikipedia - MCAS. Accessed April 2, 2026. Text is available under the Creative Commons Attribution-ShareAlike 4.0 License(opens in new window); additional terms may apply. By using this site, you agree to the Terms of Use(opens in new window) and Privacy Policy(opens in new window). Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.

This page titled 6.2: Project Scope- Steps to Recertification is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

6.3: Project Management Considerations

Project Management Triangle at Work

For the original 737 MAX 8 development project, the main pressure point appears to have been time to market, due to competitive pressures (the fear of losing customers who might adopt a new product from a competitor, in this case, Airbus). This resulted in a trade-off against scope/quality and the allocation of resources. With time decreased, scope was decreased.

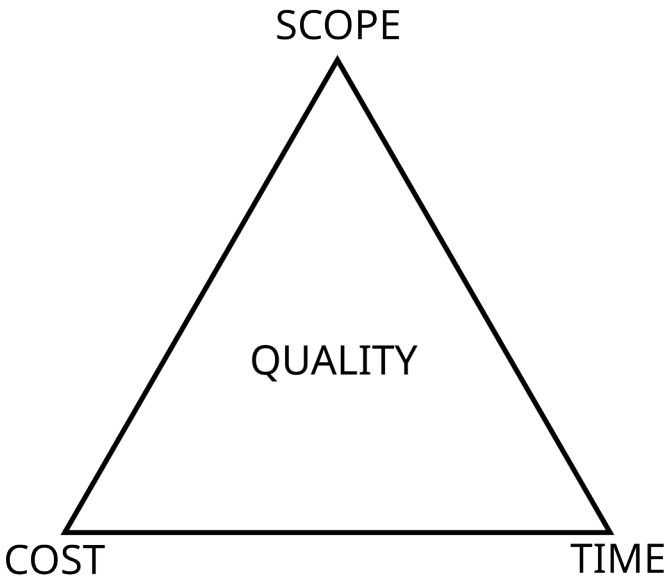


Figure 6.3.1: In project management, quality sits at the center, shaped by the balance of scope, time, and cost. (Public domain)

Accessibility Note

The project management triangle represents tradeoffs between Cost, Time, and Scope (shown at the vertices of a triangle) with the word "Quality" inside the triangle.

Corrective Actions and Recertification Effort

As previously discussed, the FAA identified various issues and corrective actions that would be requirements on project scope for implementation in order to re-certify the plane:[1]

- Use 2 sensors
- Limit MCAS to one nose-down command
- Provide pilot control of stabilizer
- Provide pilot training and procedures
- Provide a cockpit alert if AOA sensors disagree
- Modify AOA maintenance procedure

A starter list for some of the Project Management Tasks for this project might include:

Table 6.3.1: Project Management Tasks. (CC BY-NC-SA 4.0; Robert Slate CIS 79 course material)

Task	Description
Identify cause of crashes	Analyze accident reports from NTSB and FAA investigations.

Task	Description
Determine scope of corrective actions	Revise MCAS to use two sensors, limit nose-down commands, ensure pilot override capability, provide alerts for AOA sensor disagreement, update pilot training and procedures, and modify AOA sensor maintenance requirements.
Implement MCAS changes	Develop implementation schedule and conduct testing.
Implement AOA sensor disagree alert	Develop implementation schedule and conduct testing.
Pilot training	Update pilot training materials and operating procedures related to MCAS and AOA sensor disagreement alerts.
Maintenance	Revise AOA sensor inspection and maintenance procedures.
Update inventory	Modify aircraft not yet delivered to customers.
Update customer planes	Retrofit aircraft already delivered to customers.
Recertify plane	Obtain regulatory approval and recertification in the United States and other countries.

Root Cause Analysis

In order to have some assurance that the scope of work is complete, it is important to know that the root cause or causes of the accidents have been thoroughly investigated and that all contributing causes have been assessed. We have reviewed material from the NTSB and FAA for this purpose.

In the next section, we will look at how quality issues are analyzed and what types of tools are used by quality professionals such as 5 Whys, Ishikawa or fishbone diagrams, and Pareto charts.

References

1. FAA, Preliminary Summary of the FAA's Review of the Boeing 737 MAX, Return to Service of the Boeing 737 MAX Aircraft, August 3, 2020.

This page titled 6.3: Project Management Considerations is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

6.4: Managing Quality



Learning Objectives

- Understand dimensions of design quality, process quality and service quality.
- Describe the quality gurus that contributed to the modern field of quality.
- Discuss the costs of quality.
- Understand and Differentiate between TQM, ISO9001 and Six Sigma.
- Describe the tools used in quality improvement projects

Note

Remixed original OER material on 3/6/2026. Started with its "Chapter 5."

Introduction to Operations Management - Business LibreTexts.

This page titled 6.4: Managing Quality is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by LibreTexts.

6.4.1: Introduction to Managing Quality

In today's business environment, quality can be broadly defined as the extent to which a product or service meets or exceeds a customer's expectations.

Prior to the 1980s, Quality was not a main priority for North American manufacturers. As high-quality goods began flowing into North American markets from Japan, it slowly became apparent that North American companies had fallen behind. Japanese products began to be preferred by consumers who recognized their superior level of quality. This was the case especially in the automotive market.

For any product the quality of that product is determined by two primary factors. These are DESIGN quality and PROCESS quality.

Design quality can be described as the quality that a product has in terms of the actual characteristics of the product. Think about the design of your favorite cell phone. The decisions made by Marketing as well as the Design team will determine the way your phone will operate, the quality of the sound, the features it has, not to mention the way it looks, feels and lasts. Below are some facets of quality in products.[1]

Table 6.4.1.1: Design qualities (with cellphone example).

Design Quality	Description	Cellphone Example
Performance	Primary operating characteristics	Clarity of sound, speed of connection
Durability	Ability to withstand damage	A dropped phone withstands damage
Reliability	Long lasting; how long before a breakdown occurs	Several years of trouble- free performance
Features	Extra characteristics, bells, and whistles	Extra storage space, long lasting battery
Serviceability	How easy it is to fix and how willing the organization is to repair the product	Same day repairs, large network of locations
Reputation	Perceived image in the marketplace	High scores on global quality ratings
Aesthetics	The appearance of the product, feel, smell, taste	Sleek modern design, large screen

Process quality refers to the ability of the organization to produce the good or service having perfect quality at each stage of the process, or in other words, manufacturing defect-free products.

Table 6.4.1.2: Process qualities.

Element	Description
Raw Materials	Quality level of purchased inputs
Equipment	Capability, well maintained, flexibility
Employees	Experience, training, ability level
Technology	Matches the application

Measurement of **service quality** is more challenging. Each customer has a certain performance level in mind from which to compare or evaluate a service. Below are some of the commonly accepted elements in which customers evaluate service performance.

Table 6.4.1.3: Service qualities

Element	Description
Tangibles	Any physical products used during the service

Element	Description
Reliability	Dependable service, correctly done each time
Convenience	How accessible the service provider is
Responsiveness	The willingness to respond to customer requests
Time	How quickly the service is delivered
Courtesy	The politeness and friendliness of employees
Consistency	Repeated consistent performance without fail
Assurance	Employees have a high level of expertise and trust

This page titled 6.4.1: Introduction to Managing Quality is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by LibreTexts.

6.4.2: Gurus of Quality

This page titled 6.4.2: Gurus of Quality is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by LibreTexts.

6.4.2.1: Walter Shewhart (1891-1967)

Dr. Shewhart was an American physicist, engineer and statistician. He is known as the father of statistical quality control and spent much of his career researching variation and is credited with the creation of the first control chart. His work focused around the need to reduce variation in order to improve quality. He is responsible for the concepts of assignable and common variation.

The processes that produce goods and services will all have some variation. It is intuitive that the more variation in processes, the poorer the quality will be. **Assignable variation** is the type of variation where the cause can be clearly identified and corrected or managed. An example of an assignable variation might be an error by an employee, a software glitch, or a tool breakage. **Common variation**, also referred to as chance variation, is the type of variation that is inherent in the process. It is generally to be expected, and not a cause of an error. Intuitively, if we reduce or minimize either variation, we will improve product quality.

This page titled 6.4.2.1: Walter Shewhart (1891-1967) is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by LibreTexts.

6.4.2.2: Edwards Deming (1900-1993)

Dr. Deming is likely the most well known of the Quality Gurus. He was an American engineer, statistician, professor and author. Dr. Deming was recruited to Japan after WWII to assist with their national census. Beginning in 1950 Deming trained thousands of Japanese engineers, managers, and scholars in basic statistical process control. He is credited with guiding the rise of Japanese superior quality. In appreciation for Deming's guidance the Japanese named their highest quality award after Dr. Deming (The Deming Prize). Dr. Deming has an extensive list of published works but is likely most well known for **Deming's 14 points**[2] and the **Deming Cycle**.

Edward Deming's "14 Points"

1. Create constancy of purpose toward improvement of product and service, with the aim to become competitive, to stay in business and to provide jobs.
2. Adopt the new philosophy. We are in a new economic age. Western management must awaken to the challenge, must learn their responsibilities, and take on leadership for change.
3. Cease dependence on inspection to achieve quality. Eliminate the need for massive inspection by building quality into the product in the first place.
4. End the practice of awarding business on the basis of a price tag. Instead, minimize total cost. Move towards a single supplier for any one item, on a long-term relationship of loyalty and trust.
5. Improve constantly and forever the system of production and service, to improve quality and productivity, and thus constantly decrease costs.
6. Institute training on the job.
7. Institute leadership The aim of supervision should be to help people and machines and gadgets do a better job. Supervision of management is in need of overhaul, as well as supervision of production workers.
8. Drive out fear, so that everyone may work effectively for the company.
9. Break down barriers between departments. People in research, design, sales, and production must work as a team, to foresee problems of production and usage that may be encountered with the product or service.
10. Eliminate slogans, exhortations, and targets for the work force asking for zero defects and new levels of productivity. Such exhortations only create adversarial relationships, as the bulk of the causes of low quality and low productivity belong to the system and thus lie beyond the power of the work force.
    1. Eliminate work standards (quotas) on the factory floor. Substitute with leadership.
    2. Eliminate management by objective. Eliminate management by numbers and numerical goals. Instead substitute with leadership.
11. Remove barriers that rob the hourly worker of his right to pride of workmanship. The responsibility of supervisors must be changed from sheer numbers to quality.
12. Remove barriers that rob people in management and in engineering of their right to pride of workmanship.
13. Institute a vigorous program of education and self-improvement.
14. Put everybody in the company to work to accomplish the transformation. The transformation is everybody's job.

The Deming Cycle or Deming Wheel is also known as **PDCA**, or "Plan, Do, Check, Act." It is a version of continuous improvement that emphasizes the continuous nature of process improvement.

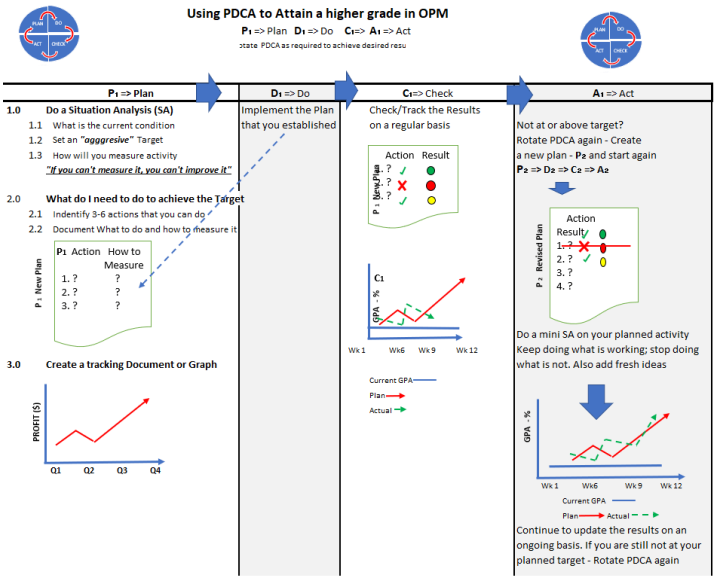


Figure 6.4.2.2.1: An example utilizing PDCA to attain a higher grade in OPM; Courtesy of Jim Philips.

Accessible format for Figure 6.4.2.2.1

This page titled 6.4.2.2: Edwards Deming (1900-1993) is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by LibreTexts.

6.4.2.3: Joseph Juran (1904-2008)

Juran was a Romanian-born American engineer. He is best known for the *Quality Control Handbook*, which was first published in 1951. He emphasized the importance of three specific factors which came to be known as the quality trilogy: quality planning, quality control and quality improvement. He authored hundreds of papers and 12 books. He is responsible for creating the concept known today the **cost of quality**. The Juran Institute in the U.S. is a leader in training and research in quality. Joseph Juran also came upon the work of Vilfredo Pareto (1848-1923) and made the Pareto Principle, also known as the 80/20 rule, well known today as a tool for problem solving and continuous improvement.

Pareto was an Italian economist and sociologist who noticed that 80% of the land in Italy was owned by about 20% of the population. This **Pareto principle** is alive and well today in the field of quality and continuous improvement. It is generally accepted that 80% of defects can be traced to a small number (20%) of the causes. Firms need to ensure that they are concentrating on fixing the correct or “root” causes.

Other examples of the 80/20 rule

- 80% of a company’s profits are generated by 20% of the products or services sold
- 80% of the continuous improvement ideas are generated by 20% of the employees

This page titled 6.4.2.3: Joseph Juran (1904-2008) is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by LibreTexts.

6.4.2.4: Philip Crosby (1926-2001) and Armand Feigenbaum (1920-2014)

Philip Crosby (1926-2001)

Crosby was an American businessman and author. He published *Quality is Free* in 1979. He believed that the costs of quality are often understated.

He coined the phrase **zero defects** and felt that there was no reason for any errors. He taught that it is less expensive to do it right the first time rather than to pay for extra inspection, scrap, rework and repairs.

Armand Feigenbaum (1920-2014)

Dr. Feigenbaum was an American quality engineer and businessman. He was the Director of Manufacturing Operations at General Electric from 1958-1968. He devised the concept of total quality control, which later became total quality management (TQM).

He is also known for his concept of a “hidden plant.” He felt that a large portion of a plant’s capacity is wasted due to the large amount of failures and defects.

Note

Changes to original OER material:

1. Accessibility fixes to headings.
2. Fixed spelling of Armand Feigenbaum.

This page titled 6.4.2.4: Philip Crosby (1926-2001) and Armand Feigenbaum (1920-2014) is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by LibreTexts.

6.4.3: Cost of Quality

Companies who provide excellent quality goods and services are obviously able to excel and differentiate themselves from competitors. They also tend to be more profitable and their losses and extra costs due to poor productivity, rework, inspection, and scrap will be negligible.

The various costs of quality can be broken down into the following four categories: **prevention costs**, **appraisal costs**, and failure costs, which are further classified as **internal failure costs** and **external failure costs**.

Prevention costs

Prevention costs include all the funds spent to prevent the occurrence of defects. Examples include quality improvement initiatives, employee training, upgrading of equipment, implementing quality procedures and making proactive design changes.

Appraisal costs

All money spent in checking and testing of product during the production process would be considered Appraisal costs. Wages of inspectors when defined as part of the process, testing labs and equipment, gauging, and process control, would be included in this category.

Internal failure costs

Once a defect has been produced, with any luck the organization will detect the error before it leaves the building and is sent to the customer. Often, defective products can be repaired, but all of the extra time spent on the rework is considered internal failure costs. Product that is unable to be repaired is classified as scrap. This also is Internal Failure costs. This can cause many other problems because customers still expect on-time deliveries. Often other orders may have to be re-manufactured and expedited in order to compensate for products that are scrap. The customer is often not aware of these issues.

External failure costs

Once a defective product has been shipped to the customer, the costs then become external failure costs. Replacement product, expedited shipping, potential law suits, product recalls, and of course loss of future business are all external failure costs. It is impossible to predict the actual external failure costs since there is no way of gauging the impact of dissatisfied customers on future business.

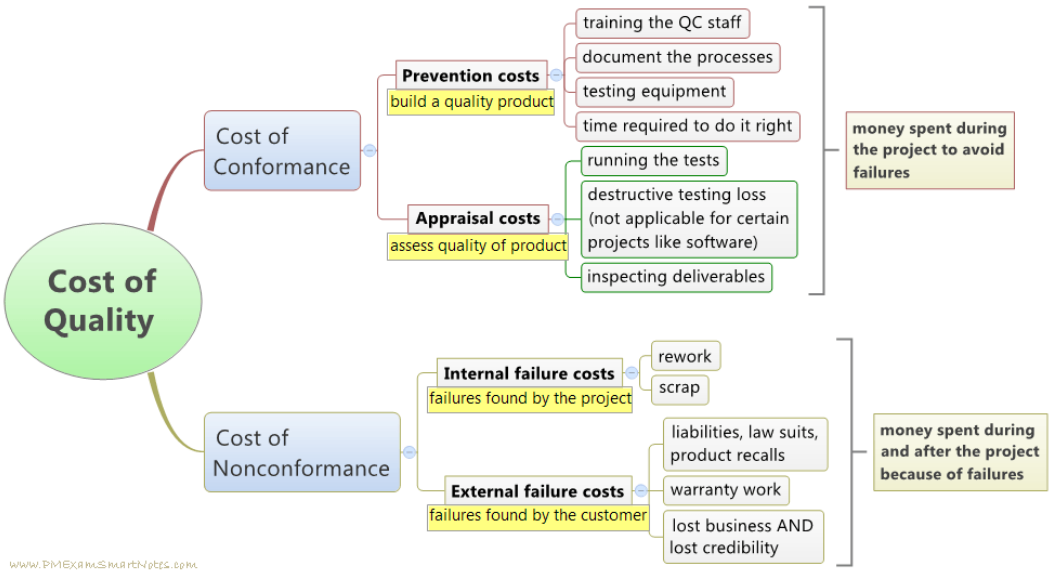


Figure 6.4.3.1: A diagram denoting the costs of conformance (prevention and appraisal costs) versus non-conformance (internal and external failure costs) in quality management; Credit: pmexamsmartnotes.com / Flickr / CC BY-NC-ND

Accessibility Note

Cost of Quality is Cost of Conformance and Nonconformance.

The Cost of Conformance has prevention costs (training the staff, documenting processes, testing equipment, time) and appraisal costs (running tests, destructive testing, inspecting deliverables). This money is spent during the project to avoid failures.

The Cost of nonconformance includes internal failure costs (rework, scrap) and external failure costs (liabilities, law suits, recalls, warranty work, lost business and lost credibility). This is money spent during and after the project because of failures.

Note

Changes made to the original OER text include:

- Added accessibility note due to length of alt text in the figure.

This page titled 6.4.3: Cost of Quality is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by LibreTexts.

6.4.4: Quality Systems

Three common quality systems that many organizations use in order to manage their quality are total quality management (TQM), ISO9001, and Six Sigma.

This page titled 6.4.4: Quality Systems is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by LibreTexts.

6.4.4.1: Total Quality Management (TQM)

To compete today, companies must deliver quality goods and services that satisfy customers' needs. This is the objective of quality management. Total quality management (TQM), or quality assurance, includes all the steps that a company takes to ensure that its goods or services *meet or exceed the customers defined specifications* and are of sufficiently high quality to meet customers' needs. Generally speaking, a company adheres to TQM principles by focusing on three tasks:

1. Customer satisfaction
2. Employee involvement
3. Continuous improvement

Let's take a closer look at these three principles.[3]

Customer Satisfaction

Companies that are committed to TQM understand that the purpose of a business is to generate a profit by satisfying customer needs. Thus, they let their customers define quality by identifying and offering those product features that satisfy customer needs. They encourage customers to tell them how to make the right products, both goods and services, that work the right way.

Armed with this knowledge, they take steps to make sure that providing quality is a factor in every facet of their operations—from design, to product planning and control, to sales and service. To get feedback on how well they're doing, many companies routinely use surveys and other methods to monitor customer satisfaction. By tracking the results of feedback over time, they can see where they need to improve.

Employee Involvement

Successful TQM requires that everyone in the organization, not simply upper-level management, commits to satisfying the customer. When customers wait too long at a drive-through window, it's the responsibility of a number of employees, not the manager alone. A defective DVD isn't solely the responsibility of the manufacturer's quality control department; it's the responsibility of every employee involved in its design, production, and even shipping. To get everyone involved in the drive for quality assurance, managers must communicate the importance of quality to subordinates and motivate them to focus on customer satisfaction. Employees have to be properly trained not only to do their jobs but also to detect and correct quality problems.

In many companies, employees who perform similar jobs work as teams, sometimes called quality circles, to identify quality, efficiency, and other work-related problems, to propose solutions, and to work with management in implementing their recommendations.

Continuous Improvement

An integral part of TQM is continuous improvement: the commitment to making constant improvements in the design, production, and delivery of goods and services. Improvements can almost always be made to increase efficiency, reduce costs, and improve customer service and satisfaction. Everyone in the organization is constantly on the lookout for ways to do things better.

This page titled 6.4.4.1: Total Quality Management (TQM) is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by LibreTexts.

6.4.4.2: ISO

The International Organization for Standardization (ISO) is an international standard-setting body composed of representatives from various standards organizations.

Founded on 23 February 1947, the organization promotes worldwide proprietary, industrial and commercial standards. It is headquartered in Geneva, Switzerland, and works in 164 countries.

ISO is an independent, non- governmental organization and is the largest developer of voluntary international standards. Use of these standards assists organization to create products and services that are safe, reliable and of good quality. The standards help businesses increase productivity while minimizing errors and waste. By enabling products from different markets to be directly compared, they facilitate companies in entering new markets and assist in the development of global trade on a fair basis. The standards also serve to safeguard consumers and the end-users of products and services, ensuring that certified products conform to the minimum standards set internationally.[4]

Steps in obtaining ISO certification

1. A firm must initially begin by documenting and implementing their systems for quality management. Add: These documents must include Policy, Procedures and work instructions. This is no small task. This documentation may take a great deal of time, effort and collaboration by all of their employees. In addition to this, the company will develop a Quality Manual that outlines how they will assure the quality of their goods and service.
2. Training of all employees involved with these processes must take place. The employees must gain full understanding and be comfortable with these new work procedures and documents
3. An audit must take place. The organization must arrange to have a certification body come to complete the audit. There are many different private companies that have been granted (by ISO) the authority to grant ISO9000 certification. These private companies are themselves accredited by ISO.

The cost for the above steps will clearly depend largely on the size and complexity of the organization. Many firms choose to use external consultants to guide and assist them in this process. Generally it is accepted that it will take somewhere from 12 to 18 months to complete this process. The usual certification process lasts for three years. Skeptics think that some organizations pursue their ISO Certification only because customers require it. In reality, a deep desire to improve internal performance should be the intrinsic motivation to implement any quality program.

There are other certifications besides ISO9001. ISO14001 is an International Standard for evaluating a firm's environmental performance.

Note

Changes made to original OER material:

- Deleted video.
- Deleted section: "What benefits are there to obtaining ISO certification?"

This page titled 6.4.4.2: ISO is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by LibreTexts.

6.4.4.3: Six Sigma

You may end up accepting a job in an organization that has a Six Sigma Program. Six Sigma is a set of techniques and tools for process improvement. It was introduced by engineer Bill Smith while working at Motorola in 1980. It is a comprehensive quality system for achieving business success by minimizing variation in business processes. Six Sigma quality is used to describe a process so well controlled that there are no more than 3.4 defects per million opportunities

Six Sigma comes from the term in statistics of sigma, meaning standard deviation. It is a measure of the amount of variation in a set of data. A usual company process might operate at a 3 sigma quality level. This indicates that 99.73% of all output will fall between plus and minus three standard deviations of the mean. The result then would be that 0.27% of all results are non-conforming, or defective. This is equivalent to 2700 defects per million opportunities. This is not so good if the business is operating in an industry where quality is critical to customer safety, such as manufacturing of aircraft, cars or medical devices. A process with a Six Sigma capability generates 99.99966% defect free outputs.

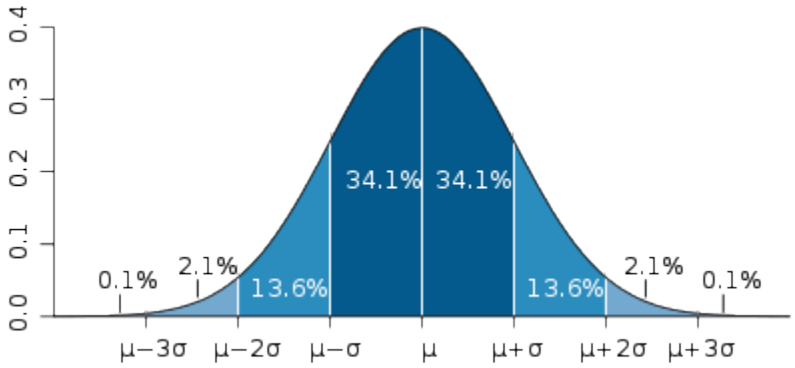


Figure 6.4.4.3.1: A diagram of a normal distribution utilizing Six Sigma; Credit: Ainali / Wikipedia / en.Wikipedia.org/wiki/File:Standard_deviation_diagram_micro.svg

Accessibility Note

Figure 6.4.4.3.1 is a chart showing a normal distribution. The center parts of the curve one sigma above and below the mean have the highest density and have 34.1% on the right hand side of the mean and 34.1% on the left hand side of the mean.

Unique to Six Sigma is the DMAIC Improvement Model. In Six Sigma, improvements are made by the use of project teams. A team may be assembled in order to fix a quality issue, to streamline an old process or to develop a new process. The team will use the **DMAIC model**, which is an acronym that stands for Define, Measure, Analyze Improve and Control. Copyright © 1993 — American Society for Quality. All rights reserved. An explanation of the steps in the DMAIC process and some of the tools used are below.[6]

1. **Define** the problem, improvement activity, opportunity for improvement, the project goals, and customer (internal and external) requirements.
2. **Measure** process performance.
3. **Analyze** the process to determine root causes of variation and poor performance (defects).
4. **Improve** process performance by addressing and eliminating the root causes.
5. **Control** the improved process and future process performance.



Figure 6.4.4.3.1: Cyclical diagram of the DMAIC model.

Also unique to Six Sigma is the use of the martial arts method of using colored belts in order to symbolize the level of expertise in Six Sigma.

- Green Belt is a designation indicating that an individual has had Six Sigma training. An individual with Green Belt status may have a full-time job in the organization but will be called upon to assist with projects part-time.
- Black Belt is a full-time position in the Quality area where the individual works as a coach / leader in projects. These individuals will have extensive experience in Six Sigma.
- Master Black Belt is a full-time position in the Six Sigma team with extensive experience in managing projects. They coach and mentor the Black Belt leaders.

Champion is usually a senior manager in the organization who oversees the other Six Sigma Quality employees. The work on determining which projects should be prioritized, ensuring proper allocation of resources and removing any roadblocks that may prevent success.

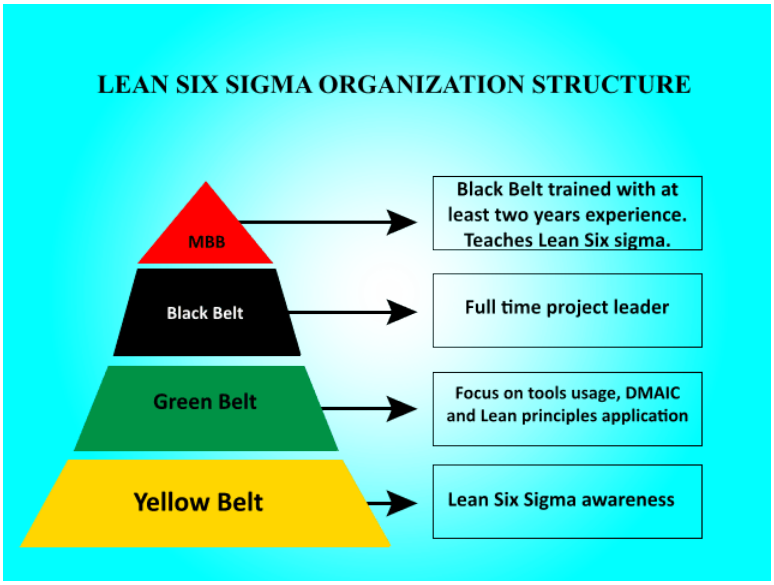


Figure 6.4.4.3.1: Variation of belt colors associated with Six Sigma; Credit: Zirguezi / Wikimedia / commons.wikimedia.org/wiki/File:Lean_Six_Sigma_Structure_Pyramid.svg

Accessibility Note

Figure 6.4.4.3.1 depicts the Lean six sigma organization structure of belt colors with Master Black Belt (MBB) at the top (Black belt trained with at least two years experience. Teaches Lean Six sigma). Black belt is next (full time project leader). Then Green Belt (Focus on tools usage, DMAIC, and Lean principles application). Finally, Yellow Belt (Lean Six Sigma awareness).

Note

Changes made to the original OER material:

- Added Accessibility Notes for two figures due to the length of the alt text.

This page titled 6.4.4.3: Six Sigma is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by LibreTexts.

6.4.5: Tools for Quality Improvement

In any Quality improvement initiative, the collection and evaluation of data is a critical step. There are a number of basic generic tools that are most commonly used. These tools include check sheets, histograms, control charts, Pareto charts, scatter diagrams, as well as cause and effect diagrams.

This page titled 6.4.5: Tools for Quality Improvement is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by LibreTexts.

6.4.5.1: Check Sheets

A Check Sheet is a custom designed form used to record the number of occurrences of a particular outcome of interest.

It may collect basic information such as how many incidents occurred, the timing, or the measurement that was non-conforming.

Motor Assembly Check Sheet

Name of Data Recorder: Lester B. Rapp

Location: Rochester, New York

Data Collection Dates: 1/17 - 1/23

Defect Types/ Event Occurrence	Dates							TOTAL
	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	
Supplied parts rusted		9	8	4	2			20
Misaligned weld			3			2		5
Improper test procedure								0
Wrong part issued		1		2				3
Film on parts								0
Voids in casting				4	2			6
Incorrect dimensions						2		2
Adhesive failure								0
Masking insufficient					1			1
Spray failure			5					5
TOTAL		10	13	10	5	4		

Figure 6.4.5.1.1: Example of a Check Sheet. (in original OER material)

Accessibility Note

Motor Assembly Check Sheet with name of data recorder, location, and data collection dates. On the left side, the Defect Types/Event Occurrences are listed (supplied parts rusted, misaligned weld, improper text procedure, wrong part issued, film on parts, voids in casting, incorrect dimensions, adhesive failure, masking insufficient, spray failure). For each day of the week, tally marks are made for each defect type. Totals are on the right and bottom.

Note

Changes made to original OER material:

- Added Accessibility Note due to the length of alt text.
- Modified text in th beginning paragraph.
- Put Figure 6.4.5.1.1 into Libretexts template form for accessibility reasons.

This page titled 6.4.5.1: Check Sheets is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by LibreTexts.

6.4.5.2: Histograms

Raw data from a check sheet may be put into a histogram. Data that is continuous in nature can be put into a Histogram that contains ranges of the data. It will show an accurate representation of the distribution of the data.

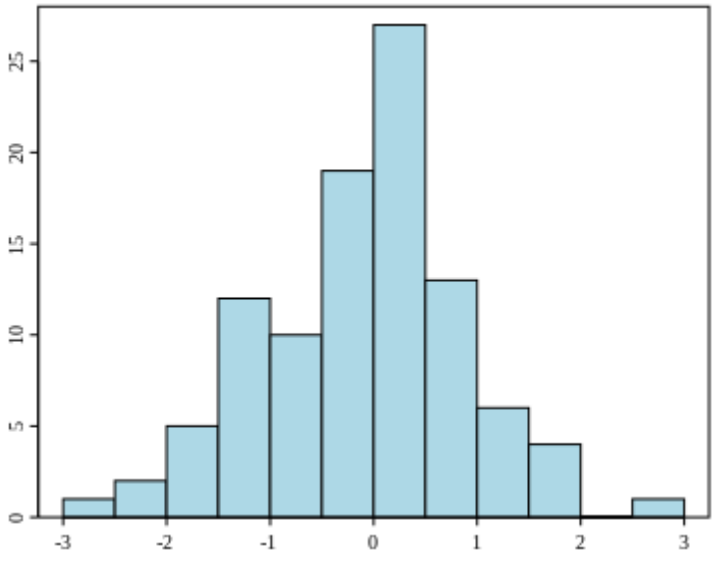


Figure 6.4.5.2.1: Example of a histogram.

This page titled 6.4.5.2: Histograms is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by LibreTexts.

6.4.5.3: Control Charts

In order to monitor the performance of a process over time, a control chart is the appropriate tool. A Control Chart includes an Upper Control limit and a lower Control limit, which are used to control the quality dimension that is measured. As long as points seem to appear randomly on both sides of the mean and they fall between the upper and lower control limits, the process is assumed to be in control.

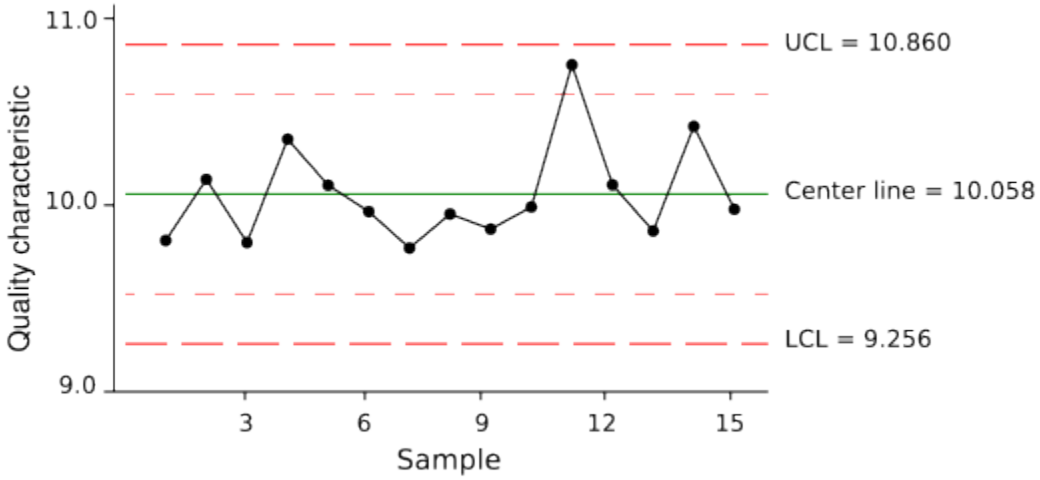


Figure 6.4.5.3.1: Example of a control chart. (in original OER source material)

Accessibility Note

Control line chart with Quality Characteristic on the left (numbered 9.0 to 11.0) and Sample on the bottom (numbered 0-15). The graphed line moves up and down over the Center line (numbered 10.058) and almost touches the UCL (numbered 10.860) at one point. The LCL is labeled as 9.256.

Note

Changes made to original OER material:

- Added accessibility note due to length of alt text.

This page titled 6.4.5.3: Control Charts is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by LibreTexts.

6.4.5.4: Pareto Charts

A special type of bar chart that shows the number of occurrences of a particular characteristic, ordered from highest to lowest. The X axis represents each characteristic and the Y axis is the number of times this occurrence was recorded. In addition, a cumulative percent line shows the cumulative percentage that each category represents. The Y axis on the right hand side of the chart corresponds to the percentage on this line. In the management of quality, managers must allocate resources to rectify the most frequently occurring problems. A Pareto analysis helps us focus our attention on the defects that occur the most frequently and to allocate the resources accordingly.

Steps in a Pareto Analysis:

1. Collect your raw data and put it into a simple table in descending order. Sum the total number of results at the bottom of the column.

Table 6.4.5.4.1: Complaints in descending order of frequency

Complaints	Number
Long wait time	81
food not hot	48
Server unknowledgeable	20
bill inaccurate	16
floor not clean	9
Menu items sold out	7

2. Include a cumulative column and calculate the cumulative percentage of each.

Table 6.4.5.4.2: Complaints in descending order of frequency with cumulative frequency and cumulative percent

Complaints	Number	Cumulative	Cumulative Percent
wait time long	81	81	44.8
food not hot	48	129	71.3
Server unknowledgeable	20	149	82.3
bill inaccurate	16	165	91.2
floor not clean	9	174	96.1
Menu items sold out	7	181	100.0

3. In EXCEL, your Pareto analysis will look like this.

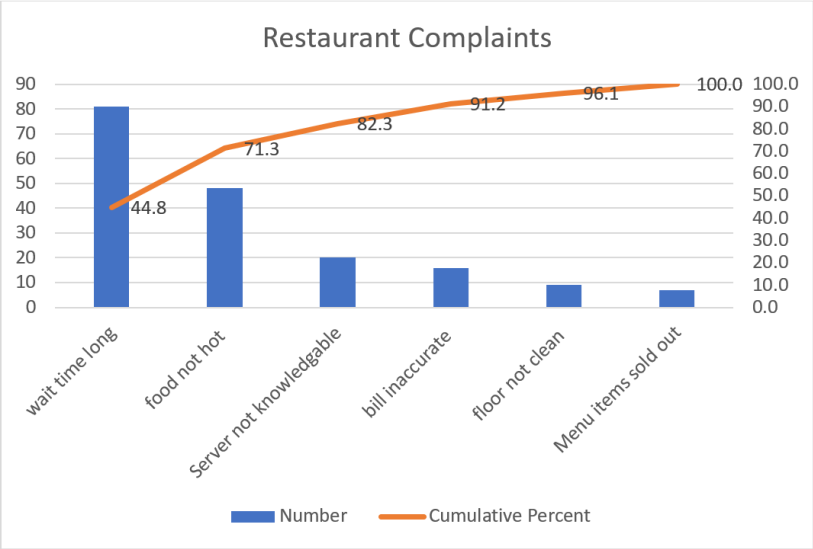


Figure 6.4.5.4.3: Example of a Pareto chart.

Accessibility Note

Figure 6.4.5.4.3 shows a vertical bar chart labeled Restaurant Complaints as its top title. The number of complaints is on the left vertical axis and vertical blue bars show the complaint count for wait time long, food not hot, server not knowledgeable, bill inaccurate, floor not clean, and Menu items sold out. An orange line going up from left to right shows the cumulative percent of complaints, which is on the vertical axis on the right hand side.

Note

Changes made:

- Accessibility fixes.
- Deleted video.

This page titled 6.4.5.4: Pareto Charts is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by LibreTexts.

6.4.5.5: Scatter Diagrams

A simple diagram helps to figure out if there is a relationship between two variables.

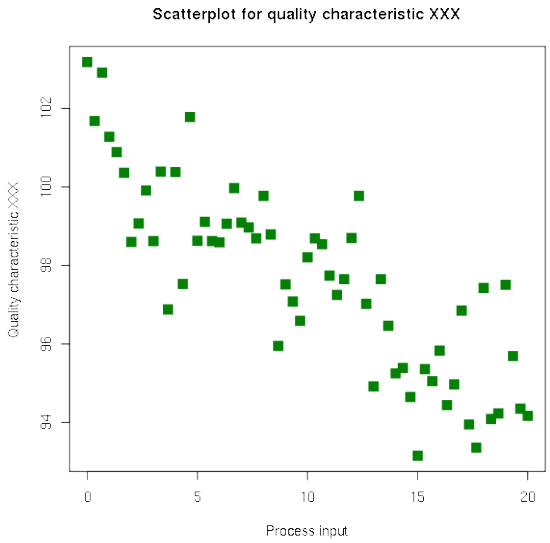


Figure 6.4.5.5.1: Example of a scatter diagram; Credit: DanielPenfield / Wikimedia / commons.wikimedia.org/wiki/File:Scatter_diagram_for_quality_characteristic_XXX.svg

Accessibility Note

Figure 6.4.5.5.1: Example of a scatter diagram. Scatterplot for quality characteristic XXX. Quality characteristic XXX is on the left side of the diagram and Process Input is on the bottom. Data points are scattered throughout the diagram field in a general downward direction as you move to the right.

Note

Changes made to original OER material:

- Added Accessibility Note due to length of alt text.

This page titled 6.4.5.5: Scatter Diagrams is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by LibreTexts.

6.4.5.6: Cause and Effect Diagrams

Also known as a Fishbone diagram, it was developed by Dr. Ishikawa[7] to help identify the causes of a problem.

The overall shape is that of a fish. The pointy end points to the 'effect' or the problem. Each of the ribs represents a major cause, or category that is a potential contributor to the problem. Commonly, the rib bones tend to be categories such as the man, method, material, machine, and environment. The actual factors that fall under each category are written on their related rib.

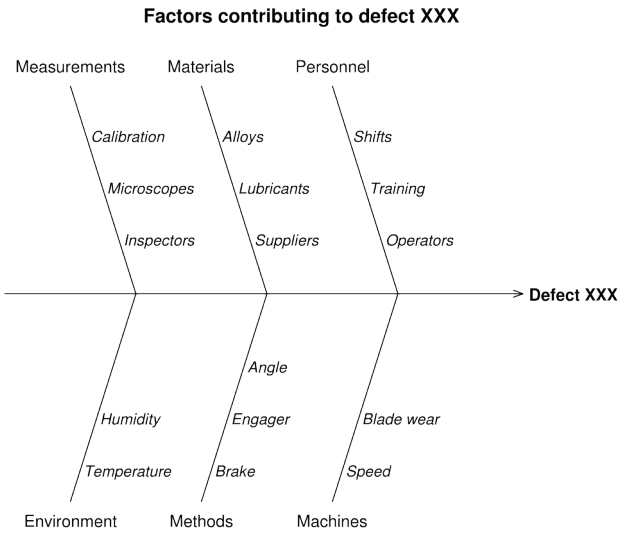


Figure 6.4.5.6.1: Example of a cause and effect diagram; Credit: DanielPenfield / Wikimedia

Accessibility Note

Figure 6.4.5.6.1: Factors contributing to defect XXX. An arrow through the middle is labeled Defect XXX. Six lines come off of that arrow (looking like fish bones). They are labeled Measurements (calibration, microscopes, inspectors), Materials (alloys, lubricants, suppliers), Personnel (shifts, training, operators), Environment (humidity, temperature), Methods (angle, engager, brake), and Machines (blade wear, speed).

Note

Changes made to original OER material:

1. Accessibility fixes.

This page titled 6.4.5.6: Cause and Effect Diagrams is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by LibreTexts.

6.4.6: References

1. Garvin, D. (1987). Competing on the Eight Dimensions of Quality. Retrieved on November 7, 2019, from HBR Garvin.
2. Wikipedia contributors. (2019). W. Edwards Deming. In *Wikipedia, The Free Encyclopedia*. Retrieved on November 9, 2019, from Wikipedia - Deming.
3. Saylor Academy. (2012). Producing for Quality. Retrieved on November 9, 2019, from Producing for Quality.
4. Wikipedia contributors. (2019). International Organization for Standardization. In Wikipedia, The Free Encyclopedia. Retrieved on January 2, 2020, from Wikipedia - International Organization for Standardization.
5. Certification.
6. American Society for Quality. (n.d.). THE DEFINE, MEASURE, ANALYZE, IMPROVE, CONTROL (DMAIC) PROCESS. Retrieved on January 22, 2020, from ASQ - DMAIC.
7. Wikiquote - Ishikawa.

This page titled 6.4.6: References is shared under a CC BY-NC 4.0 license and was authored, remixed, and/or curated by LibreTexts.

6.5: Legal Considerations

Following the two Boeing 737 Max 8 crashes described earlier in this chapter, there has been significant civil litigation (in federal courts) as well as criminal proceedings. Families of the 346 crash victims (from around 30 countries) in the Lion Air and Ethiopian Airlines crashes have filed at least 100 lawsuits against Boeing claiming that the 737 Max 8 aircraft was unsafe and that Boeing was liable for damages, seeking compensation and accountability. Allegations included negligence, wrongful death, and fraud. Many of these claims reached confidential settlements, so no details are available.

Criminal Charge

The following is taken verbatim from the criminal charges filed on January 7, 2021 in the Northern District of Texas, Fort Worth Division, in Case 4:21-cr-00005-O:[6]

CRIMINAL INFORMATION

The United States of America charges: COUNT ONE (Conspiracy to Defraud the United States)

1. From at least in or around November 2016 through at least in or around December 2018, in the Northern District of Texas and elsewhere, the Defendant, THE BOEING COMPANY, knowingly and willfully, and with the intent to defraud, conspired and agreed together with others to defraud the United States by impairing, obstructing, defeating, and interfering with, by dishonest means, the lawful function of a United States government agency, to wit, the Federal Aviation Administration Aircraft Evaluation Group ("FAA AEG") within the United States Department of Transportation, in connection with the FAA AEG's evaluation of the Boeing 737 MAX airplane's Maneuvering Characteristics Augmentation System ("MCAS"), including for purposes of the 737 MAX Flight Standardization Board Report ("FSB Report") and the 737 MAX differences-training determination.
2. In furtherance of the conspiracy and to effect its objects, in the Northern District of Texas and elsewhere, the Defendant, THE BOEING COMPANY, together with others, knowingly committed and caused the commission of the following overt acts, among others:

    a. On or about January 17, 2017, an employee of THE BOEING COMPANY emailed an employee of the FAA AEG about the 737 MAX FSB Report, stating in part, "Flight Controls: Delete MCAS, recall we decided we weren't going to cover it [. . .] since it's way outside the normal operating envelope."

    b. On or about July 7, 2017, an employee of THE BOEING COMPANY emailed employees of at least two major U.S.-based airline customers, stating in part, "Attached is the final and approved 737 FSB [R]eport which adds the 737 MAX." (In violation of Title 18, United States Code, Section 371).

After being charged with fraud by the United Stated Department of Justice (DOJ) in connection with both crashes of the 737 MAX, Boeing settled by paying over US$2.5 billion in penalties and compensation, broken down as follows in the press release that announced a Deferred Prosecution Agreement (DPA):

- a criminal monetary penalty of $243.6 million,
- $1.77 billion in damages to airline customers, and
- $500 million to a fund for the families of crash victims.[2]

A summary of this settlement appeared in a Department of Justice (DOJ) press release [3] extracted here verbatim.

Boeing Charged with 737 Max Fraud Conspiracy and Agrees to Pay over $2.5 Billion

Thursday, January 7, 2021

For Immediate Release

Office of Public Affairs

The Boeing Company (Boeing) has entered into an agreement with the Department of Justice to resolve a criminal charge related to a conspiracy to defraud the Federal Aviation Administration's Aircraft Evaluation Group (FAA AEG) in connection with the FAA AEG's evaluation of Boeing's 737 MAX airplane.

Boeing, a U.S.-based multinational corporation that designs, manufactures, and sells commercial airplanes to airlines worldwide, entered into a **deferred prosecution agreement (DPA)** in connection with a criminal information filed today in the Northern District of Texas. The criminal information charges the company with one count of conspiracy to defraud the United States. Under the terms of the DPA, Boeing will pay a total criminal monetary amount of over $2.5 billion, composed of a criminal monetary penalty of $243.6 million, compensation payments to Boeing's 737 MAX airline customers of $1.77 billion, and the establishment of a $500 million crash-victim beneficiaries fund to compensate the heirs, relatives, and legal beneficiaries of the 346 passengers who died in the Boeing 737 MAX crashes of Lion Air Flight 610 and Ethiopian Airlines Flight 302.

"The tragic crashes of Lion Air Flight 610 and Ethiopian Airlines Flight 302 exposed fraudulent and deceptive conduct by employees of one of the world's leading commercial airplane manufacturers," said Acting Assistant Attorney General David P. Burns of the Justice Department's Criminal Division. "Boeing's employees chose the path of profit over candor by concealing material information from the FAA concerning the operation of its 737 MAX airplane and engaging in an effort to cover up their deception. This resolution holds Boeing accountable for its employees' criminal misconduct, addresses the financial impact to Boeing's airline customers, and hopefully provides some measure of compensation to the crash-victims' families and beneficiaries."

"The misleading statements, half-truths, and omissions communicated by Boeing employees to the FAA impeded the government's ability to ensure the safety of the flying public," said U.S. Attorney Erin Nealy Cox for the Northern District of Texas. "This case sends a clear message: The Department of Justice will hold manufacturers like Boeing accountable for defrauding regulators – especially in industries where the stakes are this high."

"Today's deferred prosecution agreement holds Boeing and its employees accountable for their lack of candor with the FAA regarding MCAS," said Special Agent in Charge Emmerson Buie Jr. of the FBI's Chicago Field Office. "The substantial penalties and compensation Boeing will pay, demonstrate the consequences of failing to be fully transparent with government regulators. The public should be confident that government regulators are effectively doing their job, and those they regulate are being truthful and transparent."

"We continue to mourn alongside the families, loved ones, and friends of the 346 individuals who perished on Lion Air Flight 610 and Ethiopian Airlines Flight 302. The deferred prosecution agreement reached today with The Boeing Company is the result of the Office of Inspector General's dedicated work with our law enforcement and prosecutorial partners," said Special Agent in Charge Andrea M. Kropf, Department of Transportation Office of Inspector General (DOT-OIG) Midwestern Region. "This landmark deferred prosecution agreement will forever serve as a stark reminder of the paramount importance of safety in the commercial aviation industry, and that integrity and transparency may never be sacrificed for efficiency or profit."

As Boeing admitted in court documents, Boeing—through two of its 737 MAX Flight Technical Pilots—deceived the FAA AEG about an important aircraft part called the Maneuvering Characteristics Augmentation System (MCAS) that impacted the flight control system of the Boeing 737 MAX. Because of their deception, a key document published by the FAA AEG lacked information about MCAS, and in turn, airplane manuals and pilot-training materials for U.S.-based airlines lacked information about MCAS.

Boeing began developing and marketing the 737 MAX in or around June 2011. Before any U.S.-based airline could operate the new 737 MAX, U.S. regulations required the FAA to evaluate and approve the airplane for commercial use.

In connection with this process, the FAA AEG was principally responsible for determining the minimum level of pilot training required for a pilot to fly the 737 MAX for a U.S.-based airline, based on the nature and extent of the differences between the 737 MAX and the prior version of Boeing's 737 airplane, the 737 Next Generation (NG). At the conclusion of this evaluation, the FAA AEG published the 737 MAX Flight Standardization Board Report (FSB Report), which contained relevant information about certain aircraft parts and systems that Boeing was required to incorporate into airplane manuals and pilot-training materials for all U.S.-based airlines. The 737 MAX FSB Report also contained the FAA AEG's differences-training determination. After the 737 MAX FSB Report was published, Boeing's airline customers were permitted to fly the 737 MAX.

Within Boeing, the 737 MAX Flight Technical Team (composed of 737 MAX Flight Technical Pilots) was principally responsible for identifying and providing to the FAA AEG all information that was relevant to the FAA AEG in connection with the FAA AEG's publication of the 737 MAX FSB Report. Because flight controls were vital to flying modern commercial airplanes, differences

between the flight controls of the 737 NG and the 737 MAX were especially important to the FAA AEG for purposes of its publication of the 737 MAX FSB Report and the FAA AEG's differences-training determination.

In and around November 2016, two of Boeing's 737 MAX Flight Technical Pilots, one who was then the 737 MAX Chief Technical Pilot and another who would later become the 737 MAX Chief Technical Pilot, discovered information about an important change to MCAS. Rather than sharing information about this change with the FAA AEG, Boeing, through these two 737 MAX Flight Technical Pilots, concealed this information and deceived the FAA AEG about MCAS. Because of this deceit, the FAA AEG deleted all information about MCAS from the final version of the 737 MAX FSB Report published in July 2017. In turn, airplane manuals and pilot training materials for U.S.-based airlines lacked information about MCAS, and pilots flying the 737 MAX for Boeing's airline customers were not provided any information about MCAS in their manuals and training materials.

On Oct. 29, 2018, Lion Air Flight 610, a Boeing 737 MAX, crashed shortly after takeoff into the Java Sea near Indonesia. All 189 passengers and crew on board died. Following the Lion Air crash, the FAA AEG learned that MCAS activated during the flight and may have played a role in the crash. The FAA AEG also learned for the first time about the change to MCAS, including the information about MCAS that Boeing concealed from the FAA AEG. Meanwhile, while investigations into the Lion Air crash continued, the two 737 MAX Flight Technical Pilots continued misleading others—including at Boeing and the FAA—about their prior knowledge of the change to MCAS.

On March 10, 2019, Ethiopian Airlines Flight 302, a Boeing 737 MAX, crashed shortly after takeoff near Ejere, Ethiopia. All 157 passengers and crew on board died. Following the Ethiopian Airlines crash, the FAA AEG learned that MCAS activated during the flight and may have played a role in the crash. On March 13, 2019, the 737 MAX was officially grounded in the U.S., indefinitely halting further flights of this airplane by any U.S.-based airline.

As part of the DPA, Boeing has agreed, among other things, to continue to cooperate with the Fraud Section in any ongoing or future investigations and prosecutions. As part of its cooperation, Boeing is required to report any evidence or allegation of a violation of U.S. fraud laws committed by Boeing's employees or agents upon any domestic or foreign government agency (including the FAA), regulator, or any of Boeing's airline customers. In addition, Boeing has agreed to strengthen its compliance program and to enhanced compliance program reporting requirements, which require Boeing to meet with the Fraud Section at least quarterly and to submit yearly reports to the Fraud Section regarding the status of its remediation efforts, the results of its testing of its compliance program, and its proposals to ensure that its compliance program is reasonably designed, implemented, and enforced so that it is effective at deterring and detecting violations of U.S. fraud laws in connection with interactions with any domestic or foreign government agency (including the FAA), regulator, or any of its airline customers.

The department reached this resolution with Boeing based on a number of factors, including the nature and seriousness of the offense conduct; Boeing's failure to timely and voluntarily self-disclose the offense conduct to the department; and Boeing's prior history, including a civil FAA settlement agreement from 2015 related to safety and quality issues concerning the Boeing's Commercial Airplanes (BCA) business unit. In addition, while Boeing's cooperation ultimately included voluntarily and proactively identifying to the Fraud Section potentially significant documents and Boeing witnesses, and voluntarily organizing voluminous evidence that Boeing was obligated to produce, such cooperation, however, was delayed and only began after the first six months of the Fraud Section's investigation, during which time Boeing's response frustrated the Fraud Section's investigation.

The department also considered that Boeing engaged in remedial measures after the offense conduct, including: (i) creating a permanent aerospace safety committee of the Board of Directors to oversee Boeing's policies and procedures governing safety and its interactions with the FAA and other government agencies and regulators; (ii) creating a Product and Services Safety organization to strengthen and centralize the safety-related functions that were previously located across Boeing; (iii) reorganizing Boeing's engineering function to have all Boeing engineers, as well as Boeing's Flight Technical Team, report through Boeing's chief engineer rather than to the business units; and (iv) making structural changes to Boeing's Flight Technical Team to increase the supervision, effectiveness, and professionalism of Boeing's Flight Technical Pilots, including moving Boeing's Flight Technical Team under the same organizational umbrella as Boeing's Flight Test Team, and adopting new policies and procedures and conducting training to clarify expectations and requirements governing communications between Boeing's Flight Technical Pilots and regulatory authorities, including specifically the FAA AEG. Boeing also made significant changes to its top leadership since the offense occurred.

The department ultimately determined that an independent compliance monitor was unnecessary based on the following factors, among others: (i) the misconduct was neither pervasive across the organization, nor undertaken by a large number of employees, nor facilitated by senior management; (ii) although two of Boeing's 737 MAX Flight Technical Pilots deceived the FAA AEG about MCAS by way of misleading statements, half-truths, and omissions, others in Boeing disclosed MCAS's expanded operational scope to different FAA personnel who were responsible for determining whether the 737 MAX met U.S. federal airworthiness

standards; (iii) the state of Boeing's remedial improvements to its compliance program and internal controls; and (iv) Boeing's agreement to enhanced compliance program reporting requirements, as described above.

The Chicago field offices of the FBI and the DOT-OIG investigated the case, with the assistance of other FBI and DOT-OIG field offices.

Trial Attorneys Cory E. Jacobs and Scott Armstrong and Assistant Chief Michael T. O'Neill of the Fraud Section and Assistant U.S. Attorney Chad E. Meacham of the Northern District of Texas are prosecuting this case.

Individuals who believe they may be an heir, relative, or legal beneficiary of one of the Lion Air Flight 610 or Ethiopian Airlines Flight 302 passengers in this case should contact the Fraud Section's Victim Witness Unit by email at: Victimassistance email or call (888) 549-3945.

Updated February 5, 2025

2021 Deferred Prosecution Agreement Details

The announcement of the Deferred Prosecution Agreement and some high level financial penalty elements were in the press release just covered.

However, what did the 58 page Deferred Prosecution Agreement actually say about deferral of criminal prosecutions? Let's look at the first two pages of the document [7,p1-2], extracted verbatim:

DEFERRED PROSECUTION AGREEMENT

Defendant The Boeing Company (the "Company"), pursuant to authority granted by the Company's Board of Directors reflected in Attachment B, the United States Department of Justice, Criminal Division, Fraud Section (the "Fraud Section"), and the United States Attorney's Office for the Northern District of Texas (the "USAO-NDTX") enter into this deferred prosecution agreement (the "Agreement"). The terms and conditions of this Agreement are as follows:

Criminal Information and Acceptance of Responsibility

1. The Company acknowledges and agrees that the Fraud Section will file the attached one-count criminal Information in the United States District Court for the Northern District of Texas charging the Company with Conspiracy to Defraud the United States, in violation of Title 18, United States Code, Section 371 (the "Information"). In so doing, the Company: (a) knowingly waives any right it may have to indictment on this charge, as well as all rights to a speedy trial pursuant to the Sixth Amendment to the United States Constitution, Title 18, United States Code, Section 3161, and Federal Rule of Criminal Procedure 48(b); and (b) knowingly waives any objection with respect to venue to any charges by the United States arising out of the conduct described in the Statement of Facts attached hereto as Attachment A (the "Statement of Facts") and consents to the filing of the Information, as provided under the terms of this Agreement, in the United States District Court for the Northern District of Texas. The Fraud Section agrees to defer prosecution of the Company pursuant to the terms and conditions described below.
2. The Company admits, accepts, and acknowledges that it is responsible under United States law for the acts of its officers, directors, employees, and agents as charged in the Information, and as set forth in the Statement of Facts, and that the allegations described in the Information and the facts described in the Statement of Facts are true and accurate. The Company agrees that, effective as of the date it signs this Agreement, in any prosecution that is deferred by this Agreement, it will not dispute the Statement of Facts set forth in this Agreement, and, in any such prosecution, the Statement of Facts shall be admissible as: (a) substantive evidence offered by the government in its case-in-chief and rebuttal case; (b) impeachment evidence offered by the government on cross-examination; and (c) evidence at any sentencing hearing or other hearing. In addition, in connection therewith, the Company agrees not to assert any claim under the United States Constitution, Rule 410 of the Federal Rules of Evidence, Rule 11(f) of the Federal Rules of Criminal Procedure, Section 1B1.1(a) of the United States Sentencing Guidelines ("USSG" or "Sentencing Guidelines"), or any other federal rule that the Statement of Facts should be suppressed or is otherwise inadmissible as evidence in any form.

2025 Non-Prosecution Agreement

In May 2025, Boeing reached a $1.1 billion non-prosecution agreement (NPA) with the Department of Justice (DOJ), thereby

avoiding a criminal trial while admitting to obstruction and agreeing to invest in safety and compliance programs. This agreement incurred criticism from some victims' families as being insufficient and allowing Boeing to avoid full accountability. There will be more on this later in this section.

> **Note**
>
> Note that the January 2021 agreement called for the establishment of a $500 million crash-victim beneficiaries fund to compensate the heirs, relatives, and legal beneficiaries of the 346 passengers who died in the Boeing 737 MAX crashes of Lion Air Flight 610 and Ethiopian Airlines Flight 302

The specifics of the additional victim payments are as follows, taken verbatim from the details of the published non-prosecution agreement:[4,p6-7]

11. The Company agrees to pay **additional victim compensation** in the amount of $444.5 million to the heirs, relatives, and/or legal beneficiaries of the crash victims of Lion Air Flight 610 and Ethiopian Airlines Flight 302 (the "Crash-Victim Beneficiaries Compensation Amount"). No later than five (5) business days after the Agreement is executed, the Company shall deposit the full Crash-Victim Beneficiaries Compensation Amount into an escrow account ("Crash-Victim Beneficiaries Escrow Account") that it shall establish for that purpose. No monies will be paid out of the Crash-Victim Beneficiaries Escrow Account until after the entry of a final Order of the Court dismissing the Information, and only then with the prior approval of the Offices.

    (a) The Estate of each Crash-Victim will be entitled to an equal pro rata share of the $444.5 million fund. The parties agree that the appointment of a Crash-Victim Beneficiaries Compensation Claims Administrator (the "Administrator") is appropriate and necessary to determine the proper administration and disbursement of the Crash-Victim Beneficiaries Compensation Amount.

    (b) The Administrator shall make recommendations to the Offices regarding the individuals who should receive payments from the Crash-Victim Beneficiaries Compensation Amount. Only the Offices shall be empowered to make final decisions regarding the individuals who should receive the victim payments from the Crash-Victim Beneficiaries Compensation Amount.

    (c) The parties agree that the Administrator should be the Crash-Victim Beneficiaries Compensation Claims Administrator selected and engaged under the DPA. The Company agrees to pay for all costs, fees, and expenses incurred by the Administrator. The Company shall execute an engagement letter with the Administrator that must be approved, in advance of execution, by the Offices. The Offices and the Company will use their best efforts to complete the engagement within thirty (30) calendar days of the execution of this Agreement.

    (d) The Company agrees that it will not employ or be affiliated with the Administrator for a period of not less than two years from the date on which the Administrator's term expires. Nor will the Company discuss with the Administrator the possibility of further employment or affiliation during the Administrator's term. Upon agreement by the parties, this prohibition will not apply to other claims administration responsibilities that the Administrator may undertake in connection with resolutions with foreign or other domestic authorities.

    (e) The Company agrees that it will not use the fact that any beneficiary of the crash victims of Lion Air Flight 610 and Ethiopian Airlines Flight 302 seeks or receives any compensation from the Crash-Victim Beneficiaries Compensation Amount to seek to preclude such beneficiary from pursuing any other lawful claim that such beneficiary might have against the Company.

Mandamus petitions from Crash Victims

On March 31, 2026, a panel of the United States Court of Appeals for the Fifth Circuit issued a decision on the **mandamus petitions** brought by certain representatives of victims of the crashes of Lion Air Flight 610 and Ethiopian Airlines Flight 302 (together, the "Mandamus Petitions").

Definition: Mandamus petition

A mandamus petition is a legal request to a court compelling a government official, agency, or lower court to perform a duty they are legally obligated to carry out. From the Latin "we command."

The court first considered the representatives' mandamus petition that challenged the 2021 deferred prosecution agreement (DPA) between the United States and the Boeing Company. The court concluded that Boeing's 2024 breach of the DPA meant that the agreement "no longer bound the contracting parties and any challenge based on its terms became moot."

Regarding the second mandamus petition, which challenged the 2025 non-prosecution agreement (NPA), the court held that "the district court correctly concluded that the Department did not violate the [Crime Victims' Rights Act (CVRA)] by failing to confer with the families or by misleading them about the NPA." Finally, the court concluded that the CVRA did not provide the court of appeals with jurisdiction "to substantively review a district court's decision to grant a motion dismissing a criminal prosecution." Accordingly, the Fifth Circuit panel denied the Mandamus Petitions.[5]

References

1. Associated Press. "Justice Department reaches deal to allow Boeing to avoid prosecution over 737 Max crashes". The Mercury News. May 23, 2025. Accessed May 23, 2025.
2. Wikipedia - Boeing 737 Max. Accessed April 5, 2026. Text is available under the Creative Commons Attribution-ShareAlike 4.0 License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.
3. *"Boeing Charged with 737 Max Fraud Conspiracy and Agrees to Pay over $2.5 Billion" (Press release). The United States Department of Justice. January 7, 2021. Archived from the original on January 8, 2021. Retrieved January 8, 2021.* Updated February 5, 2025. Accessed April 5, 2026.
4. The United States Department of Justice, Criminal Division, Non-prosecution Agreement, United States v. The Boeing Company, 4:21-cr-00005-O, May 29, 2025. Non-prosecution Agreement. Accessed April 7, 2026.
5. United States Department of Justice. United States v. The Boeing Company, Docket Number 4:21-cr-00005-O. United States v. The Boeing Company. Accessed April 7, 2026.
6. United States v. The Boeing Company charges. Boeing Charged with 737 Max Fraud Conspiracy and Agrees to Pay over $2.5 Billion:U.S. v. The Boeing Company - Final Criminal Information. Filed as Document 1 on 01/07/21 in Case 4:21-cr-00005-O UNITED STATES OF AMERICA v. THE BOEING COMPANY, Defendant. This is the CRIMINAL INFORMATION (charges). Accessed April 8, 2026.
7. Boeing Charged with 737 Max Fraud Conspiracy and Agrees to Pay over $2.5 Billion: U.S. v. The Boeing Company - Final DPA Package. Final DPA. Filed as Document 4 on 01/07/21 in Case 4:21-cr-00005-O UNITED STATES OF AMERICA v. THE BOEING COMPANY, Defendant. This is the DEFERRED PROSECUTION AGREEMENT. Accessed April 8, 2026.
8. Richer, Alanna Durkin. "Justice Department reaches deal to allow Boeing to avoid prosecution over 737 Max crashes", May 23, 2025. The Mercury News (online version only, not in print version). Justice Department reaches deal to allow Boeing to avoid prosecution over 737 Max crashes.

This page titled 6.5: Legal Considerations is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

Terms

- **AOA:** Angle of Attack sensor on the Boeing 737
- **ANU:** Airplane Nose-up
- **ATA:** Airline Transport Association
- **ATC:** Air Traffic Control
- **CDU:** Cockpit Display Unit
- **CFR:** Code of Federal Regulations
- **COS:** Continued Operational Safety
- **CVRA:** Crime Victims' Rights Act
- **DFCS:** Digital Flight Control System
- **DFDR:** Digital Flight Data Recorder, often called the "black box"
- **DOJ:** United States Department of Justice
- **DPA:** Deferred Prosecution Agreement
- **EASA:** European-Union Aviation Safety Agency
- **FAA:** Federal Aviation Administration
- **FAA AEG** Federal Aviation Administration's Aircraft Evaluation Group
- **FCC:** Flight Control Computer
- **FDR:** Flight Data Recorder
- **FFS:** Full Flight Simulator
- **JATR:** Joint Authorities Technical Review
- **KNKT:** Komite Nasional Keselamatan Transportasi (Indonesia's National Transportation Safety Committee)
- **MCAS:** Maneuvering Characteristics Augmentation System, a flight control system on the Boeing 737
- **NASA:** National Aeronautics and Space Administration
- **NPA:** Non-Prosecution Agreement
- **NTSB:** National Transportation Safety Board
- **PFD:** Primary Flight Display
- **RTS:** Return To Service
- **SMS:** Safety Management System
- **SSA:** System Safety Analysis
- **STM:** Stabilizer Trim Motor
- **STS:** Speed Trim System

References

1. FAA, Summary of the FAA's Review of the Boeing 737 MAX, Return to Service of the Boeing 737 MAX Aircraft, November 18, 2020. Summary of the FAA's Review of the Boeing 737 MAX. Accessed March 26, 2026.

This page titled 6.6: Glossary is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

7: ARTIFICIAL INTELLIGENCE USAGE BY UTILITIES

Figure thumbnail for the chapter: Wikimedia - AI and Machine Learning. August 16, 2018. This file is licensed under the Creative Commons Attribution 2.0 Generic license. This image was originally posted to **Flickr** by mikemacmarketing at flickr. It was reviewed on 29 October 2020 by **FlickreviewR 2** and was confirmed to be licensed under the terms of the cc-by-2.0. Author according to Wikimedia.

7.1: Introduction

7.2: Specific Areas of Application of Artificial Intelligence

7.3: Artificial Intelligence Use by a California Utility (PGandE)

7.4: AI Initiatives by other California Utilities

7.4.1: Southern California Edison

7.4.2: San Diego Gas and Electric

7.5: Applying Artificial Intelligence to Wildfire Detection and Management

7.5.1: Vendor Solutions

7.5.1.1: N5 Sensors

7.5.1.2: Ladris

7.5.1.3: Pano.ai

7.6: Project Management Considerations

7.7: Legal Considerations

7.7.1: CPUC Governance

7.7.2: Wildfire Camera Privacy Issues

7.7.3: Impact of Potential AI Regulations

7.8: Glossary of Terms

This page titled 7: Artificial Intelligence Usage by Utilities is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

7.1: Introduction

The application of Artificial Intelligence (AI) has accelerated in recent years and is now part of everyday parlance. What may be surprising is that some electric utilities are actively studying the use of artificial intelligence in a number of applications. Electric utilities typically are not known as early adopters of new technology and are known to be risk averse.

In this case study, we look in general at areas where utilities could use artificial intelligence and then look specifically at how one utility, Pacific Gas and Electric in Northern California, plans to use artificial intelligence. We then look at what California's other two major utilities (Southern California Edison and San Diego Gas and Electric) are exploring.

We then zero in on one specific application that will be of interest to residents of California: the use of artificial intelligence in wildfire detection and management. Weather in the state of California can be dry and hot from late spring through the fall. Occasional high winds called Diablo winds in Northern California and Santa Ana winds in Southern California can turn wildfires into firestorms with devastating effect on property and people. Recently, some insurance companies have stopped doing business in California or have increased fire insurance policy rates to mitigate their economic risks, making it difficult for some property owners to retain or obtain fire insurance or to pay their premiums.

The severity of the problem can be seen by looking at the size of the top ten fires in the California Department of Forestry and Fire Protection incident data[1] as of September 8, 2025 processed by the authors:

Table 7.1.1*: Fire Acreage Burned*

Fire name	Sum of incident_acres_burned
August Complex (includes Doe Fire)	1,032,648
Dixie Fire	963,369
Park Fire	432,351
Ranch Fire (Mendocino Complex)	410,203
Creek Fire	402,549
SCU Lightning Complex	396,624
LNU Lightning Complex	363,220
Thomas Fire	281,917
Rim Fire	258,576
Carr Fire	229,651

Obviously, anything that can be done to detect and quickly suppress wildfires could benefit the residents of California.

Let's take a look at the definitions of a few terms that will recur in this chapter.

Definition: Artificial Intelligence

Artificial intelligence (AI) is the capability of computational systems to perform tasks typically associated with human intelligence, such as learning, reasoning, problem-solving, perception, and decision-making.[2]

Definition: Machine Learning

Machine learning (ML) is a field of study in artificial intelligence concerned with the development and study of statistical algorithms that can learn from data and generalise [generalize] to unseen data, and thus perform tasks without explicit instructions. [3]

Definition: Deep Learning

In machine learning, deep learning focuses on utilizing multilayered neural networks to perform tasks such as classification, regression, and representation learning. [4]

Definition: Generative AI

Generative artificial intelligence (Generative AI, GenAI, or GAI) is a subfield of artificial intelligence that uses generative models to produce text, images, videos, or other forms of data.[These models learn the underlying patterns and structures of their training data and use them to produce new data based on the input, which often comes in the form of natural language prompts. [5]

Definition: Predictive Analytics

Predictive analytics is a set of business intelligence (BI) technologies that uncovers relationships and patterns within large volumes of data that can be used to predict behavior and events. Unlike other BI technologies, predictive analytics is forward-looking, using past events to anticipate the future. [6]

Predictive analytics encompasses a variety of statistical techniques from data mining, predictive modeling, and machine learning that analyze current and historical facts to make predictions about future or otherwise unknown events.[6]

References

1. Cal Fire. CAL FIRE incidents. Incident data pulled on September 8, 2025 and sorted by "incident acres burned."
2. Wikipedia authors. Wikipedia - Artificial Intelligence. Information pulled 9/14/2025. Text is available under the Creative Commons Attribution-ShareAlike 4.0 License.
3. Wikipedia authors. Wikipedia - Machine Learning. Information pulled 9/14/2025. Text is available under the Creative Commons Attribution-ShareAlike 4.0 License.
4. Wikipedia authors. Wikipedia - Deep learning. Information pulled 9/14/2025. Text is available under the Creative Commons Attribution-ShareAlike 4.0 License.
5. Wikipedia authors. Wikipedia - Generative Artificial Intelligence. Information pulled 9/14/2025. Text is available under the Creative Commons Attribution-ShareAlike 4.0 License.
6. Wikipedia authors. Wikipedia - Predictive analytics. Information pulled 9/14/2025. Text is available under the Creative Commons Attribution-ShareAlike 4.0 License.

This page titled 7.1: Introduction is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

7.2: Specific Areas of Application of Artificial Intelligence

We will start by looking at some potential applications of AI by electric utilities and then take a look at the current state of adoption of AI. In the next section we will do a deeper dive into the plans to deploy AI by Pacific Gas and Electric.

We will be drawing on concepts from a recent seminal paper on the topic[1].

> Note
>
> Keep in mind that electric companies are highly regulated and need to maintain high availability to their customer base. As a result, electric utilities can be risk-averse and potentially slow in adopting new technologies such as AI.

We will be exploring some of the following concepts in this section:

- Community to grid interactions
- Managing disaster risks
- Power grid optimization
- Energy system resiliency
- Weather forecasting
- Autonomous operations on energy plants
- Utility equipment maintenance predictions
- Machine vision.

In later sections, we will explore the following:

- Applicability to California wildfire management.
- Privacy and cybersecurity considerations associated with some of these initiatives.

Community to grid interactions

Over the years, utilities such as PG&E have deployed SmartMeter™ technology to residences and businesses. "SmartMeter™ and Meter-Connector provide communication between PG&E and the grid. This two-way communication allows us to quickly identify outages and resolve other service problems, typically without visiting your home or business."[2] Customers can view the information collected by PG&E providing an online history of energy usage and costs (hourly electric and daily gas and electric consumption) and compare consumption to prior time periods. This information can be useful to both the utility and to customers.

As more customers buy electric vehicles and charge them at home, it is important to do the charging in an economically efficient manner. Managing load shifting is important. It is important economically to do the charging when the rates are most favorable.

As customers deploy solar energy, it is useful to track energy flow.

Energy System Resiliency and Managing Disaster Risks

Various weather events can severely impact a utility's infrastructure. High winds can cause trees or tree branches to fall onto utility power lines. Utilities spend a great deal of effort trying to keep their power lines clear of tree branches by surveying potential exposures proactively using human resources on foot or using drones in the field and dispatching tree trimming equipment.

Utilities have reacted to threats to their infrastructure by embarking on projects to put their transmission lines underground ("undergrounding").

Utilities can use Artificial intelligence to predict equipment failure and schedule replacements before failure.

Utilities can use Artificial intelligence to make predictions on potential high wind events such as Diablo winds in Northern

California and Santa Ana winds in Southern California. Models can be trained based on years of historical data. Utilities can selectively de-energize parts of their power infrastructure to mitigate risks.

In the event of a fire, the sooner it is detected and suppressed, the better as far as limiting damage to utility infrastructure and the property of homeowners and businesses.

Autonomous operations in energy plants

Automating work in energy plants can yield operational efficiencies.

Utility equipment maintenance predictions

Mainframe and network equipment vendors have for years used predictive algorithms to schedule equipment maintenance or part replacement. For example, by monitoring the speed of fans that are used to cool equipment, these vendors can anticipate when the fan might fail and schedule replacing it.

Utilities can do something similar for their equipment based on the rated duty cycle of each type of equipment. Utilities could collect operational performance information and use it to predict failures and schedule maintenance in advance. The operational data could be supplemented by machine vision, automatic diagnotic inspections, and the like.

Environmental impacts of AI Usage

It should be noted that even as utilities and other industries use Artificial Intelligence, AI consumes vast computing resources placed in very large data centers. These computing resources consume a great deal of power that can have an impact on the environment.

References

1. Slate, Daniel D., Parisot, Alexandre, Min, Liang, Panciatici, Patrick, Van Hentenryck; Adoption of Artificial Intelligence by Electric Utilities Energy Law Journal, Volume 45, No. 1, 2024., The Energy Bar Association, University of Tulsa College of Law.
2. Pacific Gas and Electric. SmartMeter™. Smartmeter.

This page titled 7.2: Specific Areas of Application of Artificial Intelligence is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

7.3: Artificial Intelligence Use by a California Utility (PGandE)

In this section we will be looking in detail at the efforts of Pacific Gas and Electric (PG&E) to deploy artificial intelligence in its operations.

We will be examining two key sources of information:

1. PG&E R&D Strategy Report 2024[4]. This is manageable in size at only 195 pages.
    - PG&E fashions itself as "the AI-enabled utility" and uses Artificial Intelligence (AI) in applications such as meteorology, planning, inspections, monitoring, maintenance, and customer communications, among other functions. These AI applications strengthen PG&E's detection, prevention, and response to the risk of wildfires and help to achieve greater operational performance and customer experiences. One objective is to improve grid reliability. We will look at the details of each of these areas.
    - In the report, PG&E also identifies opportunities for AI to improve data collection and analysis, expand capacity, reduce emissions, stabilize customer bills and provide improved customer service.
2. PG&E Wildfire Mitigation Plan R0 2026-2028, Volume 1 of 2[1]. This is voluminous in size at 623 pages.
3. PG&E Wildfire Mitigation Plan R1 2026-2028, Volume 2 of 2[2]. This is also large in size at 463 pages.

Because of the size of these documents, we have summarized relevant sections that discuss PG&E's efforts in many areas in machine learning and artificial intelligence. Keep in mind that if you were to be hired as a project manager of an artificial intelligence project, it would behoove you to at least skim all this material to get up to speed to understand the range of implementations and potential projects, as well as some specifics for the project to which you are assigned.

Requirements for Wildfire Mitigation Plans

California Office of Energy Infrastructure Safety (OEIS) Requirement

The California Office of Energy Infrastructure Safety (OEIS) is a state department whose purpose is to help reduce utility-related wildfire risk and ensure safe, resilient energy infrastructure. OEIS was created in response to the devastating California wildfires from 2016 through 2018. It was established on July 1, 2021 as a standalone department under the California Natural Resources Agency to oversee electrical utilities and reduce utility-related wildfire risks, with the following key responsibilities:

- **Wildfire Risk Management**: Evaluate and enforce compliance with Wildfire Mitigation Plans (WMPs) submitted by electrical corporations, ensuring system hardening, vegetation management, and quality inspections.
- **Utility Oversight**: Regulate nine electrical corporations in California, reviewing safety culture, executive compensation tied to safety metrics, and operational compliance.

The OEIS (which refers to itself by the name "Energy Safety") summarizes the requirements for Wildfire Mitigation Plans as follows:

> "Electrical corporations are required to prepare and submit Wildfire Mitigation Plans (WMPs) to Energy Safety for review and approval. WMPs should describe how the electrical corporation is constructing, maintaining, and operating its electrical lines and equipment in a manner that will minimize the risk of catastrophic wildfire. In 2018, following the devastating wildfires in 2016 and 2017, the California Legislature passed several bills increasing oversight of the electrical corporations' efforts to reduce utility-related wildfires. Energy Safety is responsible for evaluating electrical corporations' wildfire risk reduction efforts mainly through the WMP process. Electrical corporations must submit WMPs assessing their level of wildfire risk and providing plans for wildfire risk reduction to Energy Safety. Energy Safety must ensure utility wildfire mitigation efforts sufficiently address increasing utility wildfire risk." [5]

California Public Utilities Commission on WMPs

The California Public Utilities Commission reiterates this requirement as follows:

"Electrical corporations are required to prepare and submit tabular wildfire mitigation data to the Office of Energy Infrastructure Safety (Energy Safety) on a quarterly basis in quarterly data reports (QDR). QDRs have several tables comprising wildfire mitigation data, including information on the costs and progress towards the goals and objectives set in the wildfire mitigation plan (WMP). This reporting covers both individual wildfire mitigation initiatives (initiatives) and major categories aggregating mitigations, such as vegetation management or grid design. QDRs support Energy Safety in its oversight and enforcement of electrical corporations' compliance with wildfire safety. Electrical corporations are subsequently required to submit an annual report on compliance (ARC) to Energy Safety, which addresses compliance with the WMP."[6]

PG&E R&D Strategy Report 2024

The PG&E R&D Strategy Report 2024 presents new challenges facing PG&E that could be solved by the use of artificial intelligence and other technologies. In this textbook we will concentrate on the ones pertinent to AI.

Particularly useful is the chapter "The AI-enabled Utility" [4, pages 16-31] and the 2024 Problem Statement including Wildfire [4, pages 140-154].

PG&E is currently using artificial intelligence to:

- quickly identify fire ignitions,
- inspect and maintain its assets (e.g., infrastructure and equipment), and
- to prioritize field work.

Since PG&E is a conservative, risk-averse electric utility, it plans to deploy AI with appropriate human oversight.

The AI-enabled utility

PG&E intends to use:

- advanced automation in its operations,
- predictive technologies to preemptively and proactively avoid faults in their equipment, and
- to deploy AI in its interactions with its customers.

PG&E's operational definition of artificial intelligence serves as an umbrella term for a variety of machine-based technologies technologies including:

- automation,
- machine learning (ML) and optimization,
- deep learning, and
- generative AI (GenAI).

Note on GenAI

"Because of its recency, GenAI still faces limitations and challenges, necessitating significant human oversight and monitoring to ensure trustworthy outcomes. As an operator of critical infrastructure, technology maturity is a key consideration for PG&E; therefore, we aim to approach GenAI with the appropriate governance and safeguards while remaining open to its full potential." [4, page 19]

This is reflective of the general risk-aversion and conservative approach one would expect from a regulated public utility.

PG&E has already applied machine learning to mitigate wildfire risks and enhance its operational efficiency**,** which we will explore in more detail in this case study. Its goal is to make it an "AI-enabled utility." [4, page 17]

What does this mean and how is it going to be achieved? The AI-enabled utility concept will be enabled by four conceptual category-oriented pillars: [4, page 18]

1. **Process-centered AI**: Automation-enhanced operations affecting PG&E-internal processes.
2. **Asset-centered AI**: Reliable and safe grid. This will be done using predictive asset lifecycle management. PG&E wants to proactively and autonomously detect and address equipment faults or disruptions before they occur, using targeted equipment maintenance programs, for example, on transformers.
3. **Customer-centered AI**: Personal energy experience. This would leverage data patterns, such as the energy consumption patterns observed in data collected and transmitted for analysis by Smart Meters on customer sites, to potentially tailor interactions with their customers and to optimize energy solution offerings, for example, for customers with electric vehicle charging equipment. Call centers could leverage AI-enabled conversational assistants.
4. **Workforce-centered AI**. This is for PG&E employees. Note that this is not a major concern of this case study.

The desired outcomes from these AI efforts would be:

1. Lower PG&E operational costs
2. Increased asset (equipment) reliability, productivity, and lifespan
3. Higher individual customer energy IQ and savings opportunities and
4. Higher retention and improved employee productivity. These are PG&E internal human resource activities that would use AI to improve operational efficiency of employees by automating repetitive processes, use GenAI, and improve data-driven decision making. For example, GenAI could be used in creating summarized minutes of meetings or creating initial templates of project plans based on similar prior projects.

PG&E's stated overall AI-enabled Utility goals are:

1. Improve affordability
2. Reduce risk
3. Accelerate growth.[4, page 17 and 18]

Equipment monitoring and maintenance

Inspections

PG&E has made use of computer vision for remote field inspections. New tools have improved efficiency by providing inspectors that are sitting at desks (not inspectors in the field) the ability to see and mark potential equipment problems using high-resolution images that have been taken via aerial inspections using drones and helicopters. These images are also used to train ML models.

PG&E manages a large number of assets of varying types, which PG&E personnel periodically inspect to identify signs of degradation that need to to be addressed through maintenance efforts. AI may be able to improve the precision and cost effectiveness of these inspections. Applying these technologies to understand degradation patterns and timelines based on asset type, weather and grid conditions, visual signs of damage, and various failure modes could yield huge gains in optimizing asset management efforts using intelligent risk predictions.[4, page 78]

PG&E conducts vegetation management patrols across thousands of miles of transmission lines trying to identify unhealthy trees that may pose an ignition risk. Numerous plant life species respond differently to environmental conditions and demonstrate unique growth and aging patterns. Artificial intelligence could be employed to analyze various indicators of tree health by species and then predict the likelihood of failure in the context of weather patterns.[4, page 79]

Predictive Maintenance

In the area of equipment predictive maintenance, PG&E is addressing electrical transformer predictive maintenance specifically. PG&E has developed machine learning models to assess the probability of a transformer failing within the next 30 days. This analysis can help prioritize those transformer units that are at highest risk of failing for review by human analysts for further investigation and field replacement before a failure in the field actually occurs. The benefits that this predictive maintenance work can yield are:

1. safety is improved by avoiding field failures that might ignite fires,
2. grid reliability is improved by having scheduled maintenance rather than unforeseen field failures causing outages, and
3. field maintenance efforts are prioritized to address the highest risk pieces of equipment.

Wildfire prediction, monitoring, and risk assessment

In the area of predictive wildfire risk assessment, PG&E has developed machine learning models that can process real-time weather data to proactively assess wildfire risk, looking out in time about 5.5 days. PG&E can concentrate its attention on and allocate its resources to the locations of highest risk in the predicted timeframes in order to reduce potential ignitions.[4, page 22]

Safety

In places where the risk of fire is the highest, PG&E has a goal to install 10,000 miles of distribution lines underground.[4, page 66] AI and machine learning could improve efficiency efforts "to identify, optimize, and efficiently execute the planning, design, and construction of undergrounding projects." [4, page 70]

For undergrounding project selection, "PG&E utilizes a combination of predictive models that assess wildfire and reliability risk at the circuit level to inform a manual review process to select candidate circuit segments for undergrounding. Advanced AI capabilities could streamline and enhance efforts to optimize the project portfolio. By providing the ability to seamlessly evaluate network effects across multiple combinations of adjacent circuit segments, we can better identify project portfolios that not only reduce wildfire risk but also maximize our ability to mitigate the impact of outage programs, such as EPSS and PSPS, by assessing the reliability impacts of undergrounding upstream or adjacent circuit segments." [4, page 70]

PG&E has already leveraged an ML-based solution in managing underground service tickets.

Gas

PG&E has already leveraged ML-based solutions to detect underground gas pipes.

Artificial intelligence and machine learning applications have the potential to enhance monitoring, planning, field operations, compliance, and gas emissions reductions, as follows:[4, page 90]

- **Enhancing safety through proactive maintenance:** Applications of AI/ML can enable predictive maintenance in gas operations by analyzing sensor data, inspection and maintenance records, and other inputs to detect early signs of equipment failure or anomalies in operational data such as gas composition. By enabling early intervention, these capabilities could reduce the risk of gas leaks, equipment failures, and safety hazards. [4, page 90]
- **Optimizing system planning:** AI-driven tools can help **optimize project planning** for pipeline construction or maintenance by considering multiple factors such as weather patterns, traffic, and resource availability. [4, page 91]
- **Enhancing personnel training and field work:** AI-driven tools could optimize field task assignments and safety protocols, while augmented reality technology could offer real-time guidance on pipeline schematics and hazard alerts, improving both efficiency and safety. [4, page 91]
- **Predicting and detecting methane leaks:** Machine learning algorithms can help detect methane leaks in real-time and predict where future leaks may occur based on observed historical patterns. This can help reduce repair timelines.[4, page 110]

Risk Management and Oversight

The use of AI comes with risks, and PG&E is treading carefully. The California Public Utilities Commission also provides oversight on the deployment of new technologies.

> "As a provider of critical infrastructure, PG&E has the responsibility to ensure the safety, reliability, affordability, and equity of the energy system. **For this reason, we are and will continue to be conservative in our deployment of AI systems**." [4, page 25]

Note

The bold text used by PG&E highlights a utility's conservative nature. Keep this in mind in the preparation of the project manager's project risk register. This would be reflective of the relevant stakeholders's risk appetites.

PG&E Problem Statements with High AI potential

The next section of the R&D Strategy Report 2024[4] covers innovation themes and 67 problem statements. Relevant items are discussed below. For each problem statement, PG&E has made a judgement as to whether AI will factor into the solution to the problem:

- Low AI potential
- Medium AI potential
- High AI potential. [4, page 33]

The problem statements flagged as "High AI Potential" are itemized here [4, pages 35-36], with details (as appropriate) on potential AI solutions included[4, starting at page 116]:

- Problem Statement 3 Simplifying EV charging installations for single-family homes, making use of AI/ML to assist the customer on installing EV charging stations and getting insights into overall load management.
- Problem Statement 5 Increasing precision and efficacy of energy management programs, making use of AI/ML.
- Problem Statement 6 Increasing utilization of existing T&D[Transmission and Distribution] infrastructure,making use of AI/ML.
- Problem Statement 7 Maintaining operational flexibility as load factor increases, making use of AI/ML in smart grid technologies.
- Problem Statement 9 Accelerating new interconnection planning.
- Problem Statement 10 Streamlining design and estimation for routine projects, making use of AI/ML (GenAI) to reduce the manual effort and time for designing and estimating routine projects.

Note

It might be wise to reflect on how this might impact the discipline of project management.

- Problem Statement 12 Increasing visibility into electrification pathways at the customer level, making use of AI/ML to analyze customer electrical load demand.
- Problem Statement 13 Enhancing visibility into DER [Distributed Energy Resources] and EV Load Flexibility Potential, making use of AI/ML.
- Problem Statement 15 Enabling intelligent orchestration of DERs[Distributed Energy Resources].

Wildfire Related Problem Statements (through Problem Statement 33)

- Problem Statement 21 Reducing ignition risk through improved asset diagnostics and predictive intervention, making use of AI/ML:
  - To make equipment assessments regarding wear and tear across distribution and transmission systems.
  - To predict the remaining lifetime of equipment based on asset type, current age, signs of degradation, repair history, and environmental conditions.
  - To enable root cause analysis of equipment failures.

Note

Those students with an engineering background are likely thinking of using the **Failures In Time** (FIT) rate of a device or component (the number of failures that can be expected in one billion device-hours of operation and how these can be used to predict **Mean Time To Failure (MTTF)** and **Mean Time Between Failures (MTBF)** of a higher level system.

- Problem Statement 23 Expanding capabilities and coverage of continuous fault monitoring (in the electrical grid) by centrally aggregating field sensor data and analyzing the data to develop better diagnostic and predictive capabilities about the grid.
- Problem Statement 24 Crowd-sourcing third-party data.
- Problem Statement 27 Advancing ignition notification and rapid suppression technologies, making use of AI/ML to expand real-time notification capabilities across all applicable PG&E equipment to improve the allocation of resources to the ignition site.
- Problem Statement 29 Automating PSPS events to improve safety and decrease customer impacts, making use of AI/ML.
- Problem Statement 30 Enhancing diagnostic capabilities for hazard trees, making use of AI/ML.
- Problem Statement 31 Addressing gaps in moisture monitoring and interpretation, making use of AI/ML, modeling tree health to improve capabilities to predict the risk of wildfires.
- Problem Statement 32 Understanding and predicting healthy tree failure, making use of AI/ML.
- Problem Statement 33 Targeting removal of fire-damaged trees.
- Problem Statement 41 Improving accuracy of [water] well life estimations, making use of AI/ML.
- Problem Statement 42 Increasing accuracy of geo-hazard risk assessment and monitoring, making use of AI/ML to assess risks to pipelines.
- Problem Statement 56 Addressing gaps in climate data, making use of AI/ML to enhance predictions.
- Problem Statement 57 Applying climate data in utility scenario-based planning and risk assessments, making use of AI/ML to probabilistically predict which utility assets are most likely to fail due to climate-driven causes and to predict which customers and locations would be impacted.
- Problem Statement 58 Preparing for precipitation events with enhanced impact assessments and planning, making use of AI/ML to predict how storms might impact equipment deterioration and failure rates.
- Problem Statement 65 Optimizing deployment of forest management interventions, making use of AI/ML.
- Problem Statement 66 Enhancing post-wildfire rehabilitation efforts to ensure future resilience, making use of AI/ML.

Disaggregating load for load management

There is significant potential for AI to transform personally owned smart electric vehicles into utility grid assets, by having electricity flow from a charged electric vehicle onto the utility grid.

AI could also be used to streamline the EV charging station connection process (e.g., capacity check before charger installation or utility and contractor schedule management). [4, page 43 summarized]

Artificial intelligence could analyze the voluminous data generated by EV driving and charging patterns. The usage data could be used to generate customized charging profiles and personalized targeted load management programs. [4, page 44]

PG&E can use the AMI [Advanced Metering Infrastructure] data from more than 11 million Smart Meters for time-series capacity planning[4, page 50] and grid optimization. AI can analyze historical data and with future demand forecasts, can efficiently recommend changes to grid operation.[4, page 52] AI can recommend optimal interconnection points and capacity upgrades.[4, page 53]

Artificial Intelligence could revolutionize balancing supply and load (demand), thereby driving significant progress in grid management and optimization. AI could enhance real-time visibility, optimize resource orchestration [making use of Distributed Energy Resources], and improve grid reliability. [4, page 61]

AI-driven analytics could process large amounts of data from distributed energy resources, smart meters, and various sensors, driving more accurate forecasts both of demand and grid flexibility. AI's real-time analysis of complex data means that the grid could responsively adapt to fluctuations in supply and demand.[4, page 61]

Wildfire

More than half of PG&E's service area and approximately one-third of its equipment assets are located within High Fire-Threat Districts (HFTDs).[3, page 74]

Key initiatives in the area of wildfires are:

- Monitoring equipment and other assets
- Vegetation management (around power lines),
- Managing risk from external conditions such as high winds using proactive power line de-energization schemes, and
- Identifying and suppressing any ignitions that do occur. [4, page 75]

AI-enabled cameras for early ignition detection

In partnership with ALERTWest, PG&E has already deployed over 600 AI-enabled high-definition (HD) cameras over the past several years in High Fire-Threat Districts (HFTDs). The cameras enhance situational awareness in the most vulnerable geographies, enabling fire suppression agencies to respond quickly.[4, page 77]

The AI wildfire detection capabilities in this HD camera network provide alerts to Hazard Awareness and Warning Center (HAWC) analysts when potential evidence of a fire is first detected. The AI technology more accurately identify ignitions by filtering out false positives from sources like dust, fog, or haze. AI's processing speed can reduce the time between ignition and response by fire agencies. Because fire agencies can see the AI detections in real-time, it improves their response time, potentially resulting in a significant reduction in the impact on people and property.[4, page 77]

Eliminate and rapidly suppress ignitions

> **Note**
>
> The goal is to have fail-safe equipment, meaning if it fails, it fails in a safe way.

In addition to eliminating ignition risk, it is also important to have real-time alert and notification capabilities to enable rapid response to developing threats. PG&E's Hazard Awareness & Warning Center (HAWC) supports 24/7 monitoring of risks across its systems.[4, page 81]

Climate Resilience

Research and development, including applications of artificial intelligence and machine learning, are needed to accelerate climate adaptation capabilities in the following areas:

- **Optimizing infrastructure planning for climate impacts**: AI/ML-driven models can analyze climate and utility operational data to better predict potential risks to utility assets. The utility can take preemptive measures to protect vulnerable infrastructure and communities.
- **Optimizing field team operations:** AI can help anticipate the timing, location, and severity of extreme climate-driven hazards with better precision. With more accurate, predictive information, the utility can prepare and deploy resources for emergency responses. Potential advancements provided by AI/ML can improve immediate response capabilities and also inform long-term planning and policymaking.[4, page 100]

Wildfire Mitigation Plan Executive Summary

The executive summary provides a few glimpses of the use of Artificial Intelligence in Wildfire Mitigation. They will be summarized first before getting into the details of the massive two-volume Wildfire Mitigation Plan.

In the section on Risk-Informed Decision Making, the summary refers to a risk modeling tool called the Wildfire Distribution Risk Model (WDRM) which has gone through various upgrades.

- WDRM v2 used more sophisticated algorithms, machine learning, and physics-based fire simulation outputs mapped into fire size/severity tranches to quantify wildfire consequence.[3, page 6]
- WDRM v3 used more-advanced machine-learning modeling techniques, incorporated improved and updated data, added predictions of wildfire risk reduction when mitigating various sources of risk, and was enhanced to understand additional ignition sources and sub-drivers.[3, pages 6 and 7]

In the section entitled Comprehensive Monitoring and Data Collection (referencing WMP Section 8.3) the summary states:

> "we discuss our Situational Awareness and Forecasting objective to enable Artificial Intelligence (AI) processing of Wildfire Camera Data to provide automated wildfire notifications in the internal PG&E monitoring tool (Wildfire Incident Viewer – WIV). Early detection of new ignitions can help reduce the overall impact of the ignition through increased awareness and more rapid response." [3, page 11]

Wildfire Mitigation Plan Details

This section looks at the voluminous Wildfire Mitigation Plan (WMP) with 1,086 pages across two volumes, with an eye to capture the stated current and future efforts that make use of Artificial Intelligence and Machine Learning.

The 2026-2028 WMP uses the Wildfire Distribution Risk Model (WDRM) uses Machine Learning (ML), combining enhanced event probability assessments and new models for several equipment classes.[1, page 5]

One of the components of PG&E's wildfire ignition risk modeling is "LoRE – Transmission and Distribution".

Definition: LoRE

LoRe is short for Likelihood of Risk Event. LoRE, for wildfire risk models, is the probability of an ignition, or p(i), event resulting from the failure of a distribution or transmission asset during a fire season.[1,page 56]

Separate probabilities of ignition models are produced for transmission and distribution-related ignitions in order to account for differences in causal risk drivers. Both the transmission and distribution p(i) models use causal models to reflect the various threats that cause asset degradation, as well as the external hazards that can cause an electrical asset failure. Their implementations are different, as follows:

- the transmission p(i) model using a combination of engineering first-principle calculations and machine-learning algorithms**,** whereas
- the distribution p(i) model uses machine-learning algorithms only.[1, page 56]

Model Summary

Transmission Ignition Likelihood Calculation

As mentioned earlier, the transmission probability of ignition model uses a mix of first principle and machine learning (ML) probability of failure causal models.[1, page 61]

First principle causal models use "fragility curves optimized for a specific threat."[1, page 62] ML causal models estimate probability of failure in the same manner as the distribution probability models.

Inputs to the Models:

- Asset data (feeds Failure Probability Hazards Model and Failure Probability Threats model)
- Meteorology (feeds Failure Probability Hazards Model and Failure Probability Threats model)
- Vegetation (feeds Failure Probability Hazards Model)
- Event Data (feeds Failure Probability Hazards Model and Failure Probability Threats model)
- Maintenance Tags (feeds Failure Probability Threats model)

Models

- Failure Probability Hazards
- Failure Probability Threats

Intermediate Results (using feeds from the two models)

- Failure Probability Causal Results

Outputs

- Ignition probability

PG&E will evaluate next generation SmartMeter technology's capabilities for wildfire risk reduction and pursue a path to production rollout if that technology is found to be viable. These SmartMeter platforms will combine edge computing, real-time analytics, and machine learning. This will enable distributed intelligence-based applications to make scalable grid-edge decisions that can identify anomaly signatures consistent with potential ignitions.[1, page 444]

Fire Detection Systems

Early fire detection is provided by two systems:[1, page 448]

- PG&E FDAS Satellite Infrared Imagery and
- Wildfire Cameras with AI and smoke detection capabilities

Satellite Infrared Imagery

PG&E uses infrared imaging data from six satellites that are operated by National Oceanic and Atmospheric Administration (NOAA). The Space Science and Engineering Center (SSEC) processes the data and makes it available to PG&E with an update cadence of every 5 minutes.[1, page 448] Satellite fire detection provides information quickly about new fires and the spread of existing fires.

Land-based wildfire cameras with AI capability

PG&E worked with multiple camera vendors to AI to help detect new fires and enhance situational awareness. The cameras are equipped with AI technology where smoke can be detected using AI algorithms.[1, page 448]

PG&E has sponsored over 600 wildfire cameras on the Alert California network since 2019. Camera detections provide information about new fires and the spread of existing fires. These cameras are being used by first responders including CAL FIRE.[1, page 448] These cameras cover over 90% of the HFTD tier 2 and 3 areas.[1. page 447]

AI software is used to analyze and learn image characteristics (e.g., smoke location and color, direction of smoke column, etc.) that may indicate the presence of fire in an area.[1, page 449] When AI detects potential new fires, notifications to the utility and first responders can occur more quickly than relying solely on other means of detection. With the AI system, updates occur every ten seconds.[1, page 449]

The cameras use an encrypted, secure communications connection to ensure image integrity from the camera to the remote viewer.[1, page 449]

Evaluation and Selection of New Detection Systems

The evaluation of new systems follows this process:

1. Identifying new technologies or systems;
2. Conducting a Subject Matter Expert (SME) and business review for reasonableness and feasibility;
3. Evaluating alignment to program goals and objectives;

4. Benchmarking (if applicable);
5. Determining a source of funding and cost review;
6. Performing a pilot study (if needed), to evaluate effectiveness at achieving program goals and objectives and testing assumptions;
7. Implementing and deploying the technological system; and
8. Conducting a Subject Matter Expert (SME) operational review of benefits.[1, page 450]

Note

This process might be useful when we examine project management considerations later in this chapter.

Weather forecasting

Model Overview

PG&E builds, operates, and maintains core models and datasets used to train machine learning models and to forecast PSPS (Public Safety Power Shutoff) events.[1, page 452]

PG&E's Operational Mesoscale Modeling System (POMMS) is based on the Weather Research and Forecast (WRF) model, which is a mesoscale numerical weather prediction system designed for both atmospheric research and operational forecasting.[1, page 453]

Definition: Term

Mesoscale meteorology is the study of weather systems and processes at horizontal scales of approximately 5 kilometers (3 mi) to several hundred kilometers.[7]

PG&E first deployed the high resolution in-house mesoscale forecast model, POMMS, in November of 2014 and PG&E continues to improve the model.[1, page 453]

POMMS is a high-resolution weather forecasting model that generates important fire weather parameters including wind speed, temperature, Relative Humidity (RH), and precipitation. Outputs from POMMS are used as inputs to additional models to derive key fire danger indicators.[1, page 453]

Fire Potential Index (FPI) Models

The FPI 3.0 model was produced in 2019. This model coupled weather and fuels data around the ignition of each fire catalogued in the USFS's Fire Program Analysis – Fire-Occurrence Database (FPA-FOD). This was the first iteration of a machine learning model that used historical fire occurrence data with a logistic regression framework. The goal was to create an FPI model that could predict, based on forecasted weather and fuels conditions, the probability of a large fire given an ignition, instead of an FPI index value related to the risk.[1, page 464]

Definition: FPI

FPI is Fire Potential Index[1, page 14]. The FPI model classifies fire potential according to potential size in acreage as follows:

- Small: approximately 300 acres;
- Large: approximately 1,500 acres;
- Critical: 20,000 acres; and
- Catastrophic ~80,000 acres. [1, page 468]

The FPI 4.0 model was used from August 2021 through July 31, 2024. It used a novel machine learning framework, additional model features and a fire occurrence dataset developed by Sonoma Technology. Among the model-types tested were logistic regression and multiple machine-learning classification model types.[1, page 465]

The FPI 5.0 model was developed in 2022/2023 and approved for operations starting August 2024. It expanded model training data to use all fire detections rather than only the first fire detection. The model weights earlier fire detections more than later detections.[1, page 465]

The FPI model is based on a multi-classification balanced random forest framework, a state-of-the-art open-source machine learning model based on decision trees. The FPI model is trained on a fire occurrence dataset (McClure et. al., 2023) that combines agency fire information with satellite fire detections. Fire detections are derived from satellite infrared data and provide the following information: the location, intensity and time of fires.

FPI v5.0 was trained on satellite fire detections using defined classes that separate small, moderate, critical, and catastrophic fires.[1, page 466]

References

1. Pacific Gas and Electric Company. PG&E Wildfire Mitigation Plan R0 2026-2028, Volume 1 of 2, April 4, 2025. PG&E Wildfire Mitigation Plan R0 2026-2028, Volume 1 of 2.
2. Pacific Gas and Electric Company. PG&E Wildfire Mitigation Plan R1 2026-2028, Volume 2 of 2, July 28, 2025. PG&E Wildfire Mitigation Plan R1 2026-2028, Volume 2 of 2.
3. Pacific Gas and Electric Company. Pacific Gas and Electric Company 2023-2025 Wildfire Mitigation Plan Executive Summary. PG&E 2023-2025 Wildfire Mitigation Plan Executive Summary.
4. Pacific Gas and Electric, R&D Strategy Report 2024. PG&E R&D Strategy Report 2024.
5. Office of Energy Infrastructure Safety. Office of Energy Infrastructure Safety - Wildfire Mitigation Plans.
6. California Public Utilities Commission. Implementation of Wildfire Mitigation Plans. California Public Utilities Commission - Implementation of Wildfire Mitigation Plans.
7. Wikipedia authors. Mesoscale meteorology. Accessed August 29, 2026.

This page titled 7.3: Artificial Intelligence Use by a California Utility (PGandE) is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

7.4: AI Initiatives by other California Utilities

We have just covered PG&E AI initiatives. In this section we will cover AI initiatives at other California utilities:

- Southern California Edison (SCE) and
- San Diego Gas and Electric (SDGE).

This page titled 7.4: AI Initiatives by other California Utilities is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

7.4.1: Southern California Edison

In this section, we are going to look at Southern California Edison's (SCE) Wildfire Mitigation Plan[1] to get some insights into how they have already deployed machine learning and artificial intelligence and look at some of their statements about their future plans. We will see that a lot of their efforts are in deploying artificial intelligence for improving weather forecasting, both from an accuracy and timeliness perspective. They are also focused on efforts to mitigate any contributions to wildlife problems from their own equipment.

We are going to do this by inspecting the details of the SCE document[1] to simulate what you might be doing if you were to be hired as a project manager there and wanted to get up to speed quickly and gain context on their artificial intelligence efforts in preparation for running an artificial intelligence project.

The Wildfire Mitigation Plan [1] is 751 pages in size, so having these summarized extracts is intended as a time-saver when compared to what you might be confronted with when you start your background reading on the job.

Artificial intelligence is currently used, or planned to be used in the following areas:

- Risk analysis.
- Equipment analysis
- Weather forecasting
- SMART meters
- Fire Detection Systems

Risk Analysis

In a discussion about risks, SCE stated that it used machine learning to determine the topographical and climatological risk factors used for short-term and current risk modeling. [1, page 17]

Equipment Analysis

In discussing the likelihood of equipment contributing to a wildfire either through operating or failure, SCE makes use af various sub-component models that use machine learning (ML) algorithms to assess the relevance of individual ignition drivers associated with the sub-component in calculating a probability of ignition (POI).

Sub-component models include:[1, pages 71-72]

- EFF: Equipment Facility Failure. These equipment models use historical equipment outage data (outcomes), current equipment condition attributes (e.g., age, voltage, inspection results, etc.), and relevant environmental attributes (e.g., historical wind, asset loading, number of customers, temperature, relative humidity etc.). The historical attributes and outcomes can be used as a training data set in machine learning.
- CFO-Veg: Contact from Object – Vegetation. SCE uses machine learning algorithms to develop predictive models for vegetation contact.[1, Table 5-1 SCE Risk Modeling Assumptions and Limitations, page 104]
- CFO-Other: Contact from Object – Other.

SCE uses machine learning algorithms to develop predictive models for equipment failure. [1, Table 5-1 SCE Risk Modeling Assumptions and Limitations, page 104]

Weather Forecasting

In a discussion about the use of weather data, SCE says that it uses machine learning algorithms to associate applicable weather variables from the weather research and forecasting (WRF) model at the time of fault/ignition events, using 2009-2023 weather data.[1, Table 5-1 SCE Risk Modeling Assumptions and Limitations, page 103]

The output from the weather prediction systems is used as as input to the machine learning modules. The machine learning

modules and the forecast output is presented to Subject Matter Experts (SMEs). As of the end of 2024, SCE deployed machine learning to 1,624 weather station locations throughout SCE's service area.[1, page 415]

SCE's fuel moisture modeling makes use of SCE's weather forecast output in conjunction with mathematical algorithms in order to estimate dead fuel moisture. SCE developed a machine learning model trained on historic weather and fuels data to predict live fuel moisture levels.[1, page 415]

SCE has five machine-learning-based weather forecast systems to augment traditional weather forecast output. SCE has developed machine learning models for up to 1,624 weather station locations using internal weather model data and up to 1,947 weather station locations from public forecast data.[1, page 416]

The initial machine learning forecast corrections have been expanded beyond the sustained wind speed and gust speed predictions to include temperature and dew point depression variables. Existing machine learning models have been retrained each year to learn from recent weather events. [1, pages 416-417]

University of California Santa Barbara (UCSB) also developed a deep learning model to translate irregularly spaced weather station observations into a regularly spaced grid with 500-m grid length, providing insight into the real-time weather conditions over locations that do not have weather station coverage. SCE is using UCSB's Deep Learning Gridding Meteorological Model (DLGNOME).[1, page 417]

SMART Meters

Like PG&E, SCE uses Smart Meters and can utilze data from them to detect evidence of transmission wires being down and can use data to help predict the need to replace transformers. For example:

- SCE's Meter Alarm Down Energized Conductor (MADEC) is a machine-learning (ML) algorithm that uses smart meter data to detect a subset of energized wire-downs and other high-impedance faults/hazards. MADEC can generate an alarm that allows SCE to respond quickly to manually de-energize the line or circuit.[1, page 396]
- SCE's Transformer Early Damage Detection (Transformer EDD) uses SMART meter data that is processed by a custom algorithm to proactively identify one failure mode of distribution transformers. Identified transformers can be scheduled to be replaced before potential failure in order to reduce safety hazards for the public, prevent grid disruptions, and power outages.[1, page 396] Winding shorts, partially turn-to-turn shorts, can create small increases in voltage on a transformer secondary that can be detected by SMART meters. By comparing voltage data of other transformers in the vicinity, SCE can create schedules for equipment replacement maintenance actions for some transformers prior to failure, thereby reducing ignition risks due to equipment failure and limiting the effects of electric service outages to customers.[1, page 678]

Smart meters collect a large amount of data. When coupled with other data, SCE can generate alerts for various inspections that might be needed.

Fire Detection Systems

SCE uses several types of fire detection systems:

- High Definition (HD) cameras and
- Satellite and other imaging technology.

HD cameras

HD cameras are mounted on available infrastructure such as on communications towers. SCE makes use of a a network of 200 cameras with live data feeds, using wireless internet service providers.[1. page 405]

HD cameras can be used with Artificial Intelligence and with satellite imagery to confirm fires. The AI software runs on ALERTCalifornia cameras.

SCE has partnered with UCSD and thus has full access to AI that uses the camera data feeds to generate an alert associated with a specific camera location, enabling human follow-up assessments in real-time about the conditions of the detected fire (i.e., location, growth potential, nearby SCE assets, possible communities in danger, fire department resource deployment).[1, page

405]

References

1. Southern California Edison (SCE), 2026-2028 Wildfire Mitigation Plan, May 16, 2025, Revision 0. 2026-2028 Wildfire Mitigation Plan.

This page titled 7.4.1: Southern California Edison is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

7.4.2: San Diego Gas and Electric

This section examines projects and use cases relevant to this case study at San Diego Gas and Electric (SDGE) as described in their 1,020 page 2026-2028 Wildfire Mitigation Plan[1].

Some initiatives that are mentioned address wildfire mitigation, but there are no specific references to the use of artificial intelligence include:

- Grid design, including under-grounding.
- Operational changes to equipment use to reduce the chance of a risk event occurring that could ignite a fire.
- Inspection programs focusing on safety and reliability, including the use of drones to conduct inspections to drive preventive maintenance.

Initiatives that specifically mention using artificial intelligence include:

1. Customer communications and
2. Weather forecasting.

Customer communications

One potential use case that is called out is a future use of mobile applications for the utility to communicate with its customers who own property where the utility wants to install equipment such as transformers and junction boxes. The utility would make use of artificial intelligence and machine learning to show what the customer's property would look like after the installation of the equipment using an "augmented reality visual representation."[1.p19]

Weather forecasting

SDGE weather forecasting efforts already make use of artificial intelligence. AI models were trained using over 15 years of weather data. The following were specificalled called out:[1,p276]

1. The Machine Learning Wind Gust Model for the High Fire Threat District (HTFD). A three day forecast is generated two times per day using as inputs the latest weather forecasts.
2. The Fire Potential Index (FPI) seven day forecast based on weather conditions and fuel conditions.
3. The Santa Ana Wind Threat Index (SAWTI) daily forecast that rates Santa Ana wind events.

The AI-based forecasting system was trained using weather observations collected over 217 weather stations very ten minutes over 15 years. The data collection continues each year to enable improving and calibrating the models.

Definition: HFTD

High Fire Threat District [1, page xvii]

Definition: FPI

Fire Potential Index [1, page xvi] It has values such as Normal, Elevated, and Extreme.

Definition: SAWTI

Santa Ana Wind Threat Index rates Santa Ana wind events. the U.S. Forest Service issues this on a daily basis.

> Wildfire Management Plan [1, page xix]

The Santa Ana Wind Threat Index assess the potential effect of the winds on wildfire activity using the strength, extent, and duration of the wind, dryness of the air, dryness of the vegetation, and greenness of the grasses.[1] The SAWTI rating use this scale:[2]

- No Rating: Winds are either not expected, will not contribute to significant fire activity.
- Marginal: Upon ignition, fires may grow rapidly
- Moderate: Upon ignition, fires will grow rapidly and will be difficult to control
- High: Upon ignition, fires will grow very rapidly, will burn intensely, and will be very difficult to control.
- Extreme: Upon ignition, fires will have extreme growth, will burn very intensely, and will be uncontrollable.

References

1. San Diego Gas and Electric (SDGE). Wildfire Mitigation Base Plan 2026-2028, July 18, 2025, version R1. Wildfire Mitigation Base Plan 2026-2028.
2. SAWTI scale. Accessed August 26, 2026.

This page titled 7.4.2: San Diego Gas and Electric is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

7.5: Applying Artificial Intelligence to Wildfire Detection and Management

We have seen how electric utilities are currently using and have future plans for using artificial intelligence in a variety of applications. One of the most pressing applications is for wildfire detection and management.

On a smaller scale, some communities in areas prone to wildfires have begun to investigate making use of wildfire detection equipment, some of which makes use of artificial intelligence.

For example, The Mercury News published a story from Phil Diehl of The San Diego Union-Tribune that described how the city of Carlsbad, California is looking to deploy cameras whose feeds would be analyzed by artificial intelligence to detect potential wildfires and notify the appropriate authorities when a fire is detected. [1]

The University of California, San Diego, one of the partners cited by the large California utilities in the previous sections, developed the wildfire monitoring system called AlertCalifornia in 2023. On its website[3], one can view feeds from a multitude of cameras across the state.

"ALERTCalifornia has more than 1,144 high-definition, pan-tilt-zoom cameras deployed across California (as of January 2025), providing a 24-hour backcountry network with near-infrared night vision to monitor disasters such as active wildfires. ALERTCalifornia cameras can perform 360-degree sweeps approximately every two minutes and can view as far as 60 miles on a clear day and 120 miles on a clear night."[3] The cameras use Forward Looking Infrared (FLIR) technology.[3]

“During its first season, the program detected more than 1,200 fires across the state, beating 911 calls to report the fires in 30% of the incidents, the AlertCalifornia website states.”[1]

ALERTCalifornia is a partnership between UCSD, the wireless internet company DigitalPath and CalFire. It makes us of artificial intelligence to analyze camera feeds and notify emergency responders when fires are detected.[1]

"A machine-learned model has been created to accurately differentiate between smoke and other particles in the air."[1]

Use of Artificial Intelligence

The ALERTCalifornia website explains its use of Artificial Intelligence as follows:

> "ALERTCalifornia collaborated with the California Department of Forestry and Fire Protection (CAL FIRE) and industry partner Digital Path to create a fire detection AI tool with the goal of improving firefighting capabilities and response times. The value of this public-private partnership is the development of AI to aid firefighters, mitigate watchstander fatigue, reduce false positives, and confirm fire incidents in the incipient phase.
>
> When the AI spots a potential fire on ALERTCalifornia’s network of cameras, the system alerts firefighters and provides a percentage of certainty and estimated location for the incident. If the incident is vetted and confirmed by trained watchstanders, firefighters respond quickly to extinguish the fire at the incipient phase. The camera network is also used to provide actionable real-time data to quickly scale fire resources, help evacuations through enhanced situational awareness, and monitor fire behavior.
>
> Early detection and rapid response allow firefighters to combat fires before they grow. The AI tool became available to all 21 CAL FIRE 911 Dispatch Centers in September 2023. This new tool is especially effective in spotting anomalies in remote locations and is proven effective at night, even alerting firefighters before 911 calls."[3]

CalFire has invested around $24 million in the program.[1]

A second The Mercury News article “AI-boosted smoke-sniffing wildfire sensors approved for Santa Clara County” is of interest for a few reasons. De Anza College is located in Santa Clara County. Additionally, the article makes brief mention of how Artificial Intelligence is used.

> “Made in the U.S. by Maryland company N5, the solar-powered devices detect gases and particulates that smoke produces. **Artificial intelligence 'machine learning'** software analyzes the detections to determine the likelihood that a wildfire is present.”[2]

> “When a possible fire is identified, a device transmits a warning to local fire authorities, using whatever cell service has the strongest signal. If the chemical signatures of fire are strong enough to cross a threshold, the message to authorities is an alert, denoting a higher priority for response. If cell service is unavailable, as can occur when fires damage infrastructure, a backup Starlink satellite link is used.”[2]

Each N5 sensor costs about $5,000 to $7,000, with pricing dependent on whether a wind meter is included, and the status of tariffs.[2]

We will look at the N5 sensor offering in the next section.

A third The Mercury News article[4] updated the number of camera in use to 1,211 and mentioned the cameras were using artificial intelligence software as of 2023 developed by DigitalPath. "To address **privacy concerns**, UC San Diego blurs homes, parking lots and other nearby features that could track people in real time."[4] The cameras "send a message to dispatch centers across the state, which is then verified by humans."[4] In 2024 "there were 7,553 wildfires in Cal Fire’s jurisdiction. Of those, 1,668 were picked up by the cameras...and 38%, or 636, were detected by the cameras before any person called 911 to report them."[4]

DigitalPath

The DigitalPath-developed software is powered by AI trained on NVIDIA GPUs. "DigitalPath has refined a **convolutional neural network** to spot signs of fire in real time."[5] "The team began by training a **convolutional neural network** on a cloud-based system running an NVIDIA A100 Tensor Core GPU and later transitioned to a system running on eight A100 GPUs." [5]

Definition: Convolutional neural network

A **convolutional neural network** (**CNN**) is a type of feedforward neural network that learns features via filter (or kernel) optimization. This type of deep learning network has been applied to process and make predictions from many different types of data including text, images and audio. Convolution-based networks are the de-facto standard in deep learning-based approaches to computer vision and image processing.[6]

"One of DigitalPath’s breakthroughs is its system’s ability to identify the same fire captured from diverse camera angles. DigitalPath’s system efficiently filters imagery down to a human-digestible level. The system filters 8 million daily images down to just 100 alerts, or 1.25 thousandths of one percent of total images captured."[5]

Architectural Considerations

The wildfire cameras and sensors need to be placed where they can be operationally effective. Cameras are often placed at mountain-top observatories, buildings, and stand-alone towers with good line of sight. They also need sources of power that might be available at these structures, or may require solar power with battery back-up if they are located in more remote locations.

An important consideration is where the bulk of processing is going to occur. If the bulk of the processing is going to occur close to the point of data collection, the computational power and electrical power budgets are going to be high, but the amount of data to be transferred to a central facility will be reduced. If the bulk of the processing is going to occur at a central processing facility, the local computational and power budgets are going to be lighter, but the data transmission requirements will be higher. In remote locations, having a reliable data transmission mechanism might be difficult, and likely would require consideration of having a back-up communications mechanism (e.g., a satellite connection).

Another consideration is when an alert should be sent to the appropriate authorities and by whom. Should the sensor or camera send that alert? Should the alert be sent after processing in a more centralized processing center, perhaps with human oversight?

In the next section, we will take a look at some commercially available products and see what design choices they have made.

References

1. Diehl, Phil. “City Looks into camera that detect wildfires”, The Mercury News, Tuesday July 29, 2025, Section B, Page 5. Note: Phil Diehl is at The San Diego Union-Tribune.
2. Baron, Ethan. “AI-boosted smoke-sniffing wildfire sensors approved for Santa Clara County”. The Mercury News. May 23, 2025.
3. ALERTCalifornia. ALERTCalifornia.
4. Rogers, Paul. The Mercury News (online version), September 19, 2025. AI cameras spot wildfires across California, often before humans call 911
5. Caufield, Brian. How AI Helps Fight Wildfires in California. October 4, 2023. How AI Helps Fight Wildfires in California | NVIDIA Blog
6. Convolutional neural network - Wikipedia. Text is available under the Creative Commons Attribution-ShareAlike 4.0 License.

This page titled 7.5: Applying Artificial Intelligence to Wildfire Detection and Management is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

7.5.1: Vendor Solutions

In this section we take a very cursory look at several vendor solutions in the wildfire and sensor space that may be of interest to electric utilities as they pursue their wildfire mitigation efforts:

- N5 Sensors
- Ladris
- Pano.ai.

ALERTWest

We will also take a brief look at AlertWest.

ALERTWest originated from the private and public partnership between Digital Path (AlertWest's parent company) and UC San Diego who created the ALERTCalifornia platform discussed elsewhere in this textbook. ALERTWest was subsequently created to bring the same model of partnership to a broader geography.[1]

ALERTWest uses daytime AI models and night time AI models to detect wildfires.

Although AI is used, humans are in the loop in the ALERTWest Operations Center to review the AI detected fire and confirm it or designate it as a false wildfire detection.

References

1. ALERTWest. ALERTWest | AI Wildfire Detection & Situational Awareness.

This page titled 7.5.1: Vendor Solutions is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

7.5.1.1: N5 Sensors

In this section we will be surveying the product offerings of N5 Sensors in the area of Wildfire Detection and Management, in particular the incorporation of any Artificial Intelligence capabilities. We will examine material posted on their web site and examine any relevant White Papers with appropriate attribution in the References below.

N5 Sensors offers an advanced wireless sensor network for early wildfire detection and 24/7 alerting. Their real-time early warning system can provide alerts in less than 5 minutes from time of ignition).[3]

N5SHIELD™ is N5's name for its environmental intelligence platform that integrates environmental sensors (N5's and from other providers), geospatial information, and other data, and "implements different AI models to detect environmental threats and hazards."[3] N5 detects wildfires using smoke and heat, often using multiple sensors to identify different anomalies.

We now look at N5 Sensors' White Paper "Machine Learning Models and Their Effectiveness for Autonomous Detection of Wildfires Utilizing Ground-Based Sensor Networks" to get details of their approach that incorporates artificial intelligence, machine learning, and predictive analytics.[1]

The white paper does a good job addressing the architectural tradeoffs discussed earlier in this textbook of "various algorithmic approaches keeping in mind the resource requirements, namely energy, data bandwidth, implementation cost and operational requirements such as detection speed, false alarms, etc."[1]

There are three major categories of sensing and early-detection:

1. satellite-based
2. camera-based, and
3. ground-based sensor networks.

The white paper discusses innovative ML-based methods employed in the third category: ground-based sensor networks (GSNs). Ground-based sensor networks typically are composed of many low-cost and low-power sensor nodes connected to a central server. These sensors collect environmental data (e.g., temperature, wind speed, particulate matter) and transmit the data to a centralized location via wither a wired or wireless communications channel.

In some architectures, all the data collected by the sensors is sent to a centralized server. At the central server, analysts can see the data and process it or computer applications can process the data. In other types of sensor networks, there might be some computation handled at the network edge before transmission to a central site. As we know from edge computing in general, doing processing at the edge can reduce the amount of data and information uploaded to the centralized processing center.[1]

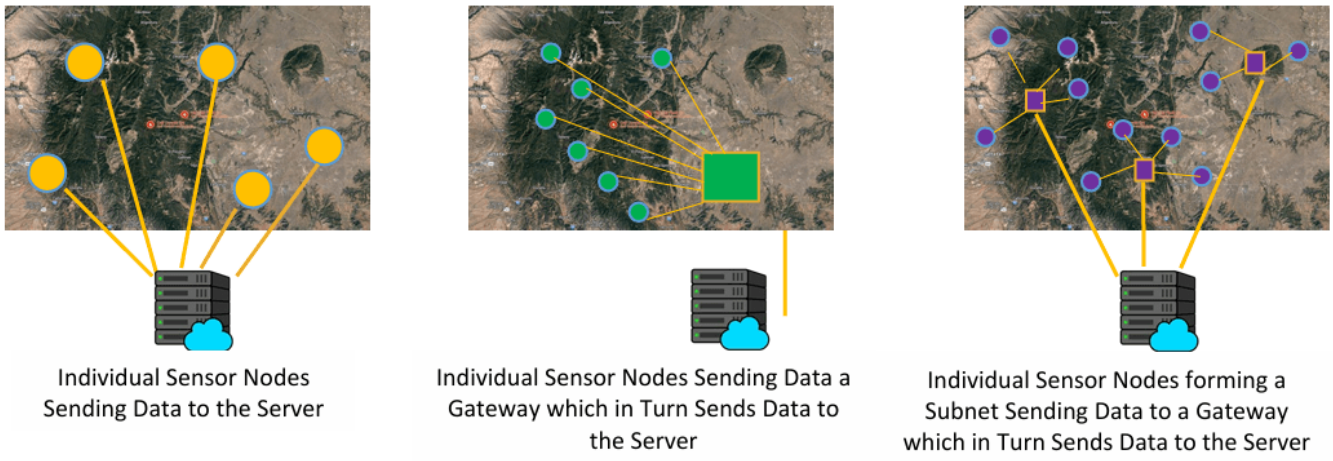


Figure 7.5.1.1.1: Sensor Architectures. (Reference 1, Home - N5 Sensors N5 Sensors, white paper "Machine Learning Models and Their Effectiveness for Autonomous Detection of Wildfires Utilizing Ground-Based Sensor Networks.")

Accessibility Note

Figure 7.5.1.1.1 shows three different topologies of sensor networks:

1. Individual Sensor Nodes Sending Data to the Server.
2. Individual Sensor Nodes Sending Data a Gateway which in Turn Sends Data to the Server.
3. Individual Sensor Nodes forming a Subnet Sending Data to a Gateway which in Turn Sends Data to the Server.

"In all these topologies, edge processing can be implemented at the sensor node or at the gateway layer and can be combined with cloud-level processing. Different architectures represent tradeoff in latency, data load, response time, and false alarm rates. In topologies where data from each node (e.g., camera) is sent to a central location, high bandwidth communication links are needed. However, often areas prone to fire do not have high bandwidth connectivity solution such as LTE. Sensor networks that use edge processing can alleviate high data transmission requirements."[1]

To achieve fully autonomous 24/7 fire detection capability, a combination of machine learning (ML) methods using cameras and other ground-based environmental sensor systems can be used with low false alarm rates.[1]

"From [a] machine learning perspective, autonomous identification of wildfires is an **anomaly detection problem**."[1] Sensor nodes at different locations gather environmental measurements (such as temperature, air pressure, humidity, wind speed, wind direction, heat/infrared signature, chemicals, particulate matter, soil moisture level), which normally would fall within certain ranges or thresholds, and measurements outside of these ranges would be classified as anomalies. "The purpose of ML models is to correctly identify these thresholds in various terrains, topographies, and macro environmental conditions and adjust accordingly to improve detection accuracy and reduce false alarms."[1]

"Deploying such wildfire detection sensor networks with ML represents the need to find energy and data efficient computing in a resource-constrained environment. With such networks measuring real-time data and fusing them with short-term and historical environmental and fuel data, we can train the ML models to not only detect but also predict the probability of ignition." The types of short-term and historical environmental and fuel data mentioned include:

- Average Rainfall for last 30 - 60 Days
- Average Humidity for last 30 - 60 days
- Average High Temperature for the last 10 days
- Fuel Type
- Fuel level[1]

"Measurement and utilizing of such different environmental parameters for detection of wildfires is a case of **multimodal sensing**. N5's wildfire detection product N5SHIELD™ uses such multimodal sensing technology."[1] A look at the product that is deployed in the field shows that it is a robust unit that integrates these various sensor units.

In the N5SHIELD™ system, the edge algorithm uses three separate sensing modalities:

1. "**Gas Sensing arrays** – An array of gas and chemical sensors responsive to chemicals released during combustion. A change point detection algorithm is used to measure the overall array response magnitude in the presence of a changing environment. Wildfires result in increased rate of change in these chemical concentrations in the air."[1] The N5 algorithm quantifies this rate of change.
2. **"Particle Sensors** – An optical-scattering based module measures a mixture of particulate sizes and their concentrations. Thresholding as well as change point detection allow us to identify particulate mixtures and behavior that are consistent with wildfire spread.
3. **IR Heat Sensor** – An Infrared sensor capable of measuring surface temperature provides a heatmap of the landscape. Changes and growth in hot spots are quantified as part of the edge detection algorithm."[1]

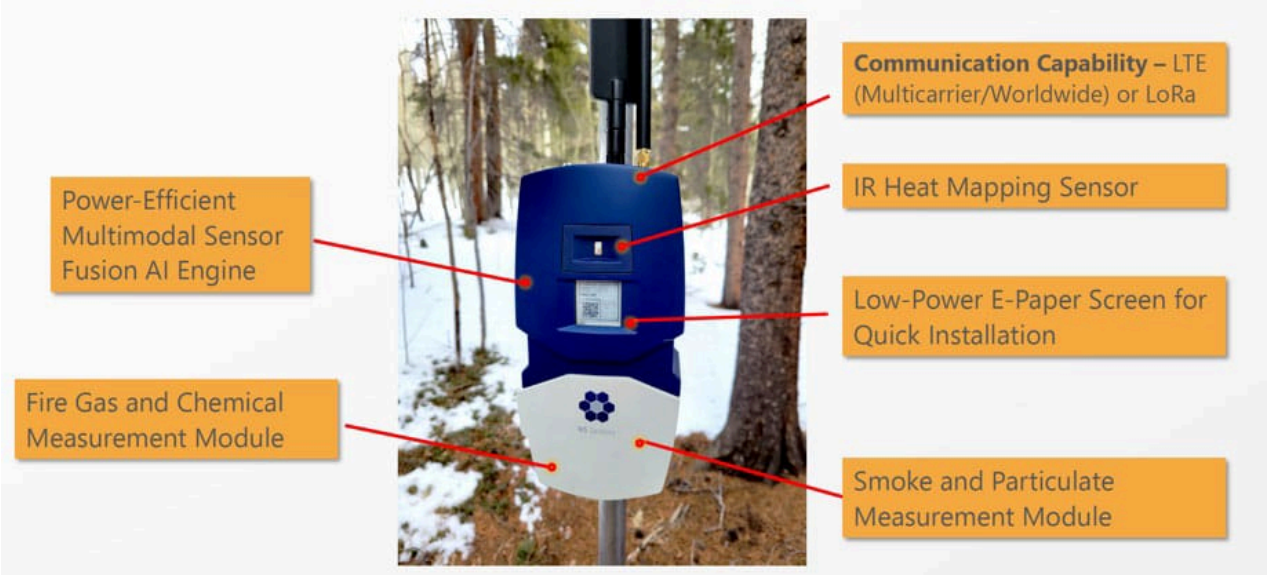


Figure 7.5.1.1.2: Field Deployment. (Reference 1, Home - N5 Sensors N5 Sensors, white paper "Machine Learning Models and Their Effectiveness for Autonomous Detection of Wildfires Utilizing Ground-Based Sensor Networks.")

Accessibility Note

Figure 7.5.1.1.2 depicts a field-deployed product that integrates the following sensor unit assemblies:

- Multi-modal Sensor Fusion AI Engine
- Fire, Gas, and Chemical Measurement Module
- Communications capability module LTE (Long Term Evolution) or LoRa (**Lo**ng **Ra**nge)
- IR (infra red) Heat Mapping Sensor
- Operator screen used during installation
- Smoke and Particulate Measurement Module

Definition: LoRa

Stands for **Lo**ng **Ra**nge. LoRa is a physical proprietary radio communication technique. It is based on spread spectrum modulation techniques derived from Chirp Spread Spectrum (CSS) technology.[4]

It should not be confused with the Machine Learning Model Low-Rank Adaptation (LoRA) that is a parameter-efficient fine tuning method, wherein instead of updating all weights in a model, LoRA inserts low-rank matrices into certain layers.

"The edge algorithm provides quantified severity measures for these modalities which are subsequently used in a network based deep-learning aggregation model to provide robust wildfire likelihood estimation."[1]

Using a group of nodes organized into a local subnet, the N5 Multi-Modal Deep Learning Architecture takes a fixed cluster of N Node features as input. Subsequent layers (Graph Convolution, Gated Recurrent Unit) extract spatial (the node's geographic coordinates) and temporal information (how the gas, particle, and IR values change over time as a function of the sensor Node's spatial orientation). "The model output is a predicted probability of active fire conditioned on a local cluster of Node units within a larger group deployed over a wider geographical area. The model is run simultaneously on disjoint or overlapping clusters of Node units. This allows for scalability across small or large deployments containing different numbers of deployed sensor Node units. A heat map of fire probabilities is produced and can be visualized in the cloud monitoring software and used to trigger an active fire alert if a probability threshold is exceeded.[1, page 6]

"The output of this network is a probability of fire detected by the input cluster of nodes, which also allows for localization of the fire event. It is an aggregation function of the node embeddings in the final layer. The framework provides the end-user with an interface that includes relevant temporal and spatial information about the fire risk which may offer useful insight in addition to

a binary fire classification. The outputs can also be used to identify fire risk conditions. Such Deep Learning based multimodal detection using sensor network has some unique advantages compared to camera-based systems."[1, page 7]

"One of the challenges of implementing any type of ground-based sensor network with deep-learning based ML models is the data bandwidth limitations in areas of deployment. For example, a typical camera system (Pan-Tilt-Zoom) requires about 24 Mbps of upload speed, whereas most typical upload speeds in 4G and 4G LTE are in the range of 20 Mbps. LoRa or other types on low-power long-distance radio communications have even lower bandwidths. Such [a] multimodal sensing approach with ground-based sensor networks represents a paradigm shift in detection that is scalable and deployable in most geographical locations."1, page 7]

There is an additional N5 Sensors White Paper: “How Satellites are used to detect and track wildfires in the US”. The white paper describes use of high altitude geostationary satellites (highest temporal resolution) and lower altitude orbital satellites (highest spatial resolution but low temporal resolution). However, the white paper does not say anything about Artificial Intelligence. The white paper references a 3-hour time lag between information collection and when the data is available to the public (NASA FIRMS Fire Information for Resource Management System system for active fire data).[2]

References

1. N5 Sensors, white paper “Machine Learning Models and Their Effectiveness for Autonomous Detection of Wildfires Utilizing Ground-Based Sensor Networks.” Home - N5 Sensors
2. N5 Sensors white paper "How Satellites are used to detect and track wildfires in the US." Home - N5 Sensors.
3. Home - N5 Sensors website.
4. LoRa - Wikipedia. Information pulled September 22, 2025. Text is available under the Creative Commons Attribution-ShareAlike 4.0 License.

This page titled 7.5.1.1: N5 Sensors is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

7.5.1.2: Ladris

Finding **specific** details about Ladris's products or about the use of Artificial Intelligence in Ladris's products on its web site was not straight forward. Since details were few, it is difficult to itemize their capabilities, much less perform a comparative assessment.

Corporate AI Ethics Statement

That having been said, their web site had a corporate statement on their AI ethics [2] related to their corporate mission to develop "AI-driven solutions that enhance emergency and risk management for local governments." "Our products are designed to predict disaster impacts, simulate evacuations, and monitor infrastructure risks."[2]

Two principles in their AI ethics statement that are worth pointing out are the following:

- "**Privacy and data protection.** Respecting user privacy is paramount. We adhere to stringent data protection laws and ensure that our AI systems handle sensitive information with utmost care, especially when dealing with personal and community data.
- **Compliance with Regulations**. We stay informed about and comply with all relevant AI regulations and standards. This commitment ensures that our products meet legal requirements and uphold ethical considerations in all jurisdictions we operate." [2]

Ladris (Wildfire Modeling and Monitoring)

The Wildfire Modeling and Monitoring capability is built on their "Core" platform that "integrates the full spectrum of emergency management data into a single **AI-native** system, from sensors and satellite imagery to plans, procedures, and models."[4] The Core platform "connects to thousands of real-time and historical data sources to provide situational awareness across early warning, infrastructure status, demographics, and field conditions."[4] "Core automates scenario modeling and threat detection based on live hazard inputs such as **satellite fire detections**." [4]

In a section entitled AI-Driven Risk Analysis for Planning and Response Ladris states: "Using **AI models** calibrated to public safety outcomes, Core surfaces actionable insights on population risk, mobility, and infrastructure vulnerability. It supports strategic planning, incident response, and resource allocation across all phases of emergency management."[4]

"Built on the Core platform, Ladris Fire equips emergency teams with powerful **predictive tools to simulate fire behavior** and act immediately. Built on trusted **fire spread models and enhanced with live environmental data**, this system delivers high-speed simulations of fire growth and community impact before and during a fire. It supports wildfire model planning and real-time operations alike."[3]

"Fire Spread Modeling in Live Conditions: Ladris Fire generates **predictive fire spread modeling maps** using current weather, terrain, and vegetation data. Simulations run in seconds and automatically adjust as new information becomes available, helping emergency teams anticipate growth patterns and prepare suppression strategies without needing manual setup or calibration."[3]

"**Real-Time Wildfire Monitoring and Alerts**: The system integrates real-time fire detection from **satellite and sensor feeds** to track new ignitions and changes in existing incidents. Users receive automatic updates and alerts as conditions evolve, supporting faster situational updates and more informed operational decisions."[3]

References

1. Ladris. Ladris: Smarter Decisions for Disaster Response
2. Monaghan, Steve. Ladris Technologies, Inc. "Ladris Responsible & Ethical AI Statement", April 12, 2025. Ladris.
3. Ladris. Ladris Fire.
4. Ladris. Ladris Core.

This page titled 7.5.1.2: Ladris is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

7.5.1.3: Pano.ai

Pano.ai offers advanced wildfire detection technology that makes use of artificial intelligence and an alert system for notifying relevant public safety professionals.

The Pano Rapid Detect product is comprised of the following:

- **Detection** via the Pano Station. These stations consist of dual tower-mounted cameras that capture a 360 degree view of the environment every minute. The images are merged to create a panoramic view that is transmitted to the cloud. The station can be configured with power source options and data transmission options (details not provided).
- **Confirmation**. At the Pano Intelligence Center the images are examined and analyzed looking for signs of smoke. The fire detection methodology includes the use of artificial intelligence to operate in both daytime and nighttime. If two stations are essentially reporting on the same incident at the same time, the Pano system can use a patented algorithm ("triangulation technology") to ascertain the exact GPS position of the fire. The process includes humans to confirm the detection of a fire by the AI software. algorithms It is also possible to overlay utility infrastructure maps to make the alerts for utility personnel more meaningful.
- **Information dissemination/alerts.** Upon confirmation of a fire detection, alerts (SMS, email) can be sent to a predefined set of public safety professionals, public utility personnel, and so forth. This includes information such as images of the fire and its location.

The system uses **deep learning AI.**

Pano AI charges around $50,000 per year per camera[2].

Customers include forestry operations, government agencies, and utilities.[2]

References

1. Advanced Wildfire + Bushfire Detection Technology | Pano AI. Referenced on September 22, 2025.
2. Pineda, Dorothy and Peterson, Brittany from the Associated Press. "States in wildfire areas are using AI for early detection." The Mercury News, Business section page C8 and C9, May 11, 2026 (print edition).

This page titled 7.5.1.3: Pano.ai is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

7.6: Project Management Considerations

PG&E's evaluation of new systems and technology follows this process (as discussed briefly earlier):

1. Identifying new technologies or systems;
2. Conducting a Subject Matter Expert (SME) and business review for reasonableness and feasibility;
3. Evaluating alignment to program goals and objectives;
4. Benchmarking, if applicable;
5. Determining a source of funding and cost review;
6. Performing a pilot study, if needed, to evaluate effectiveness at achieving program goals and objectives and testing assumptions;
7. Implementing and deploying the technological system; and
8. Conducting a SME operational review of benefits. [1, Section 7.3, page 450]

This process shares some similarities to a typical project life cycle as taught in CIS 79 and as used by Silicon Valley technology companies. In addition it shares elements with business analysis processes as taught in CIS 95H Business and Requirement Analysis.

In general, one needs to perform a "needs analysis" first to determine what opportunity might be pursued or what problem is to be solved. This of course needs to be consistent with current corporate strategy and operating plan. Subject Matter Experts can participate in brainstorming potential solutions and assessing their relative feasibility. In order to implement a solution, it is important to perform a capabilities analysis to determine if the organization has the needed capabilities to implement the solution internally or whether the organization needs to hire the resources externally or contract for the resources from outside the organization.

A concept review is typically done with a project steering committee that may agree with the the concept and provide initial seed funding for initial prototyping before the full-blown implementation project is planned (schedule, resources needed, budget required).

Vendor evaluation

Imagine that you are running a project to select a vendor to supply wildfire detection equipment.

One could assess whether the solution could be developed internally or whether the company needs to select a solution provider from outside the organization. This is part of a more general "Make versus buy" decision which also considers cost and schedule impacts.

Some tools for facilitating this are shown below.

Basic Decision Matrix [2]

Basic Decision Matrix (non-weighted)

Criterion	Solution Alternative #1	Solution Alternative #2	Solution Alternative
Criterion 1	A	D	G
Criterion 2	B	E	H
Criterion 3	C	F	I
Total Score	A+B+C	D+E+F	G+H+I

Table 7.6.1: Basic Decision Matrix. (Pohl, Klaus and Rupp, Chris. Requirements Engineering Fundamentals – A Study Guide for the Certified Professional for Requirements Engineering Exam Foundation Level/IREB Compliant, 2nd Edition, 2016, page 108.)

With this method, one selects the alternative with the highest total score.

Weighted Alternative Decision Matrix

Another method is to incorporate into the evaluation of vendors is the weighted alternative matrix or multi-criteria selection matrix. The key is to choose the major factors or criteria to be used in the evaluation and the weights your team believes should be assigned to each criterion. It is important to get these weights correctly decided up front as they can skew the decision making process.

		Option 1		Option 2		Option 3	
Criterion	Assigned Weight	Rating	Points	Rating	Points	Rating	Points
Cost	6	8	48	7	42	8	48
Schedule	6	8	48	7	42	8	48
Technical requirements	8	6	48	8	64	5	40
Ease of use	9	9	81	5	45	7	63
Functionality	8	6	48	9	72	6	48
Total Score			273		265		247

Figure 7.6.2: Weighted Alternative Multi-criteria Decision Matrix. (Robert Slate, CIS 78 course materials)

Accessibility Note

This Matrix shows then following by column:

- Criterion: Cost, Schedule, Technical requirements. ease of use, Functionality, Total Score.
- Assigned weight: numbers from 0 to 10
- For each option (Option 1, Option 2, Option 3) there are two sub-columns:
  - Rating: the rating the team assigns to each option (numbers from 0 to 10)
  - Points: the criterion's Assigned Weight multiplied by the Option's Rating

One calculates the "Points" simply by multiplying the criterion's Assigned Weight by the Option's Rating. One then totals the points for all the criteria for each option. Finally, the total points for each option are compared to rank the options in order of highest score (best) to lowest score (worst). The result is the team's recommendation to management.

Making the decision:

- Present the options that were evaluated to a steering committee/approval committee, showing the analysis.
- Make a recommendation on which option to pursue.
- The committee makes the decision based on the feasible options, their business judgement, and risk appetite.

Risk assessments

As mentioned earlier, public electric utilities are generally conservative in their approaches with a risk appetite that most would characterize as risk-averse. That being said, one needs to realize that emergency situations sometimes require thinking outside the box and operating accordingly. To the extent that there are nascent technologies such as Artificial Intelligence available for implementation to ameliorate the massive impacts of wildfires, they need to be considered. In the case of public utilities, the suggestion of using new technologies is subject to regulatory oversight by Public Utilities Commissions (e.g., California Public Utilities Commission).

A typical tool that a project manager employs is a risk register. The risk register would include risk attributes such as name of the risk, risk type, risk probability, risk severity (magnitude of the risk impact if the risk materializes), risk mitigations, and an identified person who i responsible for the risk.

In the next section, we will look at legal considerations.

References

1. PG&E Wildfire Mitigation Plan R0 2026-2028. PG&E Wildfire Mitigation Plan R0 2026-2028, Volume 1 of 2, April 4, 2025.
2. Pohl, Klaus and Rupp, Chris. Requirements Engineering Fundamentals – A Study Guide for the Certified Professional for

Requirements Engineering Exam Foundation Level/IREB Compliant, 2nd Edition, 2016, page 108.

This page titled 7.6: Project Management Considerations is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

7.7: Legal Considerations

We will first address the legal underpinnings of the California Public Utilities Commission. Then, we will explore some specific involvement by the California Public Utilities Commission, particularly in the area of Advanced Metering Infrastructure (AMI) or Smart Meters.

We look briefly at privacy issues associated with camera use in wildfire detection.

We conclude with a discussion of potential impacts of AI regulation.

This page titled 7.7: Legal Considerations is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

7.7.1: CPUC Governance

CPUC governance of public utilities

The California Public Utilities Commission (CPUC) regulates privately owned electric, natural gas, telecommunications, water, railroad, rail transit, and passenger transportation companies, in addition to authorizing video franchises.[4]

The Energy Division assists Commission activities in the electricity, natural gas, steam, and petroleum pipeline industries.[5]

The CPUC has safety oversight responsibilities that include: mandating risk management, infrastructure maintenance, and Wildfire Mitigation Plans for electric and gas utilities.

Public Utilities Code Section 8381

The California Public Utilities Code (PUC) is the legislative framework that governs privately owned utility services in California. It is enforced by the California Public Utilities Commission (CPUC). The Public Utilities Code contains 33 divisions. The applicable part for our purposes is:

DIVISION 4.1. PROVISIONS APPLICABLE TO PRIVATELY OWNED AND PUBLICLY OWNED PUBLIC UTILITIES [8301 - 8550] *(Heading of Division 4.1 amended by Stats. 1988, Ch. 1560, Sec. 2.)*

The sections of interest are:

- CHAPTER 4. Smart Grid Systems [8360 - 8369] *(Chapter 4 added by Stats. 2009, Ch. 327, Sec. 1.)*
- CHAPTER 5. Privacy Protections for Energy Consumption Data [8380 - 8381] *(Chapter 5 added by Stats. 2010, Ch. 497, Sec. 2.)*
- CHAPTER 6: Wildfire Mitigation [8385-8389]

Public Utilities Code section 8381 "Privacy Protections for Energy Consumption Data" reads as follows:[1]

Caution

The hierarchy and numbering of the original legal document have been fully preserved. The only change made was converting list markers to proper ordered list formatting for accessibility, (a), (b), (c) became a, b, c, and (1), (2), (3) became 1, 2, 3. This is a structural change only. It does not alter the document's content, sequence, or citations, and brings it into conformance with WCAG accessibility standards. Note that the major sections are letters with subordinate sections using numbers. We have added bold text for emphasis.

a. For purposes of this section, “electrical consumption data” means data about a customer’s electrical usage that is made available as part of an advanced metering infrastructure, and includes the name, account number, or residence of the customer.
b. (1) A local publicly owned electric utility shall not share, disclose, or otherwise make accessible to any third party a customer’s electrical consumption data, except as provided in subdivision (f) or upon the consent of the customer.
    2. A local publicly owned electric utility shall not sell a customer’s electrical consumption data or any other personally identifiable information for any purpose.
    3. A local publicly owned electric utility or its contractors shall not provide an incentive or discount to the customer for accessing the customer’s electrical consumption data without the prior consent of the customer.
    4. A local publicly owned electric utility that utilizes an advanced metering infrastructure that allows a customer to access the customer’s electrical consumption data shall ensure that the customer has an option to access that data without being required to agree to the sharing of the customer’s personally identifiable information, including electrical consumption data, with a third party.
c. If a local publicly owned electric utility contracts with a third party for a service that allows a customer to monitor the customer’s electricity usage, and that third party uses the data for a secondary commercial purpose, the contract between the local publicly owned electric utility and the third party shall provide that the third party prominently discloses that secondary

commercial purpose to the customer and secures the customer's consent to the use of the customer's data for that secondary commercial purpose prior to the use of the data.

d. A local publicly owned electric utility shall use reasonable security procedures and practices to protect a customer's unencrypted electrical consumption data from unauthorized access, destruction, use, modification, or disclosure, and to prohibit the use of the data for a secondary commercial purpose not related to the primary purpose of the contract without the customer's consent.
e. A local publicly owned electric utility shall not share, disclose, or otherwise make accessible to any immigration authority, as defined in Section 7284.4 of the Government Code, a customer's electrical consumption data without a court-ordered subpoena or judicial warrant.
f. (1) This section shall not preclude a local publicly owned electric utility from using customer aggregate electrical consumption data for analysis, reporting, or program management if all information has been removed regarding the individual identity of a customer.
    2. This section shall not preclude a local publicly owned electric utility from disclosing a customer's electrical consumption data to a third party for system, grid, or operational needs, or the implementation of demand response, energy management, or energy efficiency programs, provided, for contracts entered into after January 1, 2011, that the utility has required by contract that the third party implement and maintain reasonable security procedures and practices appropriate to the nature of the information to protect the personal information from unauthorized access, destruction, use, modification, or disclosure, and prohibits the use of the data for a secondary commercial purpose not related to the primary purpose of the contract without the customer's prior consent to that use.
    3. Except as provided in subdivision (e), this section shall not preclude a local publicly owned electric utility from disclosing electrical consumption data as required under state or federal law.
g. If a customer chooses to disclose the customer's electrical consumption data to a third party that is unaffiliated with, and has no other business relationship with, the local publicly owned electric utility, the utility shall not be responsible for the security of that data, or its use or misuse.

(Amended by Stats. 2020, Ch. 188, Sec. 3. (AB 2788) Effective January 1, 2021.)

CPUC decision regarding privacy and security of energy usage data

In 2011, the CPUC issued a decision that provide more specifics regarding privacy and security of energy usage data.[3] CPUC Decision 11-07-056 "DECISION ADOPTING RULES TO PROTECT THE PRIVACY AND SECURITY OF THE ELECTRICITY USAGE DATA OF THE CUSTOMERS OF PACIFIC GAS AND ELECTRIC COMPANY, SOUTHERN CALIFORNIA EDISON COMPANY, AND SAN DIEGO GAS & ELECTRIC COMPANY:"

"This decision adopts rules to protect the privacy and security of customer data generated by Smart Meters concerning the usage of electricity that are deployed by Pacific Gas and Electric Company (PG&E), Southern California Edison Company (SCE), and San Diego Gas & Electric Company (SDG&E). The rules adopted implement the protections ordered by Senate Bill (SB) 1476 (Chapter 497, Statutes of 2010). The adopted rules are also consistent with other sections of the Public Utilities Code and past Commission privacy policies. Attachment D, Rules Regarding Privacy and Security Protections for Energy Usage Data, lists the adopted privacy and security rules."[3,p2]

"The decision also adopts reporting and audit requirements regarding the utilities' customer data privacy and security practices, third-party access to customer usage information, and any security breaches of customer usage information. The adopted privacy and security rules and policies providing access to billing and usage data are reasonable. They will protect the privacy and security of customer usage data while ensuring customer access to usage information and enabling utilities and authorized third parties to use the information to provide useful energy management and conservation services. In addition, the rules and policies are consistent with privacy and security principles adopted by the Department of Homeland Security and with the policies adopted in SB 1476. Thus, these rules will bring California practices into conformity with the best national privacy and security practices."[3,p4-5]

Appendix D[7] endorses Fair Information Practice (FIP) Principles.[3,p10] These FIP principles were adopted by the Department of Homeland Security.[3,p12]

Rules Regarding Privacy and Security Protections for Energy Usage Data

1. DEFINITIONS

a. Covered Entity. A “covered entity” is (1) any electrical corporation,1 or any third party that provides services to an electrical corporation under contract, (2) any third party who accesses, collects, stores, uses or discloses covered information pursuant to an order of the Commission, unless specifically exempted, who obtains this information from an electrical corporation, or (3) any third party, when authorized by the customer, that accesses, collects, stores, uses, or discloses covered information relating to 11 or more customers who obtains this information from an electrical corporation.
b. Covered Information. “Covered information” is any usage information obtained through the use of the capabilities of Advanced Metering Infrastructure when associated with any information that can reasonably be used to identify an individual, family, household, residence, or non-residential customer, except that covered information does not include usage information from which identifying information has been removed such that an individual, family, household or residence, or non residential customer cannot reasonably be identified or re-identified. Covered information, however, does not include information provided to the Commission pursuant to its oversight responsibilities.
c. Primary Purposes. The “primary purposes” for the collection, storage, use or disclosure of covered information are to— (1) provide or bill for electrical power or gas, (2) provide for system, grid, or operational needs, (3) provide services as required by state or federal law or as specifically authorized by an order of the Commission, or (4) plan, implement, or evaluate demand response, energy management, or energy efficiency programs under contract with an electrical corporation, under contract with the Commission , or as part of a Commission authorized program conducted by a governmental entity under the supervision of the Commission.
d. [sic, should be (e)] Secondary Purpose. “Secondary purpose” means any purpose that is not a primary purpose.

2. TRANSPARENCY (NOTICE)

a. Generally. Covered entities shall provide customers with meaningful, clear, accurate, specific, and comprehensive notice regarding the accessing, collection, storage, use, and disclosure of covered information. Provided, however, that covered entities using covered data solely for a primary purpose on behalf of and under contract with utilities are not required to provide notice separate from that provided by the utility.
b. When Provided. Covered entities shall provide written notice when confirming a new customer account and at least once a year shall inform customers how they may obtain a copy of the covered entity’s notice regarding the accessing, collection, storage, use, and disclosure of covered information, and shall provide a conspicuous link to the notice on the home page of their website, and shall include a link to their notice in all electronic correspondence to customers.
c. Form. The notice shall be labeled Notice of Accessing, Collecting, Storing, Using and Disclosing Energy Usage Information and shall— (1) be written in easily understandable language, and (2) be no longer than is necessary to convey the requisite information.
d. Content. The notice and the posted privacy policy shall state clearly—
    1. the identity of the covered entity,
    2. the effective date of the notice or posted privacy policy,
    3. the covered entity’s process for altering the notice or posted privacy policy, including how the customer will be informed of any alterations, and where prior versions will be made available to customers, and
    4. the title and contact information, including email address, postal address, and telephone number, of an official at the covered entity who can assist the customer with privacy questions, concerns, or complaints regarding the collection, storage, use, or distribution of covered information.

3. PURPOSE SPECIFICATION

The notice required under section 2 shall provide—

a. an explicit description of—
    1. each category of covered information collected, used, stored or disclosed by the covered entity, and, for each category of covered information, the reasonably specific purposes for which it will be collected, stored, used, or disclosed,
    2. each category of covered information that is disclosed to third parties, and, for each such category, (i) the purposes for which it is disclosed, and (ii) the categories of third parties to which it is disclosed, and
    3. the identities of those third parties to whom data is disclosed for secondary purposes, and the secondary purposes for which the information is disclosed;

b. the approximate period of time that covered information will be retained by the covered entity;
c. a description of—
    1. the means by which customers may view, inquire about, or dispute their covered information, and
    2. the means, if any, by which customers may limit the collection, use, storage or disclosure of covered information and the consequences to customers if they exercise such limits.

4. INDIVIDUAL PARTICIPATION (ACCESS AND CONTROL)

a. Access. Covered entities shall provide to customers upon request convenient and secure access to their covered information—
    1. in an easily readable format that is at a level no less detailed than that at which the covered entity discloses the data to third parties.
    2. The Commission shall, by subsequent rule, prescribe what is a reasonable time for responding to customer requests for access.
b. Control. Covered entities shall provide customers with convenient mechanisms for—
    1. granting and revoking authorization for secondary uses of covered information,
    2. disputing the accuracy or completeness of covered information that the covered entity is storing or distributing for any primary or secondary purpose, and
    3. requesting corrections or amendments to covered information that the covered entity is collecting, storing, using, or distributing for any primary or secondary purpose.
c. Disclosure Pursuant to Legal Process.
    1. Except as otherwise provided in this rule or expressly authorized by state or federal law or by order of the Commission, a covered entity shall not disclose covered information except pursuant to a warrant or other court order naming with specificity the customers whose information is sought. Unless otherwise directed by a court, law, or order of the Commission, covered entities shall treat requests for real-time access to covered information as wiretaps, requiring approval under the federal or state wiretap law as necessary.
    2. Unless otherwise prohibited by court order, law, or order of the Commission, a covered entity, upon receipt of a subpoena for disclosure of covered information pursuant to legal process, shall, prior to complying, notify the customer in writing and allow the customer 7 days to appear and contest the claim of the person or entity seeking disclosure.
    3. Nothing in this rule prevents a person or entity seeking covered information from demanding such information from the customer under any applicable legal procedure or authority.
    4. Nothing in this section prohibits a covered entity from disclosing covered information with the consent of the customer, where the consent is express, in written form, and specific to the purpose and to the person or entity seeking the information.
    5. Nothing in this rule prevents a covered entity from disclosing, in response to a subpoena, the name, address and other contact information regarding a customer.
    6. On an annual basis, covered entities shall report to the Commission the number of demands received for disclosure of customer data pursuant to legal process or pursuant to situations of imminent threat to life or property and the number of customers whose records were disclosed. Upon request of the Commission, covered entities shall report additional information to the Commission on such disclosures. The Commission may make such reports publicly available without identifying the affected customers, unless making such reports public is prohibited by state or federal law or by order of the Commission.
d. Disclosure of Information in Situations of Imminent Threat to Life or Property. These rules concerning access, control and disclosure do not apply to information provided to emergency responders in situations involving an imminent threat to life or property. Emergency disclosures, however, remain subject to reporting rule 4(c)(6).

5. DATA MINIMIZATION

a. Generally. Covered entities shall collect, store, use, and disclose only as much covered information as is reasonably necessary or as authorized by the Commission to accomplish a specific primary purpose identified in the notice required under section 2 or for a specific secondary purpose authorized by the customer.
b. Data Retention. Covered entities shall maintain covered information only for as long as reasonably necessary or as authorized by the Commission to accomplish a specific primary purpose identified in the notice required under section 2 or for a specific secondary purpose authorized by the customer.

Data Disclosure. Covered entities shall not disclose to any third party more covered information than is reasonably necessary or as authorized by the Commission to carry out on behalf of the covered entity a specific primary purpose identified in the notice

required under section 2 or for a specific secondary purpose authorized by the customer.

6. USE AND DISCLOSURE LIMITATION

a. Generally. Covered information shall be used solely for the purposes specified by the covered entity in accordance with section 3.
b. Primary Purposes. An electrical corporation, a third party acting under contract with the Commission to provide energy efficiency or energy efficiency evaluation services authorized pursuant to an order or resolution of the Commission, or a governmental entity providing energy efficiency or energy efficiency evaluation services pursuant to an order or resolution of the Commission may access, collect, store and use covered information for primary purposes without customer consent. Other covered entities may collect, store and use covered information only with prior customer consent, except as otherwise provided here.
c. Disclosures to Third Parties.
    1. Initial Disclosure by an Electrical Corporation. An electrical corporation may disclose covered information without customer consent to a third party acting under contract with the Commission for the purpose of providing services authorized pursuant to an order or resolution of the Commission or to a governmental entity for the purpose of providing energy efficiency or energy efficiency evaluation services pursuant to an order or resolution of the Commission. An electrical corporation may disclose covered information to a third party without customer consent
        a. when explicitly ordered to do so by the Commission; or
        b. for a primary purpose being carried out under contract with and on behalf of the electrical corporation disclosing the data;

    provided that the covered entity disclosing the data shall, by contract, require the third party to agree to access, collect, store, use, and disclose the covered information under policies, practices and notification requirements no less protective than those under which the covered entity itself operates as required under this rule, unless otherwise directed by the Commission.

    2. Subsequent Disclosures. Any entity that receives covered information derived initially from a covered entity may disclose such covered information to another entity without customer consent for a primary purpose, provided that the entity disclosing the covered information shall, by contract, require the entity receiving the covered information to use the covered information only for such primary purpose and to agree to store, use, and disclose the covered information under policies, practices and notification requirements no less protective than those under which the covered entity from which the covered information was initially derived operates as required by this rule, unless otherwise directed by the Commission.
    3. Terminating Disclosures to Entities Failing to Comply With Their Privacy Assurances. When a covered entity discloses covered information to a third party under this subsection 6(c), it shall specify by contract, unless otherwise ordered by the Commission, that it shall be considered a material breach if the third party engages in a pattern or practice of accessing, storing, using or disclosing the covered information in violation of the third party's contractual obligations to handle the covered information under policies no less protective than those under which the covered entity from which the covered information was initially derived operates in compliance with this rule.
        - If a covered entity disclosing covered information for a primary purpose being carried out under contract with and on behalf of the entity disclosing the data finds that a third party contractor to which it disclosed covered information is engaged in a pattern or practice of accessing, storing, using or disclosing covered information in violation of the third party's contractual obligations related to handling covered information, the disclosing entity shall promptly cease disclosing covered information to such third party.
        - If a covered entity disclosing covered information to a Commission-authorized or customer-authorized third party receives a customer complaint about the third party's misuse of data or other violation of the privacy rules, the disclosing entity shall, upon customer request or at the Commission's direction, promptly cease disclosing that customer's information to such third party. The disclosing entity shall notify the Commission of any such complaints or suspected violations.
    4. Nothing in this section shall be construed to impose any liability on an electrical corporation relating to disclosures of information by a third party when i) the Commission orders the provision of covered data to a third party; or ii) a customer authorizes or discloses covered data to a third party entity that is unaffiliated with and has no other business relationship with the electrical corporation. After a secure transfer, the electrical corporation shall not be responsible for the security of the covered data or its use or misuse by such third party. This limitation on liability does not apply when a utility has acted recklessly.

d. **Secondary Purposes.** No covered entity shall use or disclose covered information for any secondary purpose without obtaining the customer's prior, express, written authorization for each type of secondary purpose. This authorization is not

required when information is—
1. provided pursuant to a legal process as described in 4(c) above;
2. provided in situations of imminent threat to life or property as described in 4(d) above; or
3. authorized by the Commission pursuant to its jurisdiction and control.

e. Customer Authorization.
   1. Authorization. Separate authorization by each customer must be obtained for all disclosures of covered information except as otherwise provided for herein.
   2. Revocation. Customers have the right to revoke, at any time, any previously granted authorization.
   3. Opportunity to Revoke. The consent of a residential customer shall continue without expiration, but an entity receiving information pursuant to a residential customer's authorization shall contact the customer, at least annually, to inform the customer of the authorization granted and to provide an opportunity for revocation. The consent of a non-residential customer shall continue in the same way, but an entity receiving information pursuant to a non-residential customer's authorization shall contact the customer, to inform the customer of the authorization granted and to provide an opportunity for revocation either upon the termination of the contract, or annually if there is no contract.
f. Parity. Covered entities shall permit customers to cancel authorization for any secondary purpose of their covered information by the same mechanism initially used to grant authorization.
g. Availability of Aggregated Usage Data. Covered entities shall permit the use of aggregated usage data that is removed of all personally-identifiable information to be used for analysis, reporting or program management provided that the release of that data does not disclose or reveal specific customer information because of the size of the group, rate classification, or nature of the information.

7. DATA QUALITY AND INTEGRITY

Covered entities shall ensure that covered information they collect, store, use, and disclose is reasonably accurate and complete or otherwise compliant with applicable rules and tariffs regarding the quality of energy usage data.

8. DATA SECURITY

a. Generally. Covered entities shall implement reasonable administrative, technical, and physical safeguards to protect covered information from unauthorized access, destruction, use, modification, or disclosure.
b. Notification of Breach. A covered third party shall notify the covered electrical corporation that is the source of the covered data within one week of the detection of a breach. Upon a breach affecting 1,000 or more customers, whether by a covered electrical corporation or by a covered third party, the covered electrical corporation shall notify the Commission's Executive Director of security breaches of covered information within two weeks of the detection of a breach or within one week of notification by a covered third party of such a breach. Upon request by the Commission, electrical corporations shall notify the Commission's Executive Director of security breaches of covered information.
c. Annual Report of Breaches. In addition, electrical corporations shall file an annual report with the Commission's Executive Director, commencing with the calendar year 2012, that is due within 120 days of the end of the calendar year and notifies the Commission of all security breaches within the calendar year affecting covered information, whether by the covered electrical corporation or by a third party.

9. ACCOUNTABILITY AND AUDITING

a. Generally. Covered entities shall be accountable for complying with the requirements herein, and must make available to the Commission upon request or audit—
   1. the privacy notices that they provide to customers,
   2. their internal privacy and data security policies,
   3. the categories of agents, contractors and other third parties to which they disclose covered information for a primary purpose, the identities of agents, contractors and other third parties to which they disclose covered information for a secondary purpose, the purposes for which all such information is disclosed, indicating for each category of disclosure whether it is for a primary purpose or a secondary purpose. (A covered entity shall retain and make available to the Commission upon request information concerning who has received covered information from the covered entity.), and
   4. copies of any secondary-use authorization forms by which the covered party secures customer authorization for secondary uses of covered data.
b. Customer Complaints. Covered entities shall provide customers with a process for reasonable access to covered information, for correction of inaccurate covered information, and for addressing customer complaints regarding covered information

under these rules.

c. Training. Covered entities shall provide reasonable training to all employees and contractors who use, store or process covered information.
d. Audits. Each electrical corporation shall conduct an independent audit of its data privacy and security practices in conjunction with general rate case proceedings following 2012 and at other times as required by order of the Commission. The audit shall monitor compliance with data privacy and security commitments, and the electrical corporation shall report the findings to the Commission as part of the utility's general rate case filing.
e. Reporting Requirements. On an annual basis, each electrical corporation shall disclose to the Commission as part of an annual report required by Rule 8.b, the following information:
    1. the number of authorized third parties accessing covered information,
    2. the number of non-compliances with this rule or with contractual provisions required by this rule experienced by the utility, and the number of customers affected by each non-compliance and a detailed description of each non-compliance.

California Consumer Privacy Act (CCPA)

AI-generated data could be viewed as personal information and thus covered by the CCPA. The California PUC, as part of its governance responsibilities, can apply the CCPA and other privacy laws to regulated public utilities that collect or process personal data via AI.

The PUC's governance can extend to public utility AI applications such as predictive maintenance and demand forecasting.

Public utilities that make use of AI-driven chatbots in their customer services operations or in automated billing systems need to comply with CCPA and privacy law requirements.

Transparency in Frontier Artificial Intelligence Act (TFAIA)

The Transparency in Frontier Artificial Intelligence Act (TFAIA) was signed into law on September 29, 2025, and took effect on January 1, 2026. Also referred to as SB-53 (Senate Bill-53), it is a 2025 California law which mandates increased transparency for companies building artificial intelligence. SB-53 is primarily focused on assessing and reducing potential catastrophic risks from artificial intelligence, and is the first bill addressing such risks to be passed into law in the United States.[8]

The Transparency in Frontier Artificial Intelligence Act covers "frontier" AI models used in regulated industries such as public utilities. Public utilities that employ AI for electric grid management, predictive maintenance of equipment, or customer services operations need to comply with TFAIA transparency and safety reporting requirements.

It requires companies to create publicly accessible documents assessing potential "catastrophic risk[s]" from their AI models, as well as publishing documentation on how the model incorporates national and international safety standards. SB-53 also sets up whistleblower protections and procedures for alerting the government to a "critical safety incident."[8]

Provisions

- SB-53 applies primarily to companies making at least $500 million in yearly gross revenue. It defines a “frontier model” as any AI trained with over 10^{26} FLOPS (including fine-tuning), including unreleased internal models. Both the financial and computational thresholds must be met before most of the law is applied, although the threshold can be lowered or otherwise updated by the California Department of Technology in an annual review starting in 2027. Most of the bill's provisions are focused on "catastrophic risks" from AI, which are defined as incidents in which a model contributes to more than 50 deaths or serious injuries, or causes more than one billion dollars ($1,000,000,000) in economic damage from AI-assisted acts (such as cyber-attacks or the creation of biological weapons).[8]
- The bill requires companies to provide publicly accessible safety frameworks for frontier AI models, describing how the company tests for catastrophic risk from its AI, and how it implements protections against such risks. This includes addressing the possibility that the AI may attempt to circumvent internal guardrails or oversight mechanisms. (Certain safety incidents, such as dangerously deceptive model behavior, physical injury, or death, must be reported to California Office of Emergency Services (OES) within 15 days, unless the incident poses imminent physical risk, in which case it must be reported immediately.) The company must follow its published framework, and if any changes are made, the framework should be updated within 30 days, and justification for said changes must also be made public.[8]

- All frontier companies are required to publish basic information about newly released frontier models (such as terms of service, supported languages, and intended use), although only large companies (making over $500 million annually) need to publish full safety frameworks.[8]
- SB-53 also establishes various whistleblower protections for covered employees. Large companies must have anonymous whistle-blowing channels in place which protect employees from retaliation from reporting risks to state or federal authorities if they have reasonable cause to believe that their employer is substantially risking public health and safety.[8]

References

1. California Public Utilities Code Section 8381 (2025). Accessed May 11, 2026.
2. California PUC 8381. Accessed May 11, 2026.
3. California Public Utilities Commission.Decision d.11-07-056: Rules regarding privacy and security protections for energy usage data, 2011a. DECISION ADOPTING RULES TO PROTECT THE PRIVACY AND SECURITY OF THE ELECTRICITY USAGE DATA OF THE CUSTOMERS OF PACIFIC GAS AND ELECTRIC COMPANY, SOUTHERN CALIFORNIA EDISON COMPANY, AND SAN DIEGO GAS & ELECTRIC COMPANY. July 28, 2011. 171 pages. It does not include Appendices (only a link), of which Appendix D is most relevant. California Public Utilities Commission.Decision d.11-07-056: Rules regarding privacy and security protections for energy usage data, 2011a.
4. California Public Utilities Commission. California Public Utilities Commission. Accessed May 13, 2026.
5. California Public Utilities Commission. CPUC Divisions. Accessed May 13, 2026.
6. California Legislature. California Legislature Codes. Accessed May 13, 2026.
7. California Public Utilities Commission. Appendix A-D associated with CPUC Decision D1107056. Appendix A-D associated with CPUC Decision D1107056. Accessed May 8, 2026.
8. Wikipedia authors. Transparency in Frontier Artificial Intelligence Act - Wikipedia. Accessed May 15, 2026. Text is available under the Creative Commons Attribution-ShareAlike 4.0 License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.
9. California Legislature. SB-53 Artificial intelligence models: large developers. Accessed May 15, 2026.

This page titled 7.7.1: CPUC Governance is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

7.7.2: Wildfire Camera Privacy Issues

Previously we have mentioned that there could be privacy issues associated with the use of wildfire cameras. While the purpose of the cameras is to scan the terrain for evidence of potential wildfires, the cameras are coincidentally picking up non-fire related images of people going about their daily lives in public or in and around their homes.

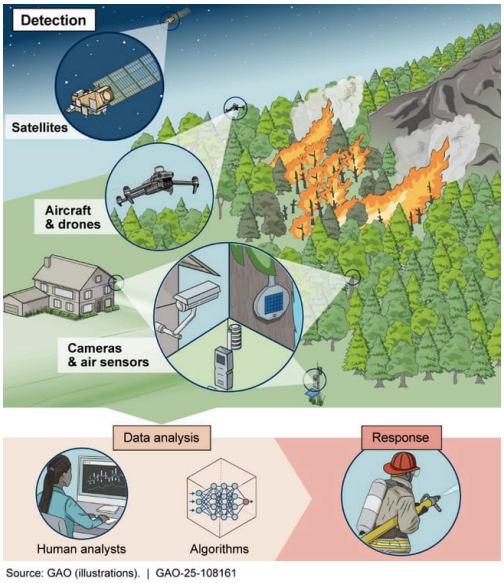


Figure 7.7.2.1: Current and Emerging Wildfire Detection Technologies. (Brian Bothwell, U.S.Government Accountability Office, Science & Tech Spotlight: Wildfire Detection Technologies | U.S. GAO [2])

Accessibility Note

Figure 7.7.2.1: Current and Emerging Wildfire Detection Technologies depicts detection mechanisms such as satellites, Aircraft & drones, Cameras & air sensors. This is followed by Data analysis (Human analysts, algorithms) and a response (deploying firefighting teams).

"Privacy and data security concerns. Cameras and drones near residences may raise privacy concerns. If these technologies collect private data, those data may then need to be managed and safeguarded."[2]

For the ALERTCalifornia wildfire sensor network[4], one mitigation that has been cited is: "To address **privacy concerns**, UC San Diego [which manages ALERTCalifornia] blurs homes, parking lots and other nearby features that could track people in real time."[1]

Generally speaking, under US constitutional law, people have a reduced expectation of privacy when going about in public.[5] That federal constitutional floor, however, can be raised by states' constitutional privacy rights, such as in Article 1 of the California Constitution[6], and by the so-called comprehensive privacy statutes passed by over twenty US states in recent years.[7]. For example, the California Constitution opens with the declaration that, "All people are by nature free and independent and have inalienable rights. Among these are enjoying and defending life and liberty, acquiring, possessing, and protecting property, and pursuing and obtaining safety, happiness, and privacy."[6] Privacy actions brought under the California Constitution generally need to show three things:

1. The existence of a legally protected privacy interest
2. There's a reasonable expectation of privacy under the circumstances
3. There was a serious invasion of the privacy interest.[8]

Potential privacy violations are detailed in California Civil Code Section 1708.8. For example, Cal. Civ. Code Section 1708.8(a)

provides that "A person is liable for physical invasion of privacy when the person knowingly enters onto the land or into the airspace above the land of another person without permission or otherwise commits a trespass in order to capture any type of visual image, sound recording, or other physical impression of the plaintiff engaging in a private, personal, or familial activity and the invasion occurs in a manner that is offensive to a reasonable person" [3]. However, the use of such data to save people from death or grave injury or to protect their assets might well prove inoffensive to a reasonable person (a standard assessed and applied by a court). Indeed, in other contexts, state privacy law makes allowances for emergency conditions, but these are typically intended to be limited to the duration of the emergency.[9]

References

1. AI cameras spot wildfires across California, often before humans call 911 Rogers, Paul. The Mercury News (online version), September 19, 2025.
2. Brian Bothwell, U.S.Government Accountability Office, Science & Tech Spotlight: Wildfire Detection Technologies, GAO-25-108161. Published May 1, 2025. Science & Tech Spotlight: Wildfire Detection Technologies | U.S. GAO, Accessed May 21, 2026.
3. Cal. Civil Code Section 1708.8. Accessed September 7, 2026.
4. UC San Diego, ALERTCalifornia. Accessed September 7, 2026.
5. See, e.g., United States v. Knotts, 460 U.S. 276 (1983). Accessed September 7, 2026.
6. CA Constitution art I § 1. Accessed September 7, 2026.
7. IAPP US State Privacy Legislation Tracker. Accessed September 7, 2026.
8. Int'l Fed'n of Prof'l & Technical Engineers, Local 21, AFL-CIO v. Superior Court, 165 P.3d 488, 499 (Cal. 2007).
9. See, e.g., Cal. Pen. Code Section 1546 et seq. (CalECPA). Accessed September 7, 2026.

This page titled 7.7.2: Wildfire Camera Privacy Issues is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

7.7.3: Impact of Potential AI Regulations

As we have shown in this chapter, the three major electric utilities in California are pursuing the use of artificial intelligence in a variety of applications. These artificial intelligence activities may expose them to data privacy laws and regulations and in particular to future AI regulations. As we have seen in the case studies in this textbook covering GDPR and RoHS, the EU is often in the forefront of establishing directives and regulations in areas of environmental protection and data privacy. The United States may follow at the federal and more likely at state levels with comparable laws. As such, it is important to keep an eye on regulation activities in the EU.

An example is the EU's Artificial Intelligence Act of 2024[2], which has "implications for both providers and organizations that deploy AI systems.[1]. The EU's AI Act[2] addresses high risk use cases in Annex III: "Critical infrastructure: AI systems intended to be used as safety components in the management and operation of critical digital infrastructure, road traffic, or in the supply of water, gas, heating or electricity."[2, page 127] After an amendment in 2026, the EU AI Act's obligations for high-risk AI systems such as those for electricity are scheduled to take effect December 2, 2027.[2]

Colorado's state AI Act, originally passed in 2024, was also amended in 2026 to delay its effective date to January 1, 2027 and to focus more on automated decision-making technologies (ADMT) rather than the original act's focus on AI risk compliance procedures and reporting.[4]

It is always important to distinguish between the parties and organizations to which a law and applies, and to which it does not apply. Some laws, such as California's Transparency in Frontier Artificial Intelligence Act of 2025, applies to AI developers and not AI deployers such as electric utilities.[3]

Nonetheless, California's privacy regulator, the California Privacy Protection Agency (CPPA), promulgated new rules in 2025 that specifically dealt with automated decision-making technology, defining ADMT as "any technology that processes personal information and uses computation to replace human decision-making or substantially replace human decision-making," and where "'substantially replace human decision-making' means a business uses the technology's output to make a decision without human involvement."[5] The ADMT regulations are relatively new and may encompass AI-powered camera alerts or machine learning scoring of the condition of critical infrastructure assets like power lines.

Current oversight of AI deployments is handled by public utility commissions such as the California Public Utilities Commission (CPUC) as part of oversight of the deployment of new technologies. "Utilities report their AI use cases to the commissions, such as intelligent image processing and machine learning to handle millions of images collected by drones deployed to inspect energy systems for issues."[1, page 10]

We have also seen that public utilities can require regulatory compliance documents such as Wildfire Mitigation Plans (WMPs) that we have covered earlier in this textbook in great detail.

On the U.S. federal side, the North American Electric Reliability Corporation (NERC) presented a report to the Federal Energy Regulatory Commission (FERC), recommending updates to reliability standards and collaboration with other federal agencies including NOAA and NASA to use "AI engines" so as "to improve forecast accuracy, modeling, and risk insights" [6, page 9].

> **Note**
>
> At the time this book went to press, California's legislature had passed several new AI and privacy bills, which were awaiting the governor's signature. Since we do not know which (if any) of these will prove relevant, we omit them here in this edition.

References

1. Slate, Daniel D., Parisot, Alexandre, Min, Liang, Panciatici, Patrick, Van Hentenryck; Adoption of Artificial Intelligence by Electric Utilities Energy Law Journal, Volume 45, No. 1, 2024., The Energy Bar Association, University of Tulsa College of Law.
2. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence

Act) Text with EEA relevance. EU Artificial Intelligence Act. Note that the AI Act has now been amended by Regulation (EU) 2026/1744 (the Digital Omnibus on AI).
3. California Senate Bill 53 (approved by California's Governor on September 29, 2025). Available on Legiscan as CA SB 53. Accessed September 6, 2026.
4. Colorado Senate Bill 189 (approved by Colorado's Governor on May 14, 2026). Available at Colorado Senate Bill.
5. California Consumer Privacy Act Regulations, Section 7001(e). Text of Regulations. Accessed September 7, 2026.
6. Candice Castaneda, North American Electric Reliability Corporation Report on Reducing the Risk of Wildfire Ignition by the Bulk Power System, Docket No. RD25-9-000, May 1, 2026. NERC Report to FERC. Accessed September 7, 2026.

This page titled 7.7.3: Impact of Potential AI Regulations is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

Terms

- **AMI:** Advanced Metering Infrastructure
- **Anonymization:** Removing or masking personal identifiers in the data[1]
- **CPUC:** California Public Utilities Commission
- **DER:** Distributed Energy Resource. [3, page 58] These resources can consume or produce energy, such as an Electric Vehicle. The hope is to use them to manage peak and dynamic demand conditions.
- **DERMS:** Distributed Energy Resource Management System. This is a system to integrate and optimize the use of DERs, hopefully with standard interfaces for interoperability.[3, page 63]
- **EPSS:** Enhanced Powerline Safety Settings. These settings detect hazards on a power transmission line and shut off power quickly to prevent an ignition and thereby stop wildfires before they can start. [3]
- **EV:** Electric Vehicle
- **FIPPS:** Fair Information Practice Principles. These are Federal Trade Commission guidelines for fair information practices in electronic marketplaces: Notice, Choice, Access, and Security (with enforcement). [5, pages 15-16]
- **FPI:** Fire Potential Index [1, page 14]. The FPI model estimates potential fire size in acreage. "Landscape scale index used as a proxy for assessing real-time risk of a wildfire under current and forecasted weather conditions."[4, page 608]
- **HE:** Homomorphic Encryption is a form of encryption that allows some computations to be performed on ciphertexts, producing an encrypted result that matches the result of the same operations performed on the plaintexts. In simpler terms, with homomorphic encryption, one can *compute on encrypted data without decrypting it*. [1]
- **HFTD:** High Fire Threat District. "Areas of the state designated by the California Public Utilities Commission (CPUC or Commission) as having elevated wildfire risk, where each utility must take additional action... to mitigate wildfire risk." [4,page 609]
- **LoRa:** Stands for **Lo**ng **Ra**nge. LoRa is a physical proprietary radio communication technique.
- **LoRE:** Likelihood of Risk Event[4, page 631]
- **Pseudonymization:** Pseudonymization involves the **r**eplacement of actual identifiers (like account numbers) with random codes. The mapping from code to actual identity is kept secret by the utility.[1]
- **PSPS:** Public Safety Power Shutoff [3]. A Public Safety Power Shutoff is a risk-mitigation event occurring when an electric utility temporarily shuts off power to reduce the potential risk of a wildfire, for example when high winds might cause vegetation (trees) to come in contact with by utility equipment such as transmission lines.
- **Random Forest:** "Random forests or random decision forests is an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time.: [4, page 626]
- **Risk event:** An event with probability of ignition, such as wire down, contact with objects, line slap, event with evidence of heat generation, or other event that causes sparking or has the potential to cause ignition. The following all qualify as risk events:
  - Ignitions;
  - Outages not caused by vegetation;
  - Outages caused by vegetation;
  - Wire-down events;
  - Faults; and
  - Other events with potential to cause ignition.[4, page 614]
- **SME:** Subject Matter Expert
- **SMPC:** Secure Multiparty Computation. Using cryptographic protocols that allow multiple parties to jointly compute results (e.g. a sum or trained model) from their combined data without exposing their individual inputs. [1]
- **V2X:** Vehicle-to-Everything. [3, page 41] V2X technology enables customers to use their EVs as backup power sources during grid stress, outages, or emergencies by exporting power from their vehicle to power their homes, businesses, and communities.[3, page 46]
- **VM:** Vegetation Management
- **WDRM:** Wildfire Distribution Risk Model [2, page 5]
- **WIV:** Wildfire Incident Viewer
- **WMP:** Wildfire Mitigation Plan [2]
- **WTRM:** Wildfire Transmission Risk Model [2, page 5]

References

1. Westrich, Benjamin; Privacy-Preserving Analytics for Smart Meter (AMI) Data - A Hybrid Approach to Comply with CPUC

Privacy Regulations. May 13, 2025. Privacy-Preserving Analytics for Smart Meter (AMI) Data - A Hybrid Approach to Comply with CPUC Privacy Regulations. License: CC BY 4.0. Create Commons Attribution: This license allows reusers to distribute, remix, adapt, and build upon the material in any medium or format, so long as attribution is given to the creator. The license allows for commercial use.
2. Pacific Gas and Electric. PG&E Wildfire Mitigation Plan R0 2026-2028, Volume 1 of 2, April 4, 2025.
3. Pacific Gas and Electric. R&D Strategy Report 2024.
4. Pacific Gas and Electric. PG&E Wildfire Mitigation Plan R1 Volume 2 of 2. PG&E Wildfire Mitigation Plan R1 2026-2028, Volume 2 of 2, July 28, 2025.
5. Federal Trade Commission, Privacy Online: A Report to Congress, June 1998. Pages 15-16.

This page titled 7.8: Glossary of Terms is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

8: COVID-19 CONTACT TRACING APP

In this chapter we will cover COVID-19 contact tracing apps, or exposure notification apps. These were discussed during the COVID-19 pandemic with the goal to apply technology to a previously manual-only process of back-tracing the contacts that an infected person had while they were potentially infectious.

Using some combination of GPS location information, Bluetooth Low Energy (BLE)[1] proximity information, and cell phone tower location information, the goal was to develop a system to gather those contacts in order to notify them that they might have recently been exposed to an infected person (someone who tested positive for the virus), so that they could test themselves for COVID-19 and follow up to seek further medical help as required.

Collecting time-stamped location information presented a variety of privacy issues that, in jurisdictions including the USA, would typically require gathering the consent of the individuals in opting in to the program (by contrast, Article 6 of the EU GDPR allows data collection based on consent or on a different, predefined lawful basis; see chapter 3). Some governments mandated the use of their nation's contact tracing app. Perhaps in normal times people would be unwilling to grant such consent, but in a global pandemic, people might approach their decision analysis and risk assessment differently and thus might be more willing to take the privacy risk in order to seek therapeutic treatment as soon as possible, especially if they are in a high risk group (e.g., older, immunocompromised).

We will look into how these contact tracing apps were accepted by the public and where they were deployed. In order to be effective, a high percentage of the population needs to be using contract tracing apps, so overcoming privacy concerns is important to increasing the adoption rate.

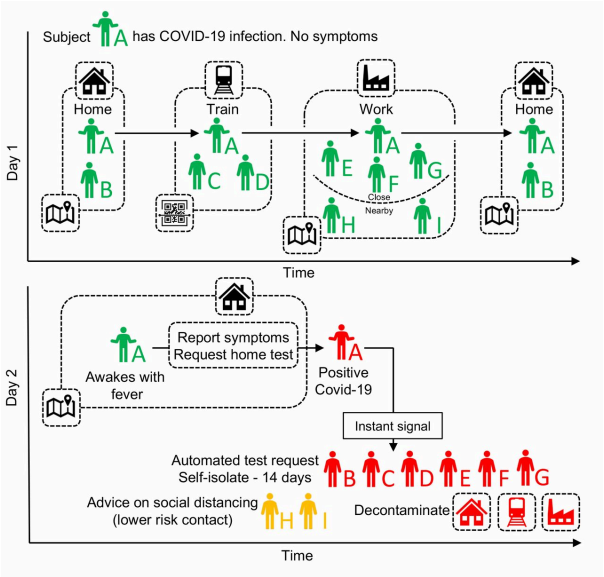


Figure 8.1: COVID-19 Contract Tracing App. (CC BY 4.0; Ferretti, Luca; Wymant, Chris; Kendall, Michelle; Zhao, Lele; Nurtay, Anel; Abeler-Dörner, Lucie; Parker, Michael; Bonsall, David; Fraser, Christophe[3])

Accessibility Note

In Figure 8.1 contacts of individual A (and all individuals using the app) are traced using GPS co-localizations with other App users, supplemented by scanning QR-codes displayed on high-traffic public amenities where GPS is too coarse. Individual A requests a SARS-COV-2 test (using the app) and their positive test result triggers an instant notification to individuals who have been in close contact. The App advises isolation for the case (individual A) and quarantine of their contacts.

In Figure 8.1, on Day 1, Person A is at home with Person B. Person A then takes a train to work and is in close proximity to Persons C and D. While at work, Person A is in close proximity to persons E, F, and G, but not as close to persons H and I. Then Person A returns home (apparently not showing any train travel) and is with Person B again. On Day 2, Person A wakes up experiencing symptoms and tests positive for COVID-19. Person A reports this through the COVID-19 contact tracing app, which in turn sends alerts to the people with whom Person A was in close proximity the previous day: Persons B, C, D, E, F, and G. That notification might alert those contacts to get tested and self-isolate for the time periods recommended by their nation's health agencies.

We will start with an academic paper covering the topic, and then review project management considerations, and legal considerations.

References

1. Wikipedia authors. Wikipedia - Bluetooth Low Energy. Accessed April 17, 2026.
2. Wikipedia authors. Wikipedia - COVID-19 apps. Accessed 3/8/2026.
3. Ferretti, Luca; Wymant, Chris; Kendall, Michelle; Zhao, Lele; Nurtay, Anel; Abeler-Dörner, Lucie; Parker, Michael; Bonsall, David; Fraser, Christophe - Figure 4 from Ferretti, Luca; Wymant, Chris; Kendall, Michelle; Zhao, Lele; Nurtay, Anel; Abeler-Dörner, Lucie; Parker, Michael; Bonsall, David; Fraser, Christophe (2020-03-31). "Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing" (archived at the Wayback Machine). *Science*. DOI link. Note: The May 8, 2020 version of the paper that now lives under the same DOI and (redirected) URL contains a very different version of figure 4 (e.g. without mention of QR codes)

Thumbnail figure: COVID-19 Contract Tracing App. (Ferretti, Luca; Wymant, Chris; Kendall, Michelle; Zhao, Lele; Nurtay, Anel; Abeler-Dörner, Lucie; Parker, Michael; Bonsall, David; Fraser, Christophe, CC BY 4.0(opens in new window))

8.1: Description of the problem and potential solutions

8.2: Project Management Considerations

8.3: Legal Considerations

8.4: Glossary

This page titled 8: COVID-19 Contact Tracing App is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

8.1: Description of the problem and potential solutions

Contact Tracing Apps: Lessons Learned on Privacy, Autonomy, and the Need for Detailed and Thoughtful Implementation

Katie Hogan [1,#], Briana Macedo [2,#], Venkata Macha [3], Arko Barman [4,5], Xiaoqian Jiang [6,✉]

Abstract

The global and national response to the COVID-19 pandemic has been inadequate due to a collective lack of preparation and a shortage of available tools for responding to a large-scale pandemic. By applying lessons learned to create better preventative methods and speedier interventions, the harm of a future pandemic may be dramatically reduced. One potential measure is the widespread use of contact tracing apps. While such apps were designed to combat the COVID-19 pandemic, the time scale in which these apps were deployed proved a significant barrier to efficacy. Many companies and governments sprinted to deploy contact tracing apps that were not properly vetted for performance, privacy, or security issues. The hasty development of incomplete contact tracing apps undermined public trust and negatively influenced perceptions of app efficacy. As a result, many of these apps had poor voluntary public uptake, which greatly decreased the apps' efficacy. Now, with lessons learned from this pandemic, groups can better design and test apps in preparation for the future. In this viewpoint, we outline common strategies employed for contact tracing apps, detail the successes and shortcomings of several prominent apps, and describe lessons learned that may be used to shape effective contact tracing apps for the present and future. Future app designers can keep these lessons in mind to create a version that is suitable for their local culture, especially with regard to local attitudes toward privacy-utility tradeoffs during public health crises.

Keywords: contact tracing, COVID-19, privacy, smartphone apps, mobile phone apps, health information, electronic health, eHealth, pandemic, app, mobile health, mHealth

Introduction

At the end of 2019, a novel coronavirus was determined to be associated with a group of pneumonia cases in Wuhan, China. In 2020, this novel coronavirus spread rapidly throughout China and the globe, prompting the World Health Organization to name the disease COVID-19 and the virus associated with the disease SARS-CoV-2 [1]. By October 2020, the COVID-19 pandemic had resulted in over 200,000 deaths in the United States [2] and over 1,000,000 deaths globally [3].

Strategies to monitor and control the spread of COVID-19 have hinged around a combination of traditional and nontraditional strategies, including rapid testing, self-quarantine, and contact tracing. Contact tracing has formed a key component of the plans to control the spread of infectious diseases in recent years, yielding a wealth of literature concerning its importance and efficacy. Contact tracing traditionally involves interviewing infected individuals and following up with any close contacts to communicate increased risk due to exposure and provide information and strategies related to that risk [1]. Alongside face-to-face contact tracing, various digital strategies have been employed to mediate this process and increase efficacy, including the introduction of a wide variety of contact tracing apps. In an article modeling the efficacy of contact tracing for Ebola, Browne et al [4] identified and examined key epidemiological parameters that impact the efficacy of contact tracing, including incubation period, infectious period, and monitoring protocols. During recent Ebola outbreaks, contact tracing apps developed in Guinea and Sierra Leone have provided a means for contact tracers to increase the speed and accuracy of contact tracing and efficient centralization of real-time data, as well as the coordination of resources and interventions [5,6]. These studies suggest that technology may play a key role in increasing the efficacy and timeliness of contact tracing. In contrast to the studies above, which placed technology in the hands of contact tracers, most strategies for COVID-19 contact tracing implementations use available smartphone technology to actively monitor risk for users in the general population.

A study by Ferretti et al [7] modeled the use of contact tracing apps and concluded that widespread use of these apps could be used alongside strategies such as widespread testing and physical distancing to suppress the pandemic successfully. This model suggests that contact tracing apps could allow greater freedom of movement, but low app adoption or incorrect usage could lead to continued spread. Similarly, models developed by Yasaka et al [8] determined that although some disease suppression could

be seen with 25% adoption of their app, higher population uptake (75%) would be required for more substantial reductions of the infection curve. A primary concern, therefore, in determining the potential efficacy of a contact tracing app is the number of people using the app and the efficiency and accuracy of its information distribution. Rowe [9] outlined three conditions necessary for the success of a contact tracing app: (1) correctness of information including diagnosis, (2) high likelihood of smartphone presence during contact, and (3) a high proportion of people using the app. As a result of the myriad technological changes that have occurred in recent years, utilizing digital technology to perform contact tracing is both a truly promising solution and an unprecedented problem. Therefore, extracting lessons from the current pandemic and the role technology can play will serve as crucial components to optimizing public health strategies during current circumstances and future outbreaks.

There has been a wide variety of COVID-19 contact tracing app reviews published over the course of the pandemic across a range of subjects, including technical analysis [10,11], privacy concerns [12-14], and ethics [15-18], each of which present valuable information on a specific aspect of digital contact tracing. This viewpoint is designed to combine information from across the expanding COVID-19 digital contact tracing literature and address key considerations that must be taken into account for developing and refining future contact tracing app design. This paper aims to accomplish this goal by providing background on common strategies for app-based contact tracing, discussing the advantages and limitations of several prominent COVID-19 contact tracing apps, and elucidating the privacy and nonprivacy concerns that have affected their adoption and reliability during the COVID-19 pandemic. Table 8.1.1 summarizes the key points of each section. This paper is intended as a hermeneutic literature review, with analysis that provides a specific viewpoint on the subject at hand. As such, a literature review was conducted over the course of July 2020 to March 2021 via PubMed and Google Scholar using terms including COVID-19, contact tracing app, and digital contact tracing, etc.

Table 8.1.1: Key considerations for the design of contact tracing apps, including those that relate to structure, technology, adoption requirements, and privacy.

App characteristic	Considerations
App structure	Centralized • Generated user data stored or managed in a central server • Direct oversight of user data Decentralized • Generated user data stored or managed on user devices • User privacy and security advantages
App technology	GPS and location-based tracking • Additional potential for privacy concerns • Higher noise compared to Bluetooth • May be paired with proximity-based tracing for increased accuracy Bluetooth and proximity-based tracing • Fewer privacy concerns, particularly when paired with a decentralized structure • High noise—signal attenuation and accuracy issues due to environmental signal absorption and reflection
Adoption requirements	Mandatory • May be viewed as a privacy violation • Higher adoption rates may increase the accuracy and efficacy of the app Opt-in • Voluntary use with specific permissions to address privacy concerns • May have lower adoption rates that decrease the accuracy and efficacy of the app

App characteristic	Considerations
Privacy	• An app should offer privacy from other users, the app manager, and snoopers • The privacy-utility tradeoff of an app must be shaped around the local cultural attitudes • Privacy and security have been repeatedly stated as primary concerns for app users: successful apps should adopt a privacy-by-design structure
Other	• An app should be easy to use and reduce user fatigue • Battery drainage, interference with other medical apps, and incompatibility with some phone models have proved to be barriers for successful global deployment of some contact tracing apps • Users must be encouraged to follow all other precautions to limit the spread of disease (ie, recommendations from public health authorities like mask wearing and physical distancing)

Key Considerations in App Design and Categories

In this section, we outline some of the common considerations for developing contact tracing apps, namely strategies and technologies employed. We discuss the advantages and disadvantages of each of these strategies, particularly with regard to efficacy and frequency of use of contact tracing apps globally.

App Parameterization Using Epidemiological Data or Disease and User Characteristics

When developing new apps for contact tracing, several factors must be accounted for within algorithms and interfaces to ensure accurate information and notifications for all parties involved. For instance, rates and types of transmission must be incorporated into the app design in order to determine the number and time frame of contacts that must be notified and tracked. The basic reproduction number (R_0) accounts for baseline disease transmissibility without immunity from exposure or vaccination or any intervention to prevent transmission [19]. Transmission for COVID-19 falls into the categories of symptomatic, presymptomatic, asymptomatic, and environmental transmission [7]. Contact tracing generally most accurately accounts for symptomatic and presymptomatic transmission, as asymptomatic and environmental transmission may not be readily identifiable. However, when contact tracing is paired with large-scale community testing, there is an enhanced ability to model transmissibility, accounting more accurately for asymptomatic and environmental transmission.

Incubation and infection times must also be taken into account to narrow time windows over which contacts should be identified for potential exposure. For COVID-19, the average time to symptom development is 5.1 days, with 99% of symptomatic cases displaying symptoms by 14 days [20]. In a study of infectiousness profiles for COVID-19 infector-infectee transmission pairs, the highest viral load was observed at symptom onset, translating to increased transmission risk [21]. However, 44% of secondary cases developed as a result of exposure during the index patient's presymptomatic stage, allowing COVID-19 to spread rapidly and highlighting a heightened risk when the perceived threat is low. Indeed, the disease's highest transmissibility has been reported to be before or immediately after symptom development [22]. Therefore, when modeling the timescale over which a patient may have been infectious, contact tracing apps must account for possible transmission up to 2 weeks before symptom development or a positive test result for asymptomatic carriers. With the incorporated passive observation of individual location or contacts, this strategy may help to account and partially compensate for diagnosis delay, which was observed to lead to increased spread, particularly as communities developed and launched containment strategies early in the pandemic [23]. These parameters are essential in initial app development and determining individual risk but should also be considered when determining the amount of time necessary for data storage, particularly given the privacy concerns voiced over the long-term collection of location and health data [14].

Additionally, some thought should be given to the direction of contact tracing with the collected data. Forward tracing, used almost exclusively within current contact tracing apps, works forward from a current positive diagnosis to identify the contacts, particularly during peak infectiousness, who are now at risk of contracting the disease [24]. Backward tracing, however, works back from the current diagnosis to identify the secondary origin of transmission and find additional contacts who are now at risk of having the disease [24,25]. This method of tracing seeks to find the source of an outbreak and has been successful at

identifying clusters around an individual known to be contagious. For instance, backward tracing was used to identify clusters and prevent further community spread of COVID-19 in both Japan and Singapore early in the pandemic [24,26]. Likewise, combining these techniques in bidirectional tracing has been shown to be particularly important with a long incubation period, with models showing a 2-fold reduction in effective reproduction number versus forward contact tracing alone [24]. Currently, most contact tracing apps exclusively use forward contact tracing, which fails to take advantage of the full range of data available. Digital tracing allows for easier extension of the tracing window in a bidirectional manner because there is no need to rely on patients' memories. However, if app usage is not high enough within a local population, digital tracing in either direction may be disrupted by network fragmentation and insufficient data [24]. Therefore, in addition to the epidemiological data necessary for app parameterization, due diligence must be given to increasing app usage numbers to increase the efficiency of a particular strategy, even those shown successful in manual tracing and modeling.

When developing app parameterization and settling on technological strategies and techniques for implementation such as those discussed below, app developers must ultimately choose strategies that result in the highest efficacy and accuracy. This strategizing must consider in particular app false positive and false negative rates for the app. False positives (type I error), referring to users incorrectly notified of increased exposure risk, can put undue strain on health care infrastructure by increasing demand for testing, result in increased levels of highly negatively impactful enforced quarantine, and decrease app utility by decreasing user attentiveness to app notifications [10,27]. False negatives (type II error), conversely, refer to individuals who have had close contact and are at high risk but are not identified by the contact tracing app, which may be the result of low app sensitivity or improper tracing [10]. These high-risk individuals are ignorant of their risk and may contract the disease and spread it further in the community. The best policies and parameters based on this reasoning seek to minimize false negatives first, as these will result in further untracked community spread. These strategies, however, will likely result in a high number of false positives due to prioritization of high-sensitivity, low-specificity methods [27-29].

False positives may result through a variety of means, but initial planning to prevent false positives must begin with the technical strategy chosen and the policy-based design of initial parameters. The definition of close contact (6 feet or 2 meters, 15 minutes) most commonly put forward by public health entities has been argued to be too coarse for mass tracing, as evidenced by the high number of false positives seen with manual contact tracing and that this definition has resulted in decreased accuracy with digital contact tracing as well [27,28]. Likewise, as will be discussed further below, the technical strategies employed, such as GPS or Bluetooth, will directly impact the accuracy and efficacy of an app [11,30]. Signal strength and duration selections for the app, as well as firmware and software compatibility with the app and other users' devices, will play a role [27]. These considerations make it imperative that apps undergo significant real-world testing to determine efficacy, data from which has not yet been revealed for most available COVID-19 contact tracing apps.

Centralized Versus Decentralized Architecture

The decision to utilize a centralized, decentralized, or hybrid overall structure or strategy is a key initial consideration when designing and implementing contact tracing apps and requires balancing privacy, security, and efficacy concerns. Centralized apps use strategies that employ a main server for data storage and analysis. Conversely, decentralized apps feature data storage that is distributed across the user network, with no individual entity having complete control or information access [31]. A hybrid architecture may have a component of both approaches, with some information handled on individual devices with a central server analyzing data and sending notifications. Contact tracing apps using each of these architectures have been employed for COVID-19, with the choice of structure highly dependent on government and cultural norms in the region of use and the needs of public health officials.

A centralized contact tracing app architecture may require significant trust in the beneficence of government investment and national or regional data infrastructure. In a centralized approach employing technology like Bluetooth LE, for example, a trusted third party (TTP) such as a government or public health entity may assign users an encrypted identifier that is broadcasted during app use [32]. These encrypted identifiers are broadcast to other users, with apps storing identifier lists that may be sent to the main TTP server in the event of a positive diagnosis. Then, users who appear on this list will receive notification of risk from the TTP through their app. Some notable examples of COVID-19 contact tracing apps with a centralized architecture include Singapore's TraceTogether and China's Health Code apps [33,34].

Conversely, decentralized architectures remove the role of data accumulation and analysis from a central server and instead place these functionalities on user devices. Anonymous user identifiers are generated as random seeds with a short lifetime (chirps) that are exchanged with other user devices. Upon a positive diagnosis, a user may upload seeds and accompanying temporal data to a central server. By analyzing seeds and temporal data from positive users on the central server, exposure notifications are generated on user devices. In this way, the central server does not serve as the primary driver of risk assessment, nor, due to the anonymized or pseudoanonymized nature of seeds and chirps, can identifying details be derived from information

available on the central server [10]. There has been a significant shift in interest toward decentralized or hybrid app architectures in recent years for privacy and security reasons, with a model focused on location privacy also indicating lower data retrieval times compared to centralized architectures [35]. Centralized approaches may be at risk of a variety of privacy and security attacks, including deanonymization of personal data through direct breaches of the main server. Likewise, decentralized app structures may also yield privacy and security concerns. For example, while Bluetooth LE and other technologies that have been applied in decentralized structures may record proximity-based data instead of positional data, personal medical information may be extrapolated via a linkage attack that uses data concerning specific, close contact notifications and known positional information from other devices to determine disease status, as explored by Bengio et al [12] in an excellent recent review of security concerns for decentralized app designs. Therefore, the decision to utilize a centralized versus decentralized app strategy centers on tradeoffs between culturally specific privacy concerns and structural trust needed for high population uptake as well as security concerns.

Additional security measures can and should be developed for both centralized and decentralized structures. Alongside encryption and anonymization techniques, blockchain technology has been suggested as a tool for increasing security and data privacy, particularly for decentralized strategies. Blockchain technologies seek to decentralize data by duplicating and distributing information across a global computer systems network, making it difficult to spoof or manipulate data within these distributed databases [36]. Xu et al [37] have developed a blockchain method called BeepTrace, designed as a bridge between users and central servers that reduces the vulnerability of sensitive data such as user identification and location, with the ability to predetermine and regulate the length of data storage. This proposed blockchain would also allow for public accountability of governments and corporations via transparency and ease of information verification, as in the case of manipulated efficacy data. However, in practical implementation, strategies would require refinement to reduce the intensive computational requirement for large populations, particularly dense populations such as in India and China.

Location- and Proximity-Based Tracking

GPS

GPS and global information systems (GIS) are routinely used for large-scale disease monitoring and predicting disease spread, as seen with features such as Google Flu Trends, which have been used to predict the spread of seasonal flu in the United States [38]. Similarly, for COVID-19, online macroscale tracking systems were developed early in the pandemic to allow for up-to-date information on the global and regional spread of the virus. The timely updates of these online GIS and mapping dashboards, including the Johns Hopkins University Center for Systems Science and Engineering dashboard and Who Health Organization Dashboard, provided a means of data sharing concerning outbreak events for the public [39].

Alongside macroscale information, GPS and GIS technology may be employed for individual contact tracing via GPS and social media mapping. The ubiquitous use of GPS-enabled smartphones in many regions of the world provides an opportunity to collect spatiotemporal trajectory data for individuals. One example of this is the STRONG (Spatiotemporal Reporting Over Network and GPS) strategy, proposed and published by Wang et al [40], which analyzed the backend GPS spatiotemporal data collected through the social media app WeChat to trace the close contacts of users, primarily in China. This strategy offers a means of integrating cell phone–based GPS positioning with voluntary real-world transaction data that provides accurate timestamps and position information, although this also requires a substantial relinquishment of individual privacy. Similar to this strategy and adopted soon after the STRONG system was published, the Chinese government began employing a national monitoring system, which, as opposed to GPS data, uses cell phone base station positioning data to evaluate individual exposure and risk [40,41]. Other countries had also employed government-sponsored cell phone location data collection to track individuals and prevent COVID-19 spread, including South Korea, which notified users before entry into “high-risk” zones, and Israel, which used location data to inform contacts of confirmed infected individuals and for quarantine enforcement [42]. Additional apps developed by countries that involve location-based tracking include the Rilevatore Teramoto app released by Italy and apps from Austria (NOVID20), India (Aarogya Setu), Norway (Institute of Public Health app), and Spain (Open Coronavirus), which combine location and proximity-based tracing [32].

Therefore, GPS and location-based mapping have been widely viewed as a potential tool within the context of COVID-19 contact tracing apps. However, further refinement is needed to reduce “noise” within GPS-based systems, which may reduce the efficacy of the system to identify high-risk contacts or areas with sufficient specificity [40]. The granularity of location-based GPS data may not be sufficient to determine a distance of 6 feet (2 meters), with GPS positioning error at approximately 10 meters indoors [11]. This limitation may lead to significantly increased false-positive and false-negative rates, and hence reduce the accuracy and efficacy for GPS-based contact tracing strategies. Bluetooth proximity-based systems are generally considered to be more accurate, with fewer false positives reported [32]. Additionally, privacy concerns arise with the thought of highly

specific spatiotemporal tracing, particularly when combined with timestamped real-world interactional data available through social media networks, which may generate problems with adoption. Security concerns must also be considered, as GPS systems are particularly vulnerable to spoofing attacks, in which a fabricated GPS trail may send incorrect spatiotemporal data to a receiver [43].

One example of a privacy-forward, GIS-based app is the COVI Canada app, initially proposed for government use, although the Canadian government has moved in another direction [44]. The proposed app employs GeoIP services to yield coarse location data that maps risk-stratified zones through which users have moved. Additionally, Bluetooth-based contacts are recorded among individuals. COVI Canada claims a decentralized approach in which pseudonymized personal data such as the number of contacts and diagnosis status are encrypted and sent to the main server where risk is calculated for each user daily via artificial intelligence (AI). The coarse location data yielded by this approach may also provide heat maps of outbreaks to public health authorities.

Bluetooth

In contrast to GPS mapping, Bluetooth records interactions between individuals based on device proximity. When individuals are in close proximity, Bluetooth-based token sharing allows for a precise record of the interaction. In this case, the limited range of Bluetooth-based technologies becomes advantageous for recording only close-range interactions to approximate a 6-foot distance. Additionally, the relative signals' strength can be used to determine the approximate distance between individuals, allowing for proximity tracing of individuals in high-risk settings, such as indoor environments or public transportation [32,45]. However, this strategy requires high adoption for accurate proximity tracking and risk assessment, as it requires a direct interaction between users. Bluetooth has an approximately 10-meter location granularity, with visibility between devices possible up to 30 meters apart [11]. Signal attenuation may be used to indirectly assess distances between devices, although this is not linear. Additionally, high standard deviation in received signal strength between 2 and 6 meters decreases accuracy within this range, and with 2 meters used to indicate close contact, it may lead to decreased app efficacy and increased false positives using Bluetooth technology.

Additionally, although Bluetooth-based tracing is generally considered more accurate than GPS-based strategies, significant issues are present with signal attenuation and accuracy. Signal absorption and reflection by the surrounding environment can lead to inaccuracies in reported distances between users [46,47]. In a recent paper on this topic, Leith and Farrell [48] evaluated the efficacy of Bluetooth LE technology for COVID-19 contact tracing in real-world environments, including users walking in the city, at a meeting table, in a train carriage, and grocery shopping, as well as assessing the impact of device orientation, use of a handbag, and type of indoor wall on signal attenuation. After observing significant impacts from all of these factors and no corresponding decrease in signal strength with increasing distance, the authors called for extensive real-world testing of Bluetooth-based COVID-19 contact tracing apps as well as data to inform the efficacy of such apps compared to manual contact tracing.

Bluetooth-based strategies have become common, with some notable examples, including national apps in several countries. Singapore's TraceTogether app was the first such Bluetooth-based, government-sponsored contact tracing app [33], and since then, the list has grown to include the Pan-European Privacy-Preserving Proximity Tracing (PEPP-PT) app, released by a European consortium, as well as Australia's COVIDSafe [49] and Canada's ABTraceTogether [44]. In contrast, the industry-based joint Apple-Google contact tracing app also relies on Bluetooth proximity tracing, but concerns have arisen around private companies' handling of sensitive health data and this new role for large tech corporations in the public health arena [50].

The PEPP-PT app was initially touted by its creators as having a 90% true-positive and 10% false-negative rate [51]. However, several studies since have indicated that Bluetooth-based proximity tracing may have significantly high error rates. For instance, in a study by Girolami et al [52] involving high school students, a traced interaction accuracy of 81% was obtained. However, in initial planning, 42% of student devices were found to be incompatible with active Bluetooth beaconing and unusable for contact tracing, which would pose a significant issue in the population at large. In a real-world setting, devices will have different versions of Bluetooth technologies and may receive, transmit, and damp signals at different levels. Likewise, at least 50% beacon loss between device dyads was observed, posited to be due to device positioning and unpredictable environmental signal attenuation as discussed above. This study, alongside real-world testing of the Bluetooth-based Google/Apple Exposure Network (GAEN) by Leif et al [48], points to somewhat unreliable and unpredictable efficacy and utility for Bluetooth-based apps and demonstrates the need for further innovation and design.

Contact Points

One alternative to location-based GPS or Bluetooth tracking, which allows for more user privacy, uses an opt-in system of contact

points. Yasaka et al [8] proposed the idea of an app that allows users to host or join checkpoints at which other users may also check in by scanning a generated location quick response (QR) code. Ideally, users would generate check-in QR codes for any gathering or public place, which poses a risk for COVID-19 transmission, and when a user tests positive, all users who have checked in at locations with them over the potential infectious period will receive an updated risk level. High user adoption would be necessary for such a system to be efficacious. Additionally, apps that rely on sustained conscious use pose a problem with user fatigue, with drop-offs in check-ins providing the potential for inaccurate risk assessments.

Volunteered Health Status

An alternative to tracking or tracing apps is an app that allows for volunteered health or contact information to be entered to initiate case-based contact tracing outside of the app itself. This tactic of voluntary information entry mitigates some privacy concerns centered around tracking and tracing. The two main categories of apps falling into this category are symptom monitoring apps and case-initiated notification systems. These apps do not directly perform contact tracing but instead serve to use technology to simplify the contact-tracing process.

Symptom monitoring apps rely upon users to accurately record a log of personal recent health records to assess for COVID-19 risk based on symptomology and may be used most effectively as a screening tool for identifying symptomatic cases. These apps require regular logging of recent symptoms associated with COVID-19, including fever, dry cough, shortness of breath, fatigue, muscle aches, and loss of taste or smell. Yamamoto et al [53] reported an example of this system integrated into an already established health screening app, K-note, which significantly decreased the follow-up burden on health care providers for monitoring close contacts of COVID-19–positive cases, allowing for efficient algorithmic identification of symptomatic users. Additional personal data that might be logged includes COVID-19–positive contacts, travel history, and visits to health care institutions, allowing for further risk stratification. Another such app includes the one implemented by Yap et al [54], COVID-19 Symptom Monitoring and Contact Tracking Record (CoV-SCR), which provides a designated space for users to track symptoms (rated 1-5) as well as keep a 14-day record of travel and close contacts whom the user may notify in the event of a positive diagnosis. Medical and academic institutions have already employed such efforts to keep a regular log of potential symptoms as a means of preventing spread, providing a way to do a first-pass screening of individuals and identify those who may need to be tested or isolated. Additionally, symptom monitoring may be paired with in-person screening, such as temperature checks upon entry to an institution. While many such apps are institution-specific, such symptom logging may be integrated into currently available personal health record apps.

Early identification of potentially COVID-19–positive individuals may enable efficient tracing of contacts and early isolation. However, such apps rely upon the memory and honesty of users to accurately portray health status. Additionally, this strategy fails to account for asymptomatic individuals, who may slip through the holes in this system and further spread the virus. Users may also be unable to account for strangers or individuals encountered in public settings such as public transportation. Passive tracing and tracking apps have therefore become the preferred avenue of exploration and implementation for apps intended for large-scale use, as discussed in the examples below, most of which seek to combine symptom monitoring with location- or proximity-based tracing.

Mandatory Versus Voluntary Use

There are three potential strategies for enforcing such an app at a federal level: opt-in, opt-out, or mandatory use. An opt-in model allows people to download the app if they so choose and has been advocated for on the basis of consent for acquisition, use, and sharing of personal information [23]. An opt-out model would automatically provide the app for all, but users would have the option to delete the app if they preferred not to use it. The final model is a mandatory download of the app for all people without deleting it. Mandatory app usage may be enforced by preventing people from using public services or entering buildings if they do not have the app. China, Hong Kong, Taiwan, and South Korea have used the mandatory model [55]. Many ethical concerns have been raised about such a design, including rights to personal autonomy and informed consent [16]. However, another concern is widening disparities between those who have access to relevant technology such as the internet and Bluetooth-capable smartphones and those who do not. In addition to an inability to move freely or engage meaningfully in public life, those who do not have smartphones or new operating systems do not have access to incentives to increase app use, such as those Parker et al [18] suggest like funds for charity donations and free mobile phone credits. Overall, the ability of a country to enforce a mandatory download system will depend on the socioeconomic status and cultural values of its citizens, with some nations more likely to experience significant pushback against or widespread inefficiencies in a mandatory system.

Furthermore, once the app is downloaded, behavior on the app may be mandatory/opt-in/opt-out. For example, once the app is downloaded, even if the app itself is optional, location sharing may be nonvoluntary and continuous. Other apps may allow for

opt-in for location sharing. Although this may minimize the app's efficacy, this approach allows for greater user privacy if they are venturing somewhere that is private to them. Additionally, apps may require mandatory sharing of infection status, as is the case in Singapore.

Specific App Examples of Successes and Limitations

In this section, we outline some of the large-scale implementations of contact tracing apps, most notably those created by national governments. We will discuss the successes and downfalls of some of these implementations and touch on how local cultural attitudes influenced the design of each app and how it was received by the local population.

Singapore

Singapore's TraceTogether app was the first national deployment of a Bluetooth-based contact tracing system in the world. The app was presented as being "for the people," with the ultimate goal of protecting the population [33]. This Singaporean technology provides several lessons for contact tracing, including concerns about Bluetooth, the privacy-utility tradeoff, as well as centralized and decentralized systems. Most importantly, the Singaporean contact tracing strategy reveals how cultural attitudes and norms must be taken into account within a community in order to achieve maximal success.

Aimed at transparency and international cooperation, the Government Technology Agency of Singapore published information about BlueTrace, the protocol that underpins TraceTogether as well as OpenTrace, an open-source repository for other countries that heavily influenced Australia's national app design [56]. The app features a hybrid decentralized-centralized, proximity-based approach and functions through contacts exchanging non–personally identifiable messages, with frequently rotated identifiers for security and privacy. Encounter history is kept on local storage and thus is decentralized. While the app itself is not mandatory, once a user is tested positive, they are legally required to release their stored data to the government per the Infectious Disease Act [55], after which health officials reach out to contacts based on personal identifiers. At this point, a user's data is stored on a centralized government server [57], and the government retains the right to share this information with other groups, including other governments [55,58]. It should be noted that these requirements would not be the same across countries and communities. The government generally advertised the TraceTogether app as being a privacy-by-design system because of the decentralized setup [59]. However, this is only true for healthy individuals who do not have to share their data to the centralized server for government control.

The BlueTrace authors also note other privacy-by-design implementation choices for the app. The TraceTogether app claims to keep proximity data only for 21 days, therefore limiting the amount of unnecessary data stored [56]. Furthermore, the BlueTrace report claims that users have control over their data, and all of their information is deleted upon request [56,59]. The TraceTogether system requires a health official to confirm that a case is legitimate to avoid false self-reported positives that may stir up panic and decrease the program's legitimacy and trust. However, the additional step of health official approval may be difficult to implement in other nations with high rates of uninsured citizens, who may not be able to seek medical care. This concern will inevitably affect some countries more than others, notably based on the presence or absence of universal health care.

Since the system requires interaction between two users, an increased percentage of people who download the app increases the program's effectiveness quadratically [56]. While Singapore was reaching relatively high usage as compared to other opt-in/opt-out programs, the 17% download rate was still insufficient to reach maximal efficiency [60]. However, the government hesitated to make the app mandatory to prevent pushback regarding surveillance and control concerns. There were some concerns that the Singaporean government was collecting cell phone data through the app, although the government denied these claims [61].

The app itself had several issues in terms of functionality. For example, certain types of phones, including Apple products or older models, could not download the app or experienced severely limited functionality. Further, there were lags between user contacts and device communication and logging. There are also inherent limitations to Bluetooth, including material barriers that attenuate transmission signals. Additionally, Bluetooth technology can interfere with other health-related apps and implantable devices [62]. High variance in transmission power across device types may also lead to difficulty in assigning appropriate thresholds to determine close contact distances relative to signal strength. Battery usage issues have also been reported [56].

To address some of these technical issues and privacy concerns, the government eventually pivoted to distributing a wearable device to all of its citizens, which would complement the TraceTogether app. These devices would not contain any information beyond contact history, thereby protecting mobile cell phone data. Further, since the devices were identical, they would bypass some of the aforementioned technical issues across phone models and provide access to all. However, this implementation did

not reduce concerns about location surveillance. In contrast to GPS, which tracks location, Bluetooth tracks interactions, which means that interactions between lawyers, doctors, and journalists are no longer confidential to the government [55]. Further, even with Bluetooth, it is still possible to narrow down one's location, especially as data accumulates [57].

Chua et al [61] notes that America may be unable to mimic Singapore's system because of cultural attitudes toward privacy-health tradeoffs. This is generally true across all distinct countries, which will each require individual policies that fit their population. The app developers themselves note that this system is specific to Singapore [58].

Apple/Google

On April 10, 2020, Apple and Google announced a joint effort to create a contact tracing app framework that would serve to facilitate cross-platform monitoring as public health officials globally developed apps for their respective jurisdictions [59]. Phase 1 of this effort focused on releasing an application programming interface (API) for interoperability between iOS and Android systems, which was released to public health officials in May. This release was accompanied by signs that these corporations had reached out to large nations to create some patches between their initial health app designs and the GAEN API, notably Australia [59]. Following this, phase 2 released a means of building access to public health apps into the iOS and Android platforms through an Exposure Notification app provided with software updates. This system allows for Android and iOS users to opt in to an Apple and Google initiative that also ties in local contact tracing apps [59].

The API functions allow Android and iOS devices to exchange data with each other for improved contact tracing. Additionally, this framework features an opt-in, decentralized contact tracing system that relies on Bluetooth technology. The main emphasized advantage of the GAEN API was initially privacy, with no mechanism for recording users' location data. A new Temporary Exposure Key is generated once daily, with Rolling Proximity Identifiers (RPI) generated every 10 minutes. Beacons with RPIs are broadcast every 250 ms, while devices scan for beacons every 4 minutes. The signal strength of received beacons is used to determine distances between users, while the number of beacons may determine the duration of contact exchanged [45,50]. Close contact is generally defined as 15 minutes of contact at distances of less than 2 meters. The exact parameters that qualify a signal as a significant contact may be determined by public health officials, with variable attenuation thresholds chosen by the nation [45]. GAEN API–based apps have been utilized worldwide, particularly within Europe (eg, Germany, Switzerland, and Italy).

The value of Bluetooth-based apps is a precise record with contacts who may be excluded or difficult to trace with traditional contact tracing methods, such as strangers in public settings. However, GAEN API–based apps have recently come under scrutiny for issues with efficacy in these scenarios. Leith and Farrell's [45] study sought to apply the GAEN API in a real-world public transportation setting, recording measurements on a commuter train between handsets at greater and less than 2 meters for 15 minutes. Using Swiss and German attenuation thresholds from apps released in late May and early June, no close contacts were recorded. The Italian attenuation threshold yielded 50% true positives with a 50% false-positive rate. While this study only used Android devices, it yielded data that suggests significant changes in noise levels based on individual use. Additionally, signal strength was demonstrated to have complex interactions with the environment in which measurements took place, including signal reflection by tram walls and absorption by bodies (at 2.4 Hz), with no clear trend in signal attenuation with respect to distance based on these complex effects. Although more such studies are required, this data suggests that the inherent flaws of Bluetooth-based contact tracing, with small changes such as models of devices and signal absorption and reflection having an outsized impact on outcomes, may significantly decrease the efficacy of the GAEN API. This limitation may not be unique to the GAEN API and is indeed of concern for all Bluetooth-based contact tracing apps.

Although this initiative was initially well received, in addition to issues of efficacy, several concerns surrounding the imposition of large tech corporations' influence in the spheres of public health and politics have arisen. These doubts have centered around a questionable past for both corporations regarding data management as well as the potential for mission creep for large tech companies [50]. The increasing inroads of tech corporations into biomedical fields, as in developing AI medical diagnostics, electronic medical records systems, and software kits for clinical trials, have some worried that in the rush to develop contact tracing apps, traditional medical expertise and professional knowledge has been traded for efficiency and optimization, which are the key values of technological production [50]. In addition to this, upon release of the GAEN API, Apple and Google have refused to work with nations and public health entities with apps in development that feature a centralized approach. This strategy effectively undermined government attempts at app creation in France and Latvia by refusing technical expertise [50]. Agencies have been forced to create workarounds for their apps that are unstable and battery-draining [59]. This lack of good-faith effort in these situations has been seen as evidence that public health and privacy concerns are currently being used as a means of increasing market share through contact tracing app development.

Ireland

Ireland's COVID Tracker app has been highlighted as a successful adaptation of the GAEN API for national use. As such, the app functions as a decentralized, Bluetooth-based system aimed at preserving individual privacy [62,63]. One of the most significant accomplishments of the opt-in app was its relatively rapid adoption, with 37% of the population having downloaded it within a week of release [62,63]. Over the course of July and early August, 137 users had been notified of their potential exposure to COVID-19 [62]. The success of the Irish app has led to its government working with other nations to retool the app for their use, including the United States [62].

Prior to launch, Irish health officials requested feedback from academic researchers and data and visualization experts as well as civil societies to evaluate the app, which has been noted as a potential source for public trust and perceived efficacy [64,65]. According to the prerelease report card issued by the Irish Council for Civil Liberties and Digital Rights Ireland, however, there were concerns regarding app privacy and security structure [64]. Most significantly, these groups noted the lack of efficacy data to back up claims of high accuracy, the need for timely deletion of personal data which could be extrapolated to yield user location, and concern over the use of closed-source Apple/Google software and control of health data by foreign private entities.

China

The Chinese tech group Alibaba released their Alipay contact tracing app Health Code on February 9 in Hangzhou [33,66]. Soon after, Tencent released a similar system on WeChat. The QR code-based Health Code app has a broad audience in China, with the app used in over 300 Chinese cities with at least 900 million users as of August 2020 [33]. Through public-private partnerships and centralized government monitoring, Health Code has now become mandatory for access to public spaces in many areas. The app uses a color-based QR code system that reveals user risk—green, yellow, or red. Those with a green QR code may visit public spaces and others, while those with a yellow or red QR code are subject to self-isolation for 7 to 14 days [33,67]. App data pairs actively collected data such as self-reported symptoms, address, and government ID with passively collected GPS location data, online transaction data, and surveillance via facial recognition technology, CCTV (closed-circuit television), and drones, all of which are monitored and synthesized by a central government server to determine exposure risk and generate the QR codes [66-68]. The users' status is updated daily at midnight to account for the previous day's activities [33].

The mandatory acceptance of this app has been cited for enabling greater effectiveness versus voluntary acceptance for reducing the spread of the virus. Additionally, the app's widespread use has demonstrated a much stronger tracing record versus South Korea, which only requires diagnosed individuals or close contacts to download their app [67]. However, the mandatory, centralized structure of this app may be responsible for considerable user stress. In one study by Joo and Shin [67], users reported that issues with app inaccuracy and errors were less stressful overall than privacy concerns. Because public spaces are restricted, even without a mandate, a de facto mandatory environment would exist for users who want to participate in public life, such as going to work or seeing family members, and could exclude those without access to appropriate technology or reliable internet access, highlighting the complexities of regulating movement based on exposure risk [66].

Likewise, significant concerns have been raised regarding erroneously issued yellow and red QR codes due to incorrect data entry or technical errors, which may necessitate unwarranted self-isolation [67]. Conversely, false-negative green codes have been reported in Wuhan, with COVID-19–positive individuals given a pass to public areas [33]. Users have reported difficulties in having an incorrect code corrected within the centralized system, adding to technological stress associated with the app [67]. Inconsistencies at the local level have added confusion to this process as users move between regions, with a yellow code requiring 7 days of quarantine in Hangzhou versus 14 days in Shandong [33]. These issues have been paired with a lack of transparency on the part of the government regarding the app's operation [67].

Additional concerns include a reliance on private tech corporations to provide location data, echoing concerns of public-private partnerships that ring similar to those voiced about the Apple/Google GAEN API [67]. Alarms have been raised regarding the potential for Alipay and WeChat to share information with police agencies [33]. Likewise, this public health role provides large tech companies with unprecedented access to individual health information, which concerns some users due to the overall lack of transparency in this process. It remains unclear how exactly public and private entities manage data collected through the Health Code, who owns this data, and how the government regulates the Health Code [33].

South Korea

The South Korean contact tracing app, Self-quarantine Safety Protection App, was mandatory for all citizens, and data was stored in a centralized database. Citizens could then view the database and see people's whereabouts to determine if they were at risk

[68]. One critique of this system was that users had to read through a lot of information daily, such that users stopped actively checking the updated information. Another app, Corona 100 m, created a more streamlined service to identify infection hotspots to avoid information fatigue [69]. However, one of the most common critiques of the South Korean system was the lack of privacy, security, and protection from fellow users/laypeople and protection of the business. Protection from identification is essential to avoid stigmatization of any individual or businesses, especially in the worldly context of economic strain due to the pandemic [70].

Introna and Poulodi define privacy as the protection from judgment from others, a highly relevant concept when considering the case of privacy breaches from contact tracing apps in South Korea [70-72]. In South Korea, businesses were threatened with false reports on site. Further, reporting of cases in or around the region of businesses resulted in economic strain, boycotting, and riots. This relates to Rowe's [9] first condition for a contact tracing system's success: correctness of the information on the app. This first point has two factors to consider: all diagnostic tests have some rate of false negatives and false positives, which may be reduced but not eliminated. Another factor is self- versus physician-reporting of positive cases. If users self-report, security issues arise, including the possibility of false reporting to stir anxiety and fear or even attack a specific location or business, as seen in South Korea. The latter point touches on the issue of privacy for businesses: by sharing the specific location of the infection, businesses around that area are now at risk of financial strain or even boycott. However, if physicians must report positive cases, the app would acquire less information, with those unable to go into a doctor's office or hospital unaccounted for. Further, whether a business is identified as a hotspot for infection due to blackmail, or it simply was the location of a positive contact point, the business may still be placed at an economic disadvantage. That information about the business is now being shared, potentially without their knowledge or consent, and they may be judged as a result. Methods of contact tracing which use GPS, QR codes, or any location-specific identifiers that are freely shared among users put businesses at risk.

However, businesses were not the only parties affected by South Korea's system. Personal information about individuals was discovered or speculated, leading to the stigmatization of those individuals. This result was especially damaging due to the fact that participation in South Korea's system was mandatory, rather than opt-in or opt-out. Therefore, those infected had no means of privacy from widespread stigmatization from their peers. Even without specific name sharing, individuals experienced online attacks, and events involving collective action against individuals online have been reported [73]. After an outbreak in an area associated with gay clubs, many individuals did not get tested or quarantined out of fear that they would out themselves and be judged by their community. It has been suggested that lottery-style randomized testing may protect against issues such as this [74].

Norway

Norway's Smittestopp app provides a key example of how rushed development of sensitive technologies can detrimentally backfire. The Smittestopp app used both GPS and Bluetooth, which was stored centrally on a government-controlled cloud platform [75]. The app collected data including mobile phone numbers, age, location data, and contact with infected individuals [75]. The app was said to collect anonymized movement data and would notify users of potential infected close contacts [76]. The app was not open source, which prevented community-based auditing [75]. The hasty development of the app also meant that it had several issues and was not particularly user-friendly, which discouraged its use [75]. Only 1.5 million people (out of 5.3 million) downloaded the app, which was not enough people for useful contact tracing [76]. As such, Norwegian Data Protection Authority deemed Smittestopp illegal because it collected too much personal information without providing a clear data usage policy to its users. The government was advised to shut down the app [76].

Norway's Smittestopp app was forced to undergo significant restructuring in mid-June by the government since the number of cases could not justify surveillance of the people [76]. Norway's rushed development of an app resulted in a less-than-optimal program, with significant security risks. Such an example highlights the need to be careful, thoughtful, and deliberate in the creation of such an app, especially when it comes to privacy concerns. Without doing so, privacy breaches may lead to a public that lacks trust in the app. After that point, it may be difficult, if not impossible, to acquire trust again. It is essential that as governments and organizations move forward, these apps are prioritized privacy and security prior to distribution to avoid wasted effort, resources, and time.

Despite its limitations and flaws, the app was at first successfully deployed because Norwegian culture places high trust in the government and novel technologies. What may be learned here is that full trust in novel technologies can lead to security and privacy breaches. However, a full distrust in novel technologies prevents the implementation of innovations that could be beneficial to local and global communities. Therefore, all of these digital contact tracing apps together (summarized in Table 8.1.2) show that a healthy skepticism level, as well as standardized and timely auditing, will improve future digital contact tracing technologies.

Table 8.1.2: Summary of key implementation decisions of each described contact tracing app

App name	Country/ company of origin	Centralized/ decentralized/ hybrid	GPS/ Bluetooth/ other	Google/ Apple API[a]?	Mandatory/ opt-in/ opt-out/ other
TraceTogether [33,55,56,58]	Singapore	Hybrid, centralized reporting with a confirmed case	Bluetooth	No	Opt-in
Google/Apple Exposure Notification API [45,50,59]	Apple/Google	Decentralized	Bluetooth	Yes	Opt-in
COVID Tracker [62-64]	Ireland	Decentralized	Bluetooth	Yes	Opt-in
Health Code [33,66-68]	China, Alipay/ WeChat	Centralized	GPS, symptom tracking; QR[b] code	No	Mandatory
Self-quarantine Safety Protection App [69-74]	South Korea	Centralized	GPS	No	Mandatory
Smittestopp [75,76]	Norway	Decentralized	Bluetooth	Yes	Opt-in

[a]API: application programming interface.

[b]QR: quick response.

Primary Limitations and Concerns in App Design

Privacy and Security

Privacy and Security Limitations and Concerns

A high proportion of the local and national population must use a contact tracing app for the app to be successful [9]. This notion is intimately tied with privacy, security, and autonomy. Public perception and acceptance play a key role in app downloads and usage for voluntary opt-in and opt-out systems. However, notions of privacy differ from country to country, and local perceptions will influence which implementation decisions are acceptable. All contact tracing will require some privacy loss, so each community must determine their optimal privacy-utility tradeoff. When public health and wellness are at odds with privacy, this tradeoff may be different than in times of stability. Several surveys have been conducted worldwide to understand opinions about digital contact tracing apps, and the primary cited reasons for not using apps were privacy, security, and surveillance. Given the large concern from users, privacy, security, surveillance, and transparency must be at the forefront of future and current contract tracing app designs.

In Ireland, of those surveyed who reported that they would not download the app, the most common reason was privacy [77]. They found concerns that the app host would use personal data for surveillance purposes, rather than for public health, and that surveillance would continue after the pandemic. Similarly, a survey study in Jordan found that 71.6% agree with the use of contact tracing apps, but only 37.8% used such technology. The main concerns among survey participants were privacy, voluntary status, and beneficence of the data [78].

In a recent large-scale, multicountry study with 5995 participants, the surveyors found general support for contact tracing apps [79]. However, they note that participants from the United States were generally less supportive of the app than other respondents. They found that this generally corresponded with a lack of trust in the national government. Similar to the Ireland study, the main reasons given for not downloading the app were concerns related to government surveillance (42%) and cybersecurity (35%). Interestingly, they found that 74.8% would definitely or probably download an opt-in app, but only 67.7% would probably or definitely keep an opt-out app. Therefore, concerns of privacy, security, and autonomy are primary concerns for users, especially in the United States.

In a Johns Hopkins study of US citizens, 82% reported that they would use a “perfectly accurate and private” tracing app, but

only 24% to 26% would want one with even “a low chance” of a data leak to the government, an employer, a tech company, or a nonprofit organization [80]. Given the benchmark that 80% of smartphone users should download a contact tracing app for it to be most successful, these results highlight the essential nature of a deliberately chosen system, which is private and full transparency by entities with an active role in its execution. Without a promise of privacy, it is far less likely that a contact tracing app will be adopted to the degree necessary for efficacy, making it significantly less useful in public health efforts.

Given that even a low chance of data leak is enough to dissuade over 50% of US users, these apps must be reliable and able to maintain a positive reputation. Even a single privacy leak event or controversy will result in a lack of credibility for the app. After such an event, the app would unlikely be viewed as “perfectly accurate and private,” preventing it from reaching the >80% margin necessary for maximal efficacy. Therefore, privacy concerns must be at the forefront of contact tracing app design. Contact tracing apps, especially in the United States, should protect users (both individuals and businesses) from fellow users, the authority hosting the app, and snoopers or hackers.

Privacy and Security Recommendations for App Development

In the time of a pandemic, certain rights may be loosened for public health and safety, for example, requirements such as mandated quarantines and mask wearing [81-84]. While disagreement has arisen with these measures, there is a general sense of compromise for the sake of public health [13,85]. When it comes to contact tracing mobile phone apps, a central question is what privacy-utility tradeoff fits within a local community’s values. Will people allow personal location information and contacts to be shared so that the pandemic could be slowed, lives could be saved, and quarantine restrictions could be relaxed sooner? If so, what are the limits of loss of privacy? Inherently, for these apps to function, some degree of location information must be shared. There is an ethical balance between reducing the havoc of the pandemic and saving lives and risking some loss of privacy [68]. Rowe [9] hypothesized that individuals might be comfortable with the risks of location sharing if that meant increased health and financial protection for themselves and their families. Ghose et al [13] found that Americans increased their rates of location sharing services during the pandemic, suggesting that people were comfortable reducing some levels of personal privacy for the sake of public health. Attitudes in South Korea show that the population is comfortable with privacy leakages for the sake of public health [68], whereas reports in American and many European countries favor privacy over public health interventions [79,80].

Nevertheless, in the design of these apps, the likelihood of an information leak should be minimized and not purposefully done. While individuals may sacrifice privacy for the public, unnecessary risks must be minimized. This requires purposeful security design, transparency about data use that should be exclusively for public health purposes, and a promise of nondiscrimination both at the hand of the app host and fellow users. To ensure these requirements, there must be a balance between minimizing information sharing while maximizing the usefulness of the app itself [18]. Yannakourou et al [86] outlined several guidelines. For example, the minimum amount of individual information must be gathered that still allows for efficacious operation, meaning the removal of outdated data that is older than the incubation period of approximately 2 weeks. Unnecessary additional information is deemed unethical since it serves no purpose for public health but is instead collected for the sake of data collection itself. Singapore’s reported 21-day limit follows this guideline. Furthermore, the app should cease use after its benefits are no longer needed (e.g., at the end of the pandemic). If these practices were implemented and advertised, this would also further garner trust and, therefore, increase the population’s usage.

While these concerns relate to privacy and transparency from the central host of the app or from fellow users, the app should also be safe from snoopers. For instance, only 16 of 50 apps reviewed prioritized data encryption and security via data anonymization and aggregate online reporting in one study [87]. This concern reinforces the importance of not rushing to deployment but rather carefully confirming the protection of the app from any type of information leakage. Open-source technology and community-based auditing have been suggested as one means of achieving thorough security [80]. Such transparency encourages trust, allows third parties to confirm the intentions of the app host, and allows the technology to be further improved, including cryptographic methods. The Massachusetts Institute of Technology (MIT) Private Kit: Safe Paths has been commended as an open-source, secure, and decentralized program [88]. The MIT implementation was designed with privacy as a priority, and their transparent, open-source model prevents abuse on the part of the app host. Many others have called for increased levels of transparency, whether that be clear and understandable statements regarding the use of data collection or open-source code available for audit [66].

Since privacy-related concerns were the main cited deterrent to app usage, privacy-by-design implementations are necessary for a successful contact tracing app. An app must be thoroughly vetted and trusted for its ability to maintain its users’ privacy and security. An opt-in system may be best in that case, especially in the United States, as trust in government is lower than in other nations [66]. Regardless, this may require trust in the government or in the authority hosting the app through transparency and decentralization. Open-source implementations allow for outside auditing and clear understanding between the developer and the user to increase transparency and accountability. This is a large concern from those who do not wish to download such apps

and feel that the government may be using their data for reasons outside of public health and slowing the virus. If the program is trustworthy by open-source practices and transparency and follows proper ethical guidelines, the app may garner more users.

Several other important ethical aspects must be considered when designing such apps, as has been emphasized by several reviews [15,16,18]. It has also been emphasized that mandatory downloading is not an appropriate solution, as not everyone has access to smartphone technology. Similarly, there are concerns about how contact tracing can be used to free people from quarantine in an equitable and fair manner, given that not everyone has access to this technology [15]. Others have brought up concerns about the use of such apps for surveillance and policing of marginalized populations [89]. There are also concerns that centralized servers may risk individual privacy from state surveillance and third-party data breaches. However, centralized information allows for future epidemiology research or public health service planning.

In sum, there are many ethical concerns surrounding contract tracing apps, many of which pertain to privacy, security, and surveillance. Given that a large number of people must actively use contact tracing apps for them to work, opt-in implementations must be trusted by their user base. There is no perfect formula for success in implementation, in large part because each community has its own values and will require a unique approach to the privacy-utility tradeoff. Local laws may also place constraints on what is possible. Surveys of local populations can be used as a springboard to app design. App designers must consider local political and cultural attitudes toward technology, privacy, and public health. Politics can determine regulations on app data collection, and so political norms may dictate the limitations of app functionality. Further, cultural attitudes are likely to influence whether local people are willing to download and use the app at all. Of course, political and cultural attitudes can influence one another. Politics are often a reflection of general cultural expectations. We have highlighted several key contact tracing apps designed specifically based on local political and cultural attitudes, such as the apps in Singapore, South Korea, and Norway. These app implementations may not have been successful or possible at all in countries other than their origin. As such, it is of utmost importance to have a good understanding of local attitudes and needs to ensure that the app is most effective. Future contact tracing apps in the United States will require high levels of privacy protection; open-source and transparent design; and a decentralized, opt-in system.

Nonprivacy Concerns Related to Apps

As previously discussed, high and consistent long-term usage is essential for contact tracing app efficacy. Although privacy is a key element limiting the extent of usage for this current generation of COVID-19 apps, several nonprivacy factors remain at play in driving users against consistent long-term usage.

App Efficacy and Perception Issues

Issues with app performance have resulted in significantly decreased usage. One of the most impactful examples is excessive battery usage due to contact tracing apps. In addition to privacy, Singaporeans have cited depletion of mobile phone battery life as one of the primary reasons against the TraceTogether app [17]. Thus, even in a nation with high levels of digital inclusion and public governance, less than a quarter of Singaporeans downloaded the app initially [55].

Additionally, in response to a software bug in the exposure notifications system developed by Google and Apple [90] caused by the recent update to the GAEN API [91], 83,000 users of Ireland's COVID Tracker app (out of 1.5 million users) deleted it from their mobile devices in an attempt to solve the underlying technical issue [92]. As a result of the Health Service Executive working with Google and NearForm (the Ireland-based company that designed the app) to solve this Android-specific problem, 10,000 users have already reinstalled the app. However, in a recent October 2020 survey, over half of respondents stated that their app's Bluetooth technology adversely affects device battery life [78]. These examples highlight the difficulty in regaining users once challenges are encountered and users are lost. Therefore, it is essential that extensive prerelease testing is performed to determine any underlying issues that could affect widespread usage.

While this technical debacle resulted in a loss of 73,000 app users, rapid responses to such issues may mitigate such nonprivacy factor effects on consistent long-term usage. Other than the bug affecting users of the COVID Tracker in Ireland, there have been no other significant reports of battery draining and overheating issues caused by a COVID-19 app. For example, nations such as Canada have seemingly experienced no battery drainage issues, even though they have also utilized the Apple-Google framework [93]. However, even if there were to be future issues, several organizations have resorted to building their own "localized" platforms with no dependence on the GAEN API [94]. For instance, Virginia's state government reports that they will forgo using the GAEN API for their system, but this may result in the need for significant workarounds for interoperating system communication that results in subpar performance [95]. The lack of transparency and comparability in app design may result in a reduced ability to resolve these issues and decrease app trust in the public.

Another possible factor in limiting the broad usage of COVID-19 apps is a lack of belief in such apps' effectiveness. One element of this distrust is evident in a survey by the Italian polling organization SWG [96]. In a survey asking 800 individuals, "Why did you choose not to download the Immuni app?," the most popular reason was "I do not consider it effective" (44%), after which was "I'm afraid for my privacy" (29%).

Finally, a significant factor against the broad usage of contact apps is an emerging sense of complacency. In a report by the Centers for Disease Control and Prevention, 41% of respondents have faced mental health challenges related to COVID-19 and steps taken to combat the pandemic, including social distancing and stay-at-home orders [97]. The literature has also revealed that stressors such as longer quarantine duration, fears of infection, frustration, boredom, inadequate supplies, inadequate information, financial loss, and stigma have contributed to negative psychological effects, including posttraumatic stress symptoms, confusion, and anger [98]. Downloading an app represents another self-sacrifice citizens have to make alongside activities like social distancing and mask wearing that are already mandated or strongly recommended. The long-term nature of the COVID-19 pandemic may lead to complacency when given the opportunity to make another COVID-19–related personal decision.

The literature has revealed that a decrease in concern, in addition to low political trust, can combine to undermine compliance with governmental restrictions during the pandemic [99]. This complacency can significantly adversely affect the adoption of contact tracing apps, even if the ruling government recommended them.

App Efficacy and Perception Recommendations

While issues like complacency may stem from the external issues dealing with pandemic fatigue, there are potential solutions for app efficacy and public perception. Bluetooth LE technology, the currently preferred technology for decentralized apps aimed at increasing user privacy, must be demonstrated to consistently and correctly measure user proximity. As of now, Bluetooth LE measurements are subject to discrepancies in signal attenuation based on the specific device used, the relative orientation of devices, and signal absorption and reflection by bodies, handbags, walls, etc [45]. Additional research must be conducted to increase the accuracy and efficacy of the employed technologies if contact tracing apps are to play an impactful role in infectious disease control moving forward. Toward this end, the National Institute of Standards and Technology, along with the MIT PACT (Private Automated Contact Tracing) project, has issued its "too close for too long" challenge to engage with research organizations worldwide concerning noise reduction and more precise distance and temporal estimation of Bluetooth LE signals [100].

Efforts to reduce unwanted false positives and false negatives have evolved in several directions. The Hamagen app used by Israel, for instance, allows users to see the location and time of potential exposure and indicate if they were not present [10]. Alternatively, data from contact tracing apps could be compared with manual tracing efforts to confirm or refute false positives and negatives, and health care providers could be called upon to report and authenticate data to reduce false negatives [10]. Additionally, big data could be used to filter out false positives in a centralized system.

For proximity-based strategies as well as location-based techniques, simulations and modeling may be used to guide design in terms of efficacy and community tailoring. For instance, Pandl et al [101] recently developed and implemented a spatial proximity simulation, which looked at the effects of both proximity detection range (0.2-10 meters) versus contact tracing app adoption (20%-100%), including simulations of decreased use with increase false-positive rates (25%-100% false positives). At higher adoption rates, longer proximity detection ranges (2 meters, 10 meters) were most effective at reducing disease spread but also resulted in increased false positives, which triggered decreased app utilization in highly reactive scenarios. The authors emphasized that this indicates the need to tailor app parameterization and strategies to match cultural expectations, indicating that less sensitive methods involving Bluetooth, GPS, and QR codes would be more acceptable in countries comfortable with a trade-off of high false positives for high efficacy.

In a recent review discussing digital technologies for contact tracing apps, Trivedi and Vasisht [11] discussed research to improve upon existing Bluetooth LE signals such as time-of-flight measurements using Bluetooth that allowed for accuracy to the foot and hybrid techniques such as those employing Bluetooth LE and acoustic-ranging in conjunction, although these have not been widely validated and are not ubiquitously deployable with today's smartphone technology. Alternatively, merging technical strategies may result in increased overall specificity, particularly for popular Bluetooth-based approaches. A recent paper from Nguyen et al [46] explored the idea of a multi–smartphone-sensor system for contact tracing using Bluetooth combined with barometer (effected by altitude and winds for indoor or outdoor environments), magnetometer (dynamic time warping used to interpret magnetic field vectors), microphone (high frequency, low amplitude short chirps emitted and time of flight measured), and WiFi data (reliant on a grid of WiFi hotspots). With added distance-based readings (microphone, WiFi), accuracy was increased from 25% to 65%, and the additional inclusion of environmental readings (barometer, magnetometer) further increased

accuracy to 87%. It should be noted, however, that the significant reduction in false positives was accompanied by a small increase in false negatives. However, these results indicate potential for increased efficacy by the synergistic leverage of multiple sensors for proximity-based digital contact tracing.

Finally, in order to increase public trust in contact tracing app efficacy, apps must undergo extensive testing prior to public release. Moreover, this testing must include rigorous evaluations of app efficacy in real-world environments, particularly those in which users are most likely to have encounters that are difficult to trace, such as public spaces. For example, Leith et al [48] have done testing of the GAEN framework in settings such as public transportation. The efficacy of these apps is not as simple as ensuring customer satisfaction, although this is important for maintaining consistent use, but has become a question of public health as such apps are embraced by national governments worldwide. As such, Bhatia et al [102] suggested that mobile health (mHealth) tools such as contact tracing apps may require regulatory intervention and need the introduction of regulatory sandboxes for extensive beta testing among diverse populations in diverse environments [102]. This strategy would ensure that efficacy data were available and rigorous before the release of a public health app in a time such as our current pandemic, which could both increase faith in the intervention and ensure that it is an appropriate allocation of resources. Appropriate and expedient evaluation of digital contact tracing apps will help determine if the benefit of this technology is worth the potential privacy and security risks previously discussed. Colizza et al [103] have called for such evaluation, and they suggest the use of surveys, epidemiological analysis, and experimental studies. At the time of writing, such analyses were minimal. With the lessons gathered from the use of contact tracing apps for COVID-19, better technology should be available for urgent and efficacious deployment in the event of another infectious disease outbreak. Because contact tracing apps call for some surrender of private data, it is necessary first to ensure that the technology used will be effective in the environments where it is most needed.

To our knowledge, the only evaluation of a digital contact tracing app was reported by Salathé et al [104] on the SwissCovid app in Switzerland, which uses the exposure network framework. They demonstrated a proof of principle that the app reached appropriate contacts who later tested positively for SARS-CoV-2. They provided evidence that notified users subsequently sought SARS-COV-2 testing. This suggests that the SwissCovid appropriately notifies people to self-isolate if they are at risk, which can limit transmission as a result. Nevertheless, there has been model-based analysis on contact tracing in general, which has found that contact tracing can reduce the effective reproduction number of a virus if the tracing is done with minimal delay, the tracing is accurate, and those informed of their potential risk follow appropriate isolation protocols [105]. This research suggests that mobile app technology can reduce the tracing delay, and thus, digital contact tracing apps may minimize viral transmission. Other models have suggested that smartphone-based contact tracing apps can become particularly effective after the first wave of an outbreak. The limitations to effectiveness, as discussed, include the rate of download and use, as well as the accuracy of the app [30]. Still, other than the analysis of the SwissCovid and these theoretical models, evaluations of specific contact tracing apps were not readily available at the time of writing.

Conclusion

In conclusion, there were many lessons learned from contact tracing apps that were designed to slow the spread of COVID-19. Many of these lessons relate to one core theme: for such an app to work, it is absolutely required that the app be used by a significant portion of the population. For this requirement to be satisfied, the app must offer three main capabilities to its user: (1) a level of trust in the efficacy of the app itself, which requires proper functionality and testing; (2) a level of security and privacy from the app host, fellow users, and malicious entities; and (3) minimal cost or effort from the user, whether that be in the form of battery usage, manual and frequent use of the app, or a host of other factors. Some of these requirements are obligatorily balanced. Many of the apps that proved unsuccessful so far did not sufficiently meet requirement 1 and potentially requirement 2. Therefore, many of these apps have been unable to gain and retain the number of users needed for accurate contact tracing, rendering them ineffective. It was proven repeatedly that these apps must be properly vetted and tested in multiple aspects prior to deployment, as technical performance issues have led to decreased public trust and usage. Likewise, any legal and ethical considerations concerning privacy must be adequately addressed before releasing an app for public use. To avoid the rushed time constraints from an emergency setting, proper development of such apps will need to happen prior to the emergency. In other words, similar to many other issues, we must be proactive rather than reactive when it comes to the use of contact tracing apps moving forward. Likewise, transparency from all actors involved in the development and management of contact tracing apps is necessary. The use of contact tracing apps during the COVID-19 pandemic will improve contact tracing apps in general by providing these real-world lessons.

Acknowledgments

This work was conducted in conjunction with the Technology Committee of the American Physician Scientists Association. KH acknowledges the Baylor College of Medicine Medical Scientist Training Program and support from the National Institute

of Dental and Craniofacial Research (F31 DE030333). XJ is a CPRIT Scholar in Cancer Research (RR180012), and he was supported in part by the Christopher Sarofim Family Professorship, UT Stars Award, UTHealth Startup, the National Institutes of Health (award number R01AG066749, R01AG066749-01S1, R41HG010978-01S1, and U01TR002062), and the National Science Foundation RAPID #2027790.

Abbreviations

AI: artificial intelligence

API: application programming interface

CCTV: closed-circuit television

CoV-SCR: COVID-19 Symptom Monitoring and Contact Tracking Record

GAEN: Google/Apple Exposure Network

GIS: global information system

mHealth: mobile health

MIT: Massachusetts Institute of Technology

PACT: Private Automated Contact Tracing

PEPP-PT: Pan-European Privacy-Preserving Proximity Tracing

QR: quick response

R0: basic reproduction number

RPI: Rolling Proximity Identifier

STRONG: Spatiotemporal Reporting Over Network and GPS

TTP: trusted third party

Footnotes

Conflicts of Interest: None declared.

References

1. *Wu Y, Chen C, Chan Y. The outbreak of COVID-19. Journal of the Chinese Medical Association. 2020;83(3):217–220.*
2. *Rossen LM, Branum AM, Ahmad FB, Sutton P, Anderson RN. Excess Deaths Associated with COVID-19, by Age and Race and Ethnicity - United States, January 26-October 3, 2020. MMWR Morb Mortal Wkly Rep. 2020 Oct 23;69(42):1522–1527. doi: 10.15585/mmwr.mm6942e2.*
3. *Ioannidis JPA. Global perspective of COVID-19 epidemiology for a full-cycle pandemic. Eur J Clin Invest. 2020 Dec 25;50(12):e13423. doi: 10.1111/eci.13423.*
4. *Browne C, Gulbudak H, Webb G. Modeling contact tracing in outbreaks with application to Ebola. J Theor Biol. 2015 Nov 07;384:33–49. doi: 10.1016/j.jtbi.2015.08.004.*
5. *Sacks JA, Zehe E, Redick C, Bah A, Cowger K, Camara M, Diallo A, Gigo ANI, Dhillon RS, Liu A. Introduction of Mobile Health Tools to Support Ebola Surveillance and Contact Tracing in Guinea. Glob Health Sci Pract. 2015 Nov 12;3(4):646–659. doi: 10.9745/ghsp-d-15-00207.*
6. *Danquah LO, Hasham N, MacFarlane M, Conteh FE, Momoh F, Tedesco AA, Jambai A, Ross DA, Weiss HA. Use of a mobile application for Ebola contact tracing and monitoring in northern Sierra Leone: a proof-of-concept study. BMC Infect Dis. 2019 Sep 18;19(1):810. doi: 10.1186/s12879-019-4354-z.*

7. Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dörner Lucie, Parker M, Bonsall D, Fraser C. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. Science. 2020 May 08;368(6491):eabb6936. doi: 10.1126/science.abb6936.
8. Yasaka TM, Lehrich BM, Sahyouni R. Peer-to-Peer Contact Tracing: Development of a Privacy-Preserving Smartphone App. JMIR Mhealth Uhealth. 2020 Apr 07;8(4):e18936. doi: 10.2196/18936.
9. Rowe F. Contact tracing apps and values dilemmas: A privacy paradox in a neo-liberal world. Int J Inf Manage. 2020 Dec;55:102178. doi: 10.1016/j.ijinfomgt.2020.102178.
10. Ahmed N, Michelin RA, Xue W, Ruj S, Malaney R, Kanhere SS, Seneviratne A, Hu W, Janicke H, Jha SK. A Survey of COVID-19 Contact Tracing Apps. IEEE Access. 2020;8:134577–134601. doi: 10.1109/access.2020.3010226.
11. Trivedi A, Vasisht D. Digital contact tracing. SIGCOMM Comput Commun Rev. 2020 Oct 26;50(4):75–81. doi: 10.1145/3431832.3431841.
12. Bengio Y, Ippolito D, Janda R, Jarvie M, Prud'homme Benjamin, Rousseau J, Sharma A, Yu Y. Inherent privacy limitations of decentralized contact tracing apps. J Am Med Inform Assoc. 2021 Jan 15;28(1):193–195. doi: 10.1093/jamia/ocaa153.
13. Ghose A, Li B, Macha M, Sun C, Foutz N. Trading Privacy for the Greater Social Good: How Did America React During COVID-19? arXiv. Preprint posted online Jun 10, 2020.
14. Bengio Y, Janda R, Yu YW, Ippolito D, Jarvie M, Pilat D, Struck B, Krastev S, Sharma A. The need for privacy with public digital contact tracing during the COVID-19 pandemic. The Lancet Digital Health. 2020 Jul;2(7):e342–e344. doi: 10.1016/s2589-7500(20)30133-3.
15. Klenk M, Duijf H. Ethics of digital contact tracing and COVID-19: who is (not) free to go? Ethics Inf Technol. 2020 Aug 24;:1–9. doi: 10.1007/s10676-020-09544-0.
16. Mello MM, Wang CJ. Ethics and governance for digital disease surveillance. Science. 2020 May 29;368(6494):951–954. doi: 10.1126/science.abb9045.
17. Pagliari C. The ethics and value of contact tracing apps: International insights and implications for Scotland's COVID-19 response. J Glob Health. 2020 Dec;10(2):020103. doi: 10.7189/jogh.10.020103. doi: 10.7189/jogh.10.020103.
18. Parker MJ, Fraser C, Abeler-Dörner Lucie, Bonsall D. Ethics of instantaneous contact tracing using mobile phone apps in the control of the COVID-19 pandemic. J Med Ethics. 2020 Jul 04;46(7):427–431. doi: 10.1136/medethics-2020-106314.
19. Yuan J, Li M, Lv G, Lu ZK. Monitoring transmissibility and mortality of COVID-19 in Europe. Int J Infect Dis. 2020 Jun;95:311–315. doi: 10.1016/j.ijid.2020.03.050.
20. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, Azman AS, Reich NG, Lessler J. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. Annals of Internal Medicine. 2020 May 05;172(9):577–582. doi: 10.7326/m20-0504.
21. He X, Lau EHY, Wu P, Deng X, Wang J, Hao X, Lau YC, Wong JY, Guan Y, Tan X, Mo X, Chen Y, Liao B, Chen W, Hu F, Zhang Q, Zhong M, Wu Y, Zhao L, Zhang F, Cowling BJ, Li F, Leung GM. Temporal dynamics in viral shedding and transmissibility of COVID-19. Nat Med. 2020 May 15;26(5):672–675. doi: 10.1038/s41591-020-0869-5.
22. Cheng H, Jian S, Liu D, Ng T, Huang W, Lin H, Taiwan COVID-19 Outbreak Investigation Team Contact Tracing Assessment of COVID-19 Transmission Dynamics in Taiwan and Risk at Different Exposure Periods Before and After Symptom Onset. JAMA Intern Med. 2020 Sep 01;180(9):1156–1163. doi: 10.1001/jamainternmed.2020.2020.
23. Rong X, Yang L, Chu H, Fan M. Effect of delay in diagnosis on transmission of COVID-19. Math Biosci Eng. 2020 Mar 11;17(3):2725–2740. doi: 10.3934/mbe.2020149.
24. Bradshaw WJ, Alley EC, Huggins JH, Lloyd AL, Esvelt KM. Bidirectional contact tracing could dramatically improve COVID-19 control. Nat Commun. 2021 Jan 11;12(1):232. doi: 10.1038/s41467-020-20325-7. doi: 10.1038/s41467-020-20325-7.
25. Kojaku S, Hébert-Dufresne L, Mones E, Lehmann S, Ahn Y. The effectiveness of backward contact tracing in networks. Nat Phys. 2021 Feb 25;17(5):652–658. doi: 10.1038/s41567-021-01187-2.
26. Lai SHS, Tang CQY, Kurup A, Thevendran G. The experience of contact tracing in Singapore in the control of COVID-19: highlighting the use of digital technology. Int Orthop. 2021 Jan 14;45(1):65–69. doi: 10.1007/s00264-020-04646-2.
27. Maccari L, Cagno V. Do we need a contact tracing app? Comput Commun. 2021 Jan 15;166:9–18. doi: 10.1016/j.comcom.2020.11.007.
28. Cencetti G, Santin G, Longa A, Pigani E, Barrat A, Cattuto C, Lehmann S, Salathé M, Lepri B. Digital proximity tracing on empirical contact networks for pandemic control. Nat Commun. 2021 Mar 12;12(1):1655. doi: 10.1038/s41467-021-21809-w. doi: 10.1038/s41467-021-21809-w.
29. Amann J, Sleigh J, Vayena E. Digital contact-tracing during the Covid-19 pandemic: An analysis of newspaper coverage in Germany, Austria, and Switzerland. PLoS One. 2021;16(2):e0246524. doi: 10.1371/journal.pone.0246524.
30. Hernandez-Orallo E, Manzoni P, Calafate CT, Cano J. Evaluating How Smartphone Contact Tracing Technology Can Reduce the Spread of Infectious Diseases: The Case of COVID-19. IEEE Access. 2020;8:99083–99097. doi: 10.1109/access.2020.2998042.
31. Shubina V, Holcer S, Gould M, Lohan ES. Survey of Decentralized Solutions with Mobile Devices for User Location Tracking, Proximity Detection, and Contact Tracing in the COVID-19 Era. Data. 2020 Sep 23;5(4):87. doi: 10.3390/data5040087.
32. Zhao Q, Wen H, Lin Z, Xuan D, Shroff N. On the Accuracy of Measured Proximity of Bluetooth-Based Contact Tracing Apps.

In: Park N, Sun K, Foresti S, Butler K, Saxena N, editors. Security and Privacy in Communication Networks. SecureComm 2020. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 335. Cham: Springer; 2020. pp. 49–60.

33. *Liang F. COVID-19 and Health Code: How Digital Platforms Tackle the Pandemic in China. Soc Media Soc. 2020 Jul 11;6(3) doi: 10.1177/2056305120947657.*
34. *Dar AB, Lone AH, Zahoor S, Khan AA, Naaz R. Applicability of mobile contact tracing in fighting pandemic (COVID-19): Issues, challenges and solutions. Comput Sci Rev. 2020 Nov;38:100307. doi: 10.1016/j.cosrev.2020.100307.*
35. *Gupta R, Rao UP. Achieving location privacy through CAST in location based services. J Commun Netw. 2017;19(3):239–249. doi: 10.1109/jcn.2017.000041.*
36. *Underwood S. Blockchain beyond bitcoin. Commun ACM. 2016 Oct 28;59(11):15–17. doi: 10.1145/2994581.*
37. *Xu H, Zhang L, Onireti O, Fang Y, Buchanan WJ, Imran MA. BeepTrace: Blockchain-Enabled Privacy-Preserving Contact Tracing for COVID-19 Pandemic and Beyond. IEEE Internet Things J. 2021 Mar 1;8(5):3915–3929. doi: 10.1109/jiot.2020.3025953.*
38. *Lazer D, Kennedy R, King G, Vespignani A. Big data. The parable of Google Flu: traps in big data analysis. Science. 2014 Mar 14;343(6176):1203–5. doi: 10.1126/science.1248506.*
39. *Kamel Boulos MN, Geraghty EM. Geographical tracking and mapping of coronavirus disease COVID-19/severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) epidemic and associated events around the world: how 21st century GIS technologies are supporting the global fight against outbreaks and epidemics. Int J Health Geogr. 2020 Mar 11;19(1):8. doi: 10.1186/s12942-020-00202-8.*
40. *Wang S, Ding S, Xiong L. A New System for Surveillance and Digital Contact Tracing for COVID-19: Spatiotemporal Reporting Over Network and GPS. JMIR Mhealth Uhealth. 2020 Jun 10;8(6):e19457. doi: 10.2196/19457.*
41. *Kleinman RA, Merkel C. Digital contact tracing for COVID-19. CMAJ. 2020 Jun 15;192(24):E653–E656. doi: 10.1503/cmaj.200922.*
42. *Abeler J, Bäcker Matthias, Buermeyer U, Zillessen H. COVID-19 Contact Tracing and Data Protection Can Go Together. JMIR Mhealth Uhealth. 2020 Apr 20;8(4):e19359. doi: 10.2196/19359.*
43. *Psiaki ML, Humphreys TE, Stauffer B. Attackers can spoof navigation signals without our knowledge. Here's how to fight back GPS lies. IEEE Spectr. 2016 Aug;53(8):26–53. doi: 10.1109/mspec.2016.7524168.*
44. *Alsdurf H, Belliveau E, Bengio Y, Deleu T, Gupta P, Ippolito D, Janda R, Jarvie M, Kolody T, Krastev S, Maharaj T, Obryk R, Pilat D, Pisano V, Prud?homme B, Qu M, Rahaman N, Rish I, Rousseau J, Sharma A, Struck B, Tang J, Weiss M, Yu Y. COVI White Paper. arXiv. Preprint posted online May 18, 2020.*
45. *Leith DJ, Farrell S. Measurement-based evaluation of Google/Apple Exposure Notification API for proximity detection in a light-rail tram. PLoS One. 2020 Sep 30;15(9):e0239943. doi: 10.1371/journal.pone.0239943.*
46. *Nguyen KA, Luo Z, Watkins C. Epidemic contact tracing with smartphone sensors. Journal of Location Based Services. 2020 Sep 01;14(2):92–128. doi: 10.1080/17489725.2020.1805521.*
47. *Braithwaite I, Callender T, Bullock M, Aldridge RW. Automated and partly automated contact tracing: a systematic review to inform the control of COVID-19. The Lancet Digital Health. 2020 Nov;2(11):e607–e621. doi: 10.1016/s2589-7500(20)30184-9.*
48. *Leith DJ, Farrell S. Coronavirus contact tracing. SIGCOMM Comput Commun Rev. 2020 Oct 26;50(4):66–74. doi: 10.1145/3431832.3431840.*
49. *Currie D, Peng C, Lyle D, Jameson B, Frommer M. Stemming the flow: how much can the Australian smartphone app help to control COVID-19? Public Health Res Pract. 2020 Jun 30;30(2) doi: 10.17061/phrp30022009. doi: 10.17061/phrp30022009.*
50. *Sharon T. Blind-sided by privacy? Digital contact tracing, the Apple/Google API and big tech's newfound role as global health policy makers. Ethics Inf Technol. 2020 Jul 18;:1–13. doi: 10.1007/s10676-020-09547-x.*
51. *Chowdhury MJM, Ferdous MS, Biswas K, Chowdhury N, Muthukkumarasamy V. COVID-19 Contact Tracing: Challenges and Future Directions. IEEE Access. 2020;8:225703–225729. doi: 10.1109/access.2020.3036718.*
52. *Girolami M, Mavilia F, Delmastro F, Distefano E. Detecting Social Interactions through Commercial Mobile Devices. 2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops); Mar 19-23; Athens, Greece. 2018. pp. 125–130.*
53. *Yamamoto K, Takahashi T, Urasaki M, Nagayasu Y, Shimamoto T, Tateyama Y, Matsuzaki K, Kobayashi D, Kubo S, Mito S, Abe T, Matsuura H, Iwami T. Health Observation App for COVID-19 Symptom Tracking Integrated With Personal Health Records: Proof of Concept and Practical Use Study. JMIR Mhealth Uhealth. 2020 Jul 06;8(7):e19902. doi: 10.2196/19902.*
54. *Yap KY, Xie Q. Personalizing symptom monitoring and contact tracing efforts through a COVID-19 web-app. Infect Dis Poverty. 2020 Jul 13;9(1):93. doi: 10.1186/s40249-020-00711-5.*
55. *Lee T, Lee H. Tracing surveillance and auto-regulation in Singapore: 'smart' responses to COVID-19. Media International Australia. 2020 Aug 12;177(1):47–60. doi: 10.1177/1329878x20949545.*
56. *Bay J, Kek J, Tan A, Hau C, Yongquan L, Tan J, Quy T. BlueTrace: A privacy-preserving protocol for community-driven contact tracing across borders. Government Technology Agency, Singapore. 2020. Apr 9, [2021-07-14].*
57. *Abbas R, Michael K. COVID-19 Contact Trace App Deployments: Learnings From Australia and Singapore. IEEE Consumer*

Electron Mag. 2020 Sep 1;9(5):65–70. doi: 10.1109/mce.2020.3002490.

58. *Goggin G. COVID-19 apps in Singapore and Australia: reimagining healthy nations with digital technology. Media International Australia. 2020 Aug 14;177(1):61–75. doi: 10.1177/1329878x20949770.*
59. *Michael K, Abbas R. Behind COVID-19 Contact Trace Apps: The Google–Apple Partnership. IEEE Consumer Electron Mag. 2020 Sep 1;9(5):71–76. doi: 10.1109/mce.2020.3002492.*
60. *Loi M. How to fairly incentivise digital contact tracing. J Med Ethics. 2020 Jul 09;:medethics-2020-106388. doi: 10.1136/medethics-2020-106388.*
61. *Chua AQ, Tan MMJ, Verma M, Han EKL, Hsu LY, Cook AR, Teo YY, Lee VJ, Legido-Quigley H. Health system resilience in managing the COVID-19 pandemic: lessons from Singapore. BMJ Glob Health. 2020 Sep 16;5(9):e003317. doi: 10.1136/bmjgh-2020-003317.*
62. *Leslie M. COVID-19 Fight Enlists Digital Technology: Contact Tracing Apps. Engineering (Beijing) 2020 Oct;6(10):1064–1066. doi: 10.1016/j.eng.2020.09.001.*
63. *Skoll D, Miller JC, Saxon LA. COVID-19 testing and infection surveillance: Is a combined digital contact-tracing and mass-testing solution feasible in the United States? Cardiovascular Digital Health Journal. 2020 Nov;1(3):149–159. doi: 10.1016/j.cvdhj.2020.09.004.*
64. *HSE Covid Tracker App: Pre-Release Report Card. Irish Council for Civil Liberties, Digital Rights Ireland. 2020. Jul, [2021-07-12].*
65. *Shankar K, Jeng W, Thomer A, Weber N, Yoon A. Data curation as collective action during COVID-19. J Assoc Inf Sci Technol. 2020 Sep 02;72(3):280–284. doi: 10.1002/asi.24406.*
66. *66.Lucivero F, Hallowell N, Johnson S, Prainsack B, Samuel G, Sharon T. COVID-19 and Contact Tracing Apps: Ethical Challenges for a Social Experiment on a Global Scale. J Bioeth Inq. 2020 Dec 25;17(4):835–839. doi: 10.1007/s11673-020-10016-9.*
67. *Joo J, Shin MM. Resolving the tension between full utilization of contact tracing app services and user stress as an effort to control the COVID-19 pandemic. Serv Bus. 2020 Sep 01;14(4):461–478. doi: 10.1007/s11628-020-00424-7.*
68. *Roche S. Smile, you're being traced! Some thoughts about the ethical issues of digital contact tracing applications. Journal of Location Based Services. 2020 Aug 24;14(2):71–91. doi: 10.1080/17489725.2020.1811409.*
69. *Ryan M. In defence of digital contact-tracing: human rights, South Korea and Covid-19. IJPCC. 2020 Aug 06;16(4):383–407. doi: 10.1108/ijpcc-07-2020-0081.*
70. *Park YJ, Choe YJ, Park O, Park SY, Kim Y, Kim J, Kweon S, Woo Y, Gwack J, Kim SS, Lee J, Hyun J, Ryu B, Jang YS, Kim H, Shin SH, Yi S, Lee S, Kim HK, Lee H, Jin Y, Park E, Choi SW, Kim M, Song J, Choi SW, Kim D, Jeon B, Yoo H, Jeong EK, COVID-19 National Emergency Response Center, EpidemiologyCase Management Team Contact Tracing during Coronavirus Disease Outbreak, South Korea, 2020. Emerg Infect Dis. 2020 Oct;26(10):2465–2468. doi: 10.3201/eid2610.201315. doi: 10.3201/eid2610.201315.*
71. *Contact Transmission of COVID-19 in South Korea: Novel Investigation Techniques for Tracing Contacts. Osong Public Health Res Perspect. 2020 Feb;11(1):60–63. doi: 10.24171/j.phrp.2020.11.1.09. doi: 10.24171/j.phrp.2020.11.1.09.*
72. *Introna L, Pouloudi A. Privacy in the information age: Stakeholders, interests and values. J Bus Ethics. 1999;22:27–38. doi: 10.1023/A:1006151900807.*
73. *Park S, Choi GJ, Ko H. Information Technology-Based Tracing Strategy in Response to COVID-19 in South Korea-Privacy Controversies. JAMA. 2020 Jun 02;323(21):2129–2130. doi: 10.1001/jama.2020.6602.*
74. *Luciano F. Mind the App-Considerations on the Ethical Risks of COVID-19 Apps. Philos Technol. 2020 Jun 13;33(2):1–6. doi: 10.1007/s13347-020-00408-5.*
75. *Sandvik KB. "Smittestopp": If you want your freedom back, download now. Big Data & Society. 2020 Jul 28;7(2):205395172093998. doi: 10.1177/2053951720939985.*
76. *Ursin G, Skjesol I, Tritter J. The COVID-19 pandemic in Norway: The dominance of social implications in framing the policy response. Health Policy Technol. 2020 Dec;9(4):663–672. doi: 10.1016/j.hlpt.2020.08.004.*
77. *O'Callaghan M, Buckley J, Fitzgerald B, Johnson K, Laffey J, McNicholas B, Nuseibeh B, O'Keeffe D, O'Keeffe I, Razzaq A, Rekaner K, Richardson I, Simpkin A, Storni C, Tsvyatkova D, Walsh J, Welsh T, Glynn L. A National Survey of Attitudes to COVID-19 Digital Contact Tracing in the Republic of Ireland Internet. Research Square. doi: 10.21203/rs.3.rs-40778/v1. Preprint posted online July 10, 2020.*
78. *Abuhammad S, Khabour OF, Alzoubi KH. COVID-19 Contact-Tracing Technology: Acceptability and Ethical Issues of Use. Patient Prefer Adherence. 2020 Sep;14:1639–1647. doi: 10.2147/PPA.S276183. doi: 10.2147/PPA.S276183.*
79. *Altmann S, Milsom L, Zillessen H, Blasone R, Gerdon F, Bach R, Kreuter F, Nosenzo D, Toussaert S, Abeler J. Acceptability of App-Based Contact Tracing for COVID-19: Cross-Country Survey Study. JMIR Mhealth Uhealth. 2020 Aug 28;8(8):e19857. doi: 10.2196/19857.*
80. *The app credibility gap. Nat Biotechnol. 2020 Jul 26;38(7):768–768. doi: 10.1038/s41587-020-0610-4.*
81. *Cheng VC, Wong S, Chuang VW, So SY, Chen JH, Sridhar S, To KK, Chan JF, Hung IF, Ho P, Yuen K. The role of community-wide wearing of face mask for control of coronavirus disease 2019 (COVID-19) epidemic due to SARS-CoV-2. J Infect. 2020 Jul;81(1):107–114. doi: 10.1016/j.jinf.2020.04.024.*

32. *Yeung N, Lai J, Luo J. Face Off: Polarized Public Opinions on Personal Face Mask Usage during the COVID-19 Pandemic. arXiv. doi: 10.1109/bigdata50022.2020.9378114. Preprint posted online Oct 31, 2020,*
33. *Parmet WE, Sinha MS. Covid-19 — The Law and Limits of Quarantine. N Engl J Med. 2020 Apr 09;382(15):e28. doi: 10.1056/nejmp2004211.*
34. *Haischer MH, Beilfuss R, Hart MR, Opielinski L, Wrucke D, Zirgaitis G, Uhrich TD, Hunter SK. Who is wearing a mask? Gender-, age-, and location-related differences during the COVID-19 pandemic. PLoS One. 2020 Oct 15;15(10):e0240785. doi: 10.1371/journal.pone.0240785.*
35. *Feng S, Shen C, Xia N, Song W, Fan M, Cowling BJ. Rational use of face masks in the COVID-19 pandemic. The Lancet Respiratory Medicine. 2020 May;8(5):434–436. doi: 10.1016/s2213-2600(20)30134-x.*
36. *Yannakourou S, Jougleux P, Markou C, Tziazas S, Gregoriou G, Flouri A, Koutsonikola I, Efthymiades K, Panagiotou G. Covid-19 and Law: Legal Solutions during Covid-19 Upheavals. ENTHA. 2020;13:9–29.*
37. *Sharma T, Bashir M. Use of apps in the COVID-19 response and the loss of privacy protection. Nat Med. 2020 Aug 26;26(8):1165–1167. doi: 10.1038/s41591-020-0928-y.*
38. *Rosenkrantz L, Schuurman N, Bell N, Amram O. The need for GIScience in mapping COVID-19. Health Place. 2021 Jan;67:102389. doi: 10.1016/j.healthplace.2020.102389.*
39. *Hendl T, Chung R, Wild V. Pandemic Surveillance and Racialized Subpopulations: Mitigating Vulnerabilities in COVID-19 Apps. J Bioeth Inq. 2020 Dec 25;17(4):829–834. doi: 10.1007/s11673-020-10034-7.*
90. *OBrien C. Google to roll out software fix to remedy issue with Covid app. Irish Times. 2020. Aug 9, [2021-07-12].*
91. *Gorey C. Google to launch fix for battery drain affecting Covid Tracker Ireland app. Silicon Republic. 2020. Aug 10, [2021-07-12].*
92. *O'Brien C. HSE says fix for Covid Tracker app rolled out to all Android users. Irish Times. 2020. Aug 10, [2021-07-12].*
93. *Gagné C. Everything You Need to Know About the New COVID Alert App. Chatelaine. 2020. Aug 10, [2021-07-12].*
94. *Leith DJ, Farrell S. Measurement-based evaluation of Google/Apple Exposure Notification API for proximity detection in a commuter bus. PLoS One. 2021 Apr 29;16(4):e0250826. doi: 10.1371/journal.pone.0250826.*
95. *Morrison S. Perhaps months too late, the Apple-Google Covid-19 contact tracing tool comes to America. Vox. 2020. Aug 6, [2021-07-12].*
96. *Immuni: le motivazioni di chi non scarica l'app. Punto Informatico. 2020. Jul 9, [2021-07-14].*
97. *Lardieri A. Coronavirus Pandemic Causing Anxiety, Depression in Americans, CDC Finds. US News. 2020. Aug 13, [2021-07-12].*
98. *Brooks SK, Webster RK, Smith LE, Woodland L, Wessely S, Greenberg N, Rubin GJ. The psychological impact of quarantine and how to reduce it: rapid review of the evidence. The Lancet. 2020 Mar 14;395(10227):912–920. doi: 10.1016/S0140-6736(20)30460-8.*
99. *Lalot F, Heering MS, Rullo M, Travaglino GA, Abrams D. The dangers of distrustful complacency: Low concern and low political trust combine to undermine compliance with governmental restrictions in the emerging Covid-19 pandemic. Group Processes & Intergroup Relations. 2020 Oct 30; doi: 10.1177/1368430220967986.*
00. *Sadjadi S, Greenberg C, Reynolds D. NIST TC4TL Challenge. NIST. 2020. Jun, [2021-07-12].*
01. *Pandl K, Thiebes S, Schmidt-Kraepelin M, Sunyaev A. How detection ranges and usage stops impact digital contact tracing effectiveness for COVID-19 Internet. medRxiv. doi: 10.1101/2020.12.08.20246140. Preprint posted online Dec 11, 2020.*
02. *Bhatia A, Matthan R, Khanna T, Balsari S. Regulatory Sandboxes: A Cure for mHealth Pilotitis? J Med Internet Res. 2020 Sep 15;22(9):e21276. doi: 10.2196/21276.*
03. *Colizza V, Grill E, Mikolajczyk R, Cattuto C, Kucharski A, Riley S, Kendall M, Lythgoe K, Bonsall D, Wymant C, Abeler-Dörner Lucie, Ferretti L, Fraser C. Time to evaluate COVID-19 contact-tracing apps. Nat Med. 2021 Mar 15;27(3):361–362. doi: 10.1038/s41591-021-01236-6.*
04. *Salathé Marcel, Althaus C, Anderegg N, Antonioli D, Ballouz T, Bugnon E, Čapkun Srdjan, Jackson D, Kim S, Larus J, Low N, Lueks W, Menges D, Moullet C, Payer M, Riou J, Stadler T, Troncoso C, Vayena E, von Wyl V. Early evidence of effectiveness of digital contact tracing for SARS-CoV-2 in Switzerland. Swiss Med Wkly. 2020 Dec 14;150:w20457. doi: 10.4414/smw.2020.20457.*
05. *Kretzschmar ME, Rozhnova G, Bootsma MCJ, van Boven M, van de Wijgert JHHM, Bonten MJM. Impact of delays on effectiveness of contact tracing strategies for COVID-19: a modelling study. The Lancet Public Health. 2020 Aug;5(8):e452–e459. doi: 10.1016/s2468-2667(20)30157-2.*

Attributions

Contact Tracing Apps: Lessons Learned on Privacy, Autonomy, and the Need for Detailed and Thoughtful Implementation - PMC

"Contact Tracing Apps: Lessons Learned on Privacy, Autonomy, and the Need for Detailed and Thoughtful Implementation"

Katie Hogan [1,#], Briana Macedo [2,#], Venkata Macha [3], Arko Barman [4,5], Xiaoqian Jiang [6,✉]

Editor: Christian Lovis

Reviewed by: Konstantin Pandl, David Wong

- **Author information**
- **Article notes**
- **Copyright and License information**

©Katie Hogan, Briana Macedo, Venkata Macha, Arko Barman, Xiaoqian Jiang. Originally published in JMIR Medical Informatics, 19.07.2021.

This is an open-access article distributed under the terms of the Creative Commons Attribution License 4.0, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Medical Informatics, is properly cited. The complete bibliographic information, a link to the original publication, as well as this copyright and license information must be included.

PMC Copyright Notice.

PMCID: PMC8291141 PMID: 34254937

Link to the original publication:

JMIR Medical Informatics - Contact Tracing Apps: Lessons Learned on Privacy, Autonomy, and the Need for Detailed and Thoughtful Implementation

Published on 19.Jul.2021 in Vol 9, No 7 (2021): July

JMIR Med Inform 2021;9(7):e27449

DOI link

Note

3/8/2026: Brought in a CC BY 4.0 paper. Provided proper attribution as required at the end of the paper under "Attributions."

Edits:

- Removed bullets from references
- Fixed "eg" to "e.g."
- Added short summary to appear in table of contents.
- Edited the Reference list per the Accessibility fix list instructions from the Accessibility Consultant using Option A.
- Updated Tables/Figures for accessibility fixes. Used Libretexts numbering templates and changes references to Tables/Figures in the text.
- Removed use of headings in the Abbreviations (accessibility item).

This page titled 8.1: Description of the problem and potential solutions is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by LibreTexts.

8.2: Project Management Considerations

Project Management Triangle at Work

Given the severity of the COVID-19 pandemic, time was of the essence to slow the spread of infections. Government health organizations pushed to accelerate the development of vaccines and therapeutics. In some countries and states, people were not allowed to physically come to work unless they were essential workers and thus needed to work from home if that option existed. Some businesses and people suffered major economic impact. Nonetheless, people still needed to be out in public for some purposes such as purchasing food, going to medical appointments, and so forth. As a result, having access to a COVID-19 contact tracing app could be beneficial.

In approaching a project to develop a COVID-19 contact tracing app, the main pressure point likely was time. The apps were generally built on behalf of government health organizations/public health authorities, so cost was not going to be a major factor in an emergency situation. The likely trade-offs would be against scope.

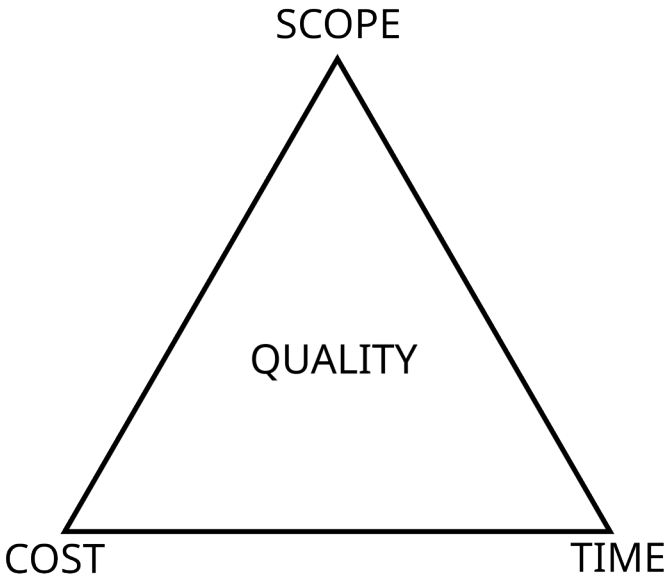


Figure 8.2.1: Project Management Triangle. (Public domain)

Accessibility Note

Figure 8.2.1: Project Management Triangle. The project management triangle represents tradeoffs between Cost, Time, and Scope (shown at the vertices of a triangle) with the word "Quality" inside the triangle.

Choose a Framework

Since time was of the essence, leveraging COVID-19 contact tracing app development frameworks, such as the Google/Apple Exposure Notification (GAEN) protocol, could save time. Making use of this framework would provide good coverage of the large number of Apple iOS and Google Android mobile devices in use worldwide. It is a decentralized approach that makes use of Bluetooth Low Energy for detecting close proximity contacts and encryption to help allay privacy concerns. It had a phased rollout:

1. Publication of the API specification
2. Rollout of tools to enable only governments (public health organizations) to create official privacy-preserving COVID-19 tracing apps
3. Integrating the functionality into the iOS and Android operating systems and distributing it via operating system updates.

After the pandemic ends, the functionality could be automatically removed.

The choice of a framework is going to affect whether the approach is centralized or decentralized.

List of frameworks[1]

Contact tracing application development frameworks

Name	Centralized / Decentralized	Author/promoter	License
ViraTrace Biohazard Containment Tools	Partially-Centralized w/ SGX backend	ViraTrace, LLC (Wayne Thornton, Andrei Taranu, Ivan Bestvina, Anjana Pai)	Restricted Public Source License (ViraTrace Public Source License 1.0.1)
Pan-European Privacy-Preserving Proximity Tracing (PEPP-PT) project	partially-centralized	Fraunhofer Institute for Telecommunications, Robert Koch Institute, Technische Universität Berlin, TU Dresden, University of Erfurt, Vodafone Germany	multiple protocols, closed source, private specifications, MPL
Coalition Network		Coalition Foundation, Nodle, French Institute for Research in Computer Science and Automation, Berkeley, California	GPL 3
Exposure Notifications System	decentralized	Google, Apple Inc.	public specification
Decentralized Privacy-Preserving Proximity Tracing (DP-3T)	decentralized	EPFL, ETHZ, KU Leuven, TU Delft, University College London, CISPA, University of Oxford, University of Torino / ISI Foundation	publicly developed Apache 2.0 reference implementation, MPL 2.0 iOS/Android code.
BlueTrace / OpenTrace	partially-centralized	Singapore Government Digital Services	public specification, GPL 3 code
Privacy-Sensitive Protocols And Mechanisms for Mobile Contact Tracing (PACT) / CovidSafe		Microsoft volunteers, University of Washington	public specification, MIT License code
PACT: Private Automated Contact Tracing		MIT Computer Science and Artificial Intelligence Laboratory, Massachusetts General Hospital, MIT Lincoln Laboratory, MIT Media Lab, Boston University, Weizmann Institute of Science, Brown University	public specification, MIT License code
Covid Watch / TCN Coalition / TCN Protocol	decentralized	Covid Watch, CoEpi, ITO, Commons Project, Zcash Foundation, Openmined, Coalition Network	public developed specification, MIT License code

Name	Centralized / Decentralized	Author/promoter	License
OpenCovidTrace	decentralized	Nebula Ventures, open source community,[124] Quantstellation, MLM Holdings, Evocativideas, 1Checkin	public developed specification, LGPL iOS/ Android code.
ReCoVer	Centralized, Hybrid	Stefano Piotto, Luigi Di Biasi,[126] Softmining, Minervas, PushApp, NexusTlc, University of Salerno, dibiasi.it Materiale Electrico,[127] BiTS Community beta volunteers	multiple protocols, partially closed source, partially private specifications, MPL

Scope

The following are some of the items that need to be addressed in finalizing project scope (assuming for the United States):

- Platforms to be supported: Apple iOS, Google Android devices
- Centralized versus decentralized architecture
- Technology choice: Bluetooth, GPS, cell phone location information (or hybrid approaches)
- Framework choice
- User interface
  - Obtain user consent
  - Enable/disable the app for when the user needs privacy
  - User provides notification of symptoms or of a positive COVID-19 test using a test kit
  - App provides notification alert to those exposed, with instructions about getting tested (getting a test kit or directions to a nearby test center) or seeking medical help.
  - Optionally, the app could provide information about hot spots of COVID-19 activity so the user could plan to avoid those locations if possible.
- Coverage of privacy and security considerations including data encryption
- Testing the app
  - Apple-device to Apple device; Android device to Android device, Apple device to Android device
  - Testing with phones in various positions (handheld, in purses, in backpacks, brief cases, etc.)
  - Test user upload of positive test and correct notification of contacts

Implementation in California

Since De Anza College is in California, let's look at what California did.

Since Google and Apple are located in California, an obvious choice would be to make use of the Google/Apple Exposure Notification System (GAEN) framework[2] GAEN requires that Exposure Notification apps be released only by public health authorities.[2] To that end, the operative authority in California is the California Department of Public Health.

The California Department of Public Health released the CA Notify system on December 10, 2020.[2] The CA Notify system used the GAEN framework. CA Notify uses Bluetooth Low-Energy for encounter logging to measure how close you are to other smartphones and for what duration. It did this by sending tracking messages to nearby devices containing unique identifiers that were encrypted. The smartphones saved these messages for 14 days.

If a user tests positive for COVID-19, that user provides the information about that positive COVID-19 test to CA Notify. The last 14 days of the infected user's encrypted keys were uploaded to a central server.

CA Notify, without sharing any information about the user who tested positive or the location, sends a user of the app a message

if the user might have been exposed to COVID-19 because they were in close proximity for a long enough period of time with information and resources about testing, how to keep others safe, and what to expect if you are sick.[3]

As of May 11, 2023, the COVID-19 digital exposure notification systems across the U.S. were turned off, on the same date that the nation's COVID-19 State of Emergency ended. Those who had CA Notify active on their smartphone received an alert informing them that exposure notifications are turned off, with the assurance to users that their privacy would be protected, since names and locations were never collected or stored.[4]

References

1. Wikipedia - COVID-19 apps. Accessed 3/8/2026. Text is available under the Creative Commons Attribution-ShareAlike 4.0 License(opens in new window); additional terms may apply. By using this site, you agree to the Terms of Use(opens in new window) and Privacy Policy(opens in new window). Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.
2. Wikipedia - Exposure Notification. Accessed May 4, 2026.
3. CA Notify. Accessed May 4, 2026.
4. CA Notify is Disabled as of May 11, 2023. CA Notify Disabled. Accessed May 4, 2026.
5. Apple and Google. Exposure Notification Cryptography Specification, Preliminary — Subject to Modification and Extension, April 2020, v1.2. Accessed May 5, 2026.

This page titled 8.2: Project Management Considerations is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Robert Slate and Daniel Slate.

8.3: Legal Considerations

Privacy Concerns

As we noted in the project management considerations, there was enormous pressure to deploy contact tracing apps as quickly as possible. Based on initial impressions and general cultural considerations, there were concerns by members of the public about privacy and security in contact tracing apps. These concerns affected app adoption rates by the public at large, and therefore affected the overall effectiveness of the app. There are lessons to be learned here that may affect how government public health agencies proactively prepare contact tracing apps with requisite scope to cover privacy and security concerns with the benefit of time for future major pandemics.

In some countries, it was culturally acceptable for the government to mandate usage of the country's contact tracing app. In most countries, usage of the country's contact tracing app was voluntary. In the case when usage of the country's contact tracing app was voluntary, it is interesting to look at the privacy-utility tradeoff that the populace is confronting. In normal times people would be unwilling to grant consent to have their data used for this purpose, but in a global pandemic, people might approach their decision analysis and risk assessment differently and thus might be more willing to take the privacy risk in order to seek therapeutic treatment as soon as possible, especially if they are in a high health risk group (e.g., older, immunocompromised). People have different risk profiles (different utility functions) and thus approach the decision differently. Public health agencies could provide guidance to people in various health risk categories to help influence the adoption rate.

In addition to the health risk tradeoff assessment, there is also the element of impact on personal freedom of moving around. Some government health agencies made relatively quick decisions to impose lock-downs which greatly impacted personal freedom of movement, employment, and created other economic impacts. The availability of a trusted contact tracing app that is widely deployed might influence public health authorities to impose less severe measures in the future.

The contact tracing apps need to limit the amount of personally identifiable information that is collected. Encrypted or anonymous identifiers for people could be transmitted using Bluetooth LE technology.

As with other medical information, users are concerned that their diagnosed health status, if it became widely known, could impact their employment or health insurance status.

Similarly, a place of business could experience severe economic impact if the business is tagged as a disease hot spot that members of the public stop patronizing for fear of contracting the disease.

Ethical Concerns[3]

COVID-19 contact tracing apps have led to concerns around privacy, the rights of app users, and governmental authority. Commentators have looked to the European Convention on Human Rights, the International Covenant on Civil and Political Rights (ICCPR), and the United Nations Siracusa Principles to outline four principles to use when considering the ethics of the mass surveillance of a country's populace involved in using COVID-19 contact tracing apps:

1. necessity,
2. proportionality,
3. scientific validity, and
4. time boundedness.[3]

Necessity is the idea that governments should only interfere with a person's rights when deemed essential for public health interests.The potential risks associated with infringements of personal privacy must be outweighed by the possibility of reducing significant harm to others. Potential benefits of contact-tracing apps that may be considered include allowing for blanket population-level quarantine measures to be lifted sooner and the minimization of people under quarantine. Hence, some proponents justify contact-tracing apps as they may be less intrusive than blanket quarantine measures.[3]

Proportionality refers to the concept that a contact tracing app's potential negative impact on a person's rights should be justifiable by the severity of the health risks that are being addressed. Apps must use the most privacy-preserving options available to achieve their goals, and the selected option should not only be a logical option for achieving the goal but also an

effective one.[3]

Scientific validity evaluates whether an app is effective, timely and accurate. Traditional manual contact-tracing procedures are not efficient enough for the COVID-19 pandemic, and do not consider asymptomatic transmission. Contact-tracing apps, on the other hand, can be effective COVID-19 contact-tracing tools that reduce R value [basic reproduction number] to less than 1, leading to sustained epidemic suppression.[3]

Definition: Basic Reproduction Number

In epidemiology, the **basic reproduction number**, or **basic reproductive number** (sometimes called **basic reproduction ratio** or **basic reproductive rate**), denoted R_0 (pronounced *R nought* or *R zero*), of an infection is the expected number of cases directly generated by one case in a population where all individuals are susceptible to infection.[4]

However, for contact-tracing apps to be effective, there needs to be a minimum 56-60% uptake (adoption of the app) in the population. Apps should be continually modified to reflect current knowledge on the diseases being monitored.[3]

Time boundedness describes the need for establishing legal and technical sunset clauses so that contact-tracing apps are only allowed to operate as long as necessary to address the pandemic situation. Contact-tracing apps should be withdrawn as soon as possible after the end of a pandemic. If the end of the pandemic cannot be predicted, the use of apps should be regularly reviewed and decisions about continued use should be made at each review. Collected data should only be retained by public health authorities for research purposes with clear stipulations on how long the data will be held for and who will be responsible for security, oversight, and ownership.[3] As previously noted, the contact tracing apps were in fact withdrawn after the state of emergency was removed. As of May 11, 2023, the COVID-19 digital exposure notification systems across the U.S. were turned off, on the same date that the nation's COVID-19 state of emergency ended.

GDPR

Since we have covered the EU GDPR in detail in this textbook, we can now briefly assess how COVID-19 contact tracing apps fit with the GDPR. Let's assume that the app is making use of the GAEN framework as discussed in Project Management considerations and that the app is meant to be deployed within the EU.

GDPR Article 4 likely covers the unique encrypted identifiers of the GAEN framework.

Let's refresh our minds with a subset of Article 4's definitions.

Definition: Personal Data

‘personal data’ means any information relating to an identified or identifiable natural person (‘data subject’); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.[5]

Definition: Processing

‘processing’ means any operation or set of operations which is performed on personal data or on sets of personal data, whether or not by automated means, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction.[5]

We previously did not discuss the specifics of data concerning health, but let us do so now since this case study deals with COVID-19 health data.

Definition: Data concerning health

‘data concerning health’ means personal data related to the physical or mental health of a natural person, including the provision of health care services, which reveal information about his or her health status.[5]

GDPR Article 9 "Processing of special categories of personal data" is applicable, with relevant clauses extracted below:

1. **Processing of personal data** revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, **data concerning health** or data concerning a natural person's sex life or sexual orientation **shall be prohibited**.
2. (a) **Paragraph 1 shall not apply if one of the following applies**:

[...]

(g) processing is necessary for reasons of substantial public interest, on the basis of Union or Member State law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject;

(h) processing is necessary for the purposes of preventive or occupational medicine, for the assessment of the working capacity of the employee, medical diagnosis, the provision of health or social care or treatment or the management of health or social care systems and services on the basis of Union or Member State law or pursuant to contract with a health professional and subject to the conditions and safeguards referred to in paragraph 3;

(i) processing is necessary for reasons of public interest in the area of public health, such as protecting against serious cross-border threats to health or ensuring high standards of quality and safety of health care and of medicinal products or medical devices, on the basis of Union or Member State law which provides for suitable and specific measures to safeguard the rights and freedoms of the data subject, in particular professional secrecy.[5]

Not discussed previously in the GDPR case study, but which is relevant here, is GDPR Recital 46. Note the added highlighting in bold:

> "The processing of personal data should also be regarded to be lawful where it is necessary to protect an interest which is essential for the life of the data subject or that of another natural person. Processing of personal data based on the **vital interest of another natural person** should in principle take place only where the processing cannot be manifestly based on another legal basis. Some types of processing may serve both important grounds of public interest and the vital interests of the data subject as for instance **when processing is necessary for humanitarian purposes, including for monitoring epidemics and their spread** or in situations of humanitarian emergencies, in particular in situations of natural and man-made disasters."[5]

The use of GAEN encryption should meet the GDPR requirements for data pseudonymisation.[6] (Data that is truly anonymised, per Recital 26 GDPR, is exempt from the GDPR's legal requirements.[5])

The contact tracing app, in particular the portion that notifies users of potential exposures, is a processing system per the Article 4 definition. This will bring into play the GDPR requirements of Article 5 "Principles relating to processing of personal data" and Article 24 "Responsibility of the controller."

Definition: Pseudonymisation

‘pseudonymisation’ means the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person;

Definition: Controller

‘controller’ means the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data; where the purposes and means of such processing are determined by Union or Member State law,

Definition: Processor

‘processor’ means a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller;

Definition: Consent

‘consent’ of the data subject means any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her;

As discussed in Project Management Considerations, the app user interface would obtain the requisite user consent. That would be sufficient to meet GDPR Article 6's lawful basis requirement to process data, without even needing to invoke Article 6(e) (public task) or Article 6(d) (protecting a vital interest).[5]

References

1. Contact Tracing Apps: Lessons Learned on Privacy, Autonomy, and the Need for Detailed and Thoughtful Implementation - (opens in new window). Katie Hogan, Briana Macedo, Venkata Macha, Arko Barman, Xiaoqian Jiang "Contact Tracing Apps: Lessons Learned on Privacy, Autonomy, and the Need for Detailed and Thoughtful Implementation."
2. COVID-19 contact tracing apps: a stress test for privacy, the GDPR, and data protection regimes | Journal of Law and the Biosciences | Oxford Academic. Accessed April 18, 2026. Laura Bradford, Mateo Aboy, and Kathleen Liddell. "COVID-19 contact tracing apps: a stress test for privacy, the GDPR, and data protection regimes", *Journal of Law and the Biosciences*, Volume 7, Issue 1, January-June 2020, DOI lsaa034. Published May 18, 2020.
3. Wikipedia - COVID-19 apps (citing, *inter alia*, Jessica Morley, Josh Cowls, Mariarosaria Taddeo & Luciano Floridi, "Ethical Guidelines for COVID-19 tracing apps," *Nature*, Volume 582 (2020), 29-31). Accessed May 4, 2026. Text is available under the Creative Commons Attribution-ShareAlike 4.0 License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.
4. Wikipedia - Basic reproduction number. Accessed May 4, 2026.
5. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 2016 O.J. (L 119) 1.
6. See European Data Protection Board, Guidelines 01/2025 on Pseudonymisation (January 16, 2025).

This page titled 8.3: Legal Considerations is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Robert Slate and Daniel Slate.

Terms

- **BLE:** Bluetooth Low Energy
- **GAEN:** Google/Apple Exposure Notification system that is Bluetooth LE based
- **GDPR:** the European Union's General Data Protection Regulation
- **GIS:** Global Information System
- **GIS:** Geographic Information System
- **GPS:** Global Positioning System
- **PACT:** MIT's Private Automated Contact Tracing project
- **PEPP-PT:** Pan-European Privacy Preserving Proximity Tracing
- **QR:** Quick Response. A QR code is a machine readable optical code that is essentially a two-dimensional bar code.
- **TTP:** Trusted Third Party

This page titled 8.4: Glossary is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Robert Slate and Daniel Slate.

INDEX

No index terms found.

This page titled Index is shared under a Public Domain license and was authored, remixed, and/or curated by Robert Slate and Daniel Slate.

GLOSSARY

S

Sample Word 1 | Sample Definition 1

This page titled Glossary is shared under a CC BY 4.0 license and was authored, remixed, and/or curated by Robert Slate and Daniel Slate.

DETAILED LICENSING

Overview

Title: Managing Technology Projects

Webpages: 196

Applicable Restrictions: Noncommercial

All licenses found:

- CC BY 4.0: 82.7% (162 pages)
- CC BY-NC 4.0: 10.2% (20 pages)
- Public Domain: 3.6% (7 pages)
- Undeclared: 1.5% (3 pages)
- Unknown License: 1% (2 pages)
- CC BY-NC-SA 4.0: 1% (2 pages)
- Unknown License: 0% (0 page)

By Page

- Managing Technology Projects - *Unknown License*
  - Front Matter - *Unknown License*
    - TitlePage - *CC BY 4.0*
    - InfoPage - *Public Domain*
    - Table of Contents - *Undeclared*
    - Licensing - *Public Domain*
    - Program Page - *Public Domain*
    - AI Use - *CC BY 4.0*
    - Accessibility Statement - *Public Domain*
    - About the Authors - *CC BY 4.0*
  - 1: Introduction - *CC BY 4.0*
  - 2: Operation Warp Speed and COVID-19 Vaccines - *CC BY 4.0*
    - 2.1: Introduction - *CC BY 4.0*
    - 2.2: Operation Warp Speed - *CC BY 4.0*
    - 2.3: COVID-19 Vaccine Development - *CC BY 4.0*
    - 2.4: Drug Development for Traditional Drugs and Biologics - *CC BY 4.0*
    - 2.5: Immune System and Vaccines - *CC BY 4.0*
      - 2.5.1: Section Goals - *CC BY 4.0*
      - 2.5.2: Review of the Anatomy of the Immune System - *CC BY 4.0*
      - 2.5.3: Vaccines - *CC BY 4.0*
      - 2.5.4: References - *CC BY 4.0*
      - 2.5.5: Glossary - *CC BY 4.0*
    - 2.6: Introduction to Pharmacology - *CC BY-NC-SA 4.0*
    - 2.7: FDA Vaccine Development 101 - *Public Domain*
    - 2.8: mRNA Vaccines - *CC BY 4.0*
    - 2.9: Investigational New Drug (IND) Application - *CC BY 4.0*
    - 2.10: Biologics License Application (BLA) - *CC BY 4.0*
    - 2.11: Emergency Use Authorization (EUA) - *CC BY 4.0*
    - 2.12: Current Good Manufacturing Practice (CGMP) - *CC BY 4.0*
    - 2.13: CDC Advisory Committee on Immunization Practices (ACIP) - *CC BY 4.0*
    - 2.14: Project Management Considerations - *CC BY 4.0*
      - 2.14.1: Project Schedule Preparation - *CC BY 4.0*
      - 2.14.2: Project Acceleration Techniques - *CC BY 4.0*
      - 2.14.3: Risk Management - *CC BY 4.0*
    - 2.15: Legal Considerations - *CC BY 4.0*
      - 2.15.1: National Emergency Powers - *CC BY 4.0*
      - 2.15.2: Contracts - *CC BY 4.0*
        - 2.15.2.1: Selecting the Type of Contract - *CC BY-NC-SA 4.0*
        - 2.15.2.2: BARDA Contracts - *CC BY 4.0*
        - 2.15.2.3: BARDA contract with Moderna - *CC BY 4.0*
        - 2.15.2.4: BARDA Contract with SiO2 Medical Products, Inc. for Vials - *CC BY 4.0*
      - 2.15.3: Risk Mitigation and Liability - *CC BY 4.0*
    - 2.16: Glossary - *CC BY 4.0*
  - 3: EU GDPR Compliance - *CC BY 4.0*
    - 3.1: Introduction - *CC BY 4.0*
    - 3.2: The Regulation - *CC BY 4.0*

- 3.2.1: Article 1 Subject-matter and objectives - *CC BY 4.0*
- 3.2.2: Article 3 Territorial Scope - *CC BY 4.0*
- 3.2.3: Article 4 Definitions - *CC BY 4.0*
- 3.2.4: Article 5 Principles related to processing of data - *CC BY 4.0*
- 3.2.5: Article 6 Lawfulness of processing - *CC BY 4.0*
- 3.2.6: Article 7 Conditions for consent - *CC BY 4.0*
- 3.2.7: Article 9 Processing of special categories of personal data - *CC BY 4.0*
- 3.2.8: Article 13 Information to be provided where personal data are collected from the data subject - *CC BY 4.0*
- 3.2.9: Article 15 Right of access by the data subject - *CC BY 4.0*
- 3.2.10: Article 16 Right to rectification - *CC BY 4.0*
- 3.2.11: Article 17 Right to erasure ('right to be forgotten') - *CC BY 4.0*
- 3.2.12: Article 18 Right to restriction of processing - *CC BY 4.0*
- 3.2.13: Article 20 Right to data portability - *CC BY 4.0*
- 3.2.14: Article 21 Right to object - *CC BY 4.0*
- 3.2.15: Article 22 Automated individual decision-making, including profiling - *CC BY 4.0*
- 3.2.16: Article 24 Responsibility of the controller - *CC BY 4.0*
- 3.2.17: Article 25 Data protection by design and by default - *CC BY 4.0*
- 3.2.18: Article 28 Processor - *CC BY 4.0*
- 3.2.19: Article 32 Security of processing - *CC BY 4.0*
- 3.2.20: Article 33 Notification of a personal data breach to the supervisory authority - *CC BY 4.0*
- 3.2.21: Article 34 Communication of a personal data breach to the data subject - *CC BY 4.0*
- 3.2.22: Articles 35-36 Data Protection Impact Assessments - *CC BY 4.0*
- 3.2.23: Article 37 Designation of the data protection officer - *CC BY 4.0*
- 3.2.24: Article 38 Position of the data protection officer - *CC BY 4.0*
- 3.2.25: Article 39 Tasks of the data protection officer - *CC BY 4.0*
- 3.2.26: Article 42 Certification - *CC BY 4.0*
- 3.2.27: Article 83 General conditions for imposing administrative fines - *CC BY 4.0*

- 3.3: Implementing GDPR Compliance - *CC BY 4.0*
  - 3.3.1: Implementation Considerations - *CC BY 4.0*
  - 3.4: Project Management Considerations - *CC BY 4.0*
  - 3.5: Legal Considerations - *CC BY 4.0*
  - 3.6: Glossary - *CC BY 4.0*

- 4: Network Switch Development - *CC BY 4.0*
  - 4.1: Introduction - *CC BY 4.0*
  - 4.2: Types of Network Switches - *CC BY 4.0*
  - 4.3: System Scope - *CC BY 4.0*
  - 4.4: Hardware System Design - *CC BY 4.0*
    - 4.4.1: Network ASICs - *CC BY 4.0*
    - 4.4.2: Chassis Design - *CC BY 4.0*
    - 4.4.3: Board Design - *CC BY 4.0*
    - 4.4.4: Signal integrity - *CC BY 4.0*
    - 4.4.5: Electro-magnetic Regulatory Compliance - *CC BY 4.0*
    - 4.4.6: Testing (HALT) - *CC BY 4.0*
  - 4.5: Software System Design - *CC BY 4.0*
  - 4.6: Project Management Considerations - *CC BY 4.0*
    - 4.6.1: Procurement - *CC BY 4.0*
      - 4.6.1.1: Bid Documents (RFI, RFP, etc.) - *CC BY 4.0*
      - 4.6.1.2: PR to PO procedure - *CC BY 4.0*
    - 4.6.2: Team Management - *CC BY 4.0*
      - 4.6.2.1: Content Theories of Motivation - *CC BY 4.0*
      - 4.6.2.2: Leadership - *CC BY 4.0*
        - 4.6.2.2.1: Introduction - *CC BY 4.0*
        - 4.6.2.2.2: The Nature of Leadership - *CC BY 4.0*
        - 4.6.2.2.3: The Leadership Process - *CC BY 4.0*
        - 4.6.2.2.4: Types of Leaders and Leader Emergence - *CC BY 4.0*
        - 4.6.2.2.5: The Trait Approach to Leadership - *CC BY 4.0*
        - 4.6.2.2.6: Behavioral Approaches to Leadership - *CC BY 4.0*
        - 4.6.2.2.7: Situational (Contingency) Approaches to Leadership - *CC BY 4.0*
        - 4.6.2.2.8: Substitutes for and Neutralizers of Leadership - *CC BY 4.0*
        - 4.6.2.2.9: Transformational, Visionary, and Charismatic Leadership - *CC BY 4.0*
        - 4.6.2.2.10: Leadership Needs in the 21st Century - *CC BY 4.0*
        - 4.6.2.2.11: Summary - *CC BY 4.0*
  - 4.7: Legal Considerations - *CC BY 4.0*
    - 4.7.1: Patents - *CC BY 4.0*
      - 4.7.1.1: The Patenting Process - *CC BY 4.0*
      - 4.7.1.2: Patent Disclosures - *CC BY 4.0*
      - 4.7.1.3: Patent cross-licensing - *CC BY*

4.0
- 4.7.2: Trade Secrets - *CC BY 4.0*
  - 4.7.2.1: Trade Secret Protections - *CC BY 4.0*
  - 4.7.2.2: The Foundations of Trade Secrets Law - *CC BY 4.0*
  - 4.7.2.3: Elements of a Trade Secret - *CC BY 4.0*
  - 4.7.2.4: The Secrecy Requirement - *CC BY 4.0*
  - 4.7.2.5: Misappropriation of Trade Secrets - *CC BY 4.0*
  - 4.7.2.6: Remedies Available for the Misappropriation of Trade Secrets - *CC BY 4.0*
  - 4.7.2.7: Assessment Questions - *CC BY 4.0*
- 4.7.3: Trademark Basics - *CC BY 4.0*
  - 4.7.3.1: Core Concepts - *CC BY 4.0*
  - 4.7.3.2: Early Trademark Systems - *CC BY 4.0*
  - 4.7.3.3: U.S. Trademark Law - *CC BY 4.0*
  - 4.7.3.4: Four Types of Trademarks - *CC BY 4.0*
  - 4.7.3.5: Subject Matter of Trademarks - *CC BY 4.0*
  - 4.7.3.6: The Spectrum of Distinctiveness - *CC BY 4.0*
  - 4.7.3.7: Bars to Trademark - *CC BY 4.0*
  - 4.7.3.8: Establishing Trademark Protection - *CC BY 4.0*
  - 4.7.3.9: Trademark Infringement - *CC BY 4.0*
  - 4.7.3.10: Trademark Remedies - *CC BY 4.0*
  - 4.7.3.11: Fair Use of Trademarks - *CC BY 4.0*
  - 4.7.3.12: Assessment Questions - *CC BY 4.0*
- 4.8: Glossary - *CC BY 4.0*

◦ 5: EU RoHS Compliance - *CC BY 4.0*
- 5.1: Introduction - *CC BY 4.0*
- 5.2: The Directive - *CC BY 4.0*
- 5.3: Implementing the Directive - *CC BY 4.0*
- 5.4: Project Management Considerations - *CC BY 4.0*
- 5.5: Legal Considerations - *CC BY 4.0*
- 5.6: Glossary - *CC BY 4.0*

◦ 6: Boeing 737 Max Quality Issue Recovery 2018 - *CC BY 4.0*
- 6.1: Introduction - *CC BY 4.0*
- 6.2: Project Scope- Steps to Recertification - *CC BY 4.0*
- 6.3: Project Management Considerations - *CC BY 4.0*
- 6.4: Managing Quality - *CC BY-NC 4.0*
  - 6.4.1: Introduction to Managing Quality - *CC BY-NC 4.0*
  - 6.4.2: Gurus of Quality - *CC BY-NC 4.0*
    - 6.4.2.1: Walter Shewhart (1891-1967) - *CC BY-NC 4.0*
    - 6.4.2.2: Edwards Deming (1900-1993) - *CC BY-NC 4.0*
    - 6.4.2.3: Joseph Juran (1904-2008) - *CC BY-NC 4.0*
    - 6.4.2.4: Philip Crosby (1926-2001) and Armand Fiegenbaum (1920-2014) - *CC BY-NC 4.0*
  - 6.4.3: Cost of Quality - *CC BY-NC 4.0*
  - 6.4.4: Quality Systems - *CC BY-NC 4.0*
    - 6.4.4.1: Total Quality Management (TQM) - *CC BY-NC 4.0*
    - 6.4.4.2: ISO - *CC BY-NC 4.0*
    - 6.4.4.3: Six Sigma - *CC BY-NC 4.0*
  - 6.4.5: Tools for Quality Improvement - *CC BY-NC 4.0*
    - 6.4.5.1: Check Sheets - *CC BY-NC 4.0*
    - 6.4.5.2: Histograms - *CC BY-NC 4.0*
    - 6.4.5.3: Control Charts - *CC BY-NC 4.0*
    - 6.4.5.4: Pareto Charts - *CC BY-NC 4.0*
    - 6.4.5.5: Scatter Diagrams - *CC BY-NC 4.0*
    - 6.4.5.6: Cause and Effect Diagrams - *CC BY-NC 4.0*
  - 6.4.6: References - *CC BY-NC 4.0*
- 6.5: Legal Considerations - *CC BY 4.0*
- 6.6: Glossary - *CC BY 4.0*

◦ 7: Artificial Intelligence Usage by Utilities - *CC BY 4.0*
- 7.1: Introduction - *CC BY 4.0*
- 7.2: Specific Areas of Application of Artificial Intelligence - *CC BY 4.0*
- 7.3: Artificial Intelligence Use by a California Utility (PGandE) - *CC BY 4.0*
- 7.4: AI Initiatives by other California Utilities - *CC BY 4.0*
  - 7.4.1: Southern California Edison - *CC BY 4.0*
  - 7.4.2: San Diego Gas and Electric - *CC BY 4.0*
- 7.5: Applying Artificial Intelligence to Wildfire Detection and Management - *CC BY 4.0*
  - 7.5.1: Vendor Solutions - *CC BY 4.0*
    - 7.5.1.1: N5 Sensors - *CC BY 4.0*
    - 7.5.1.2: Ladris - *CC BY 4.0*
    - 7.5.1.3: Pano.ai - *CC BY 4.0*
- 7.6: Project Management Considerations - *CC BY 4.0*
- 7.7: Legal Considerations - *CC BY 4.0*
  - 7.7.1: CPUC Governance - *CC BY 4.0*
  - 7.7.2: Wildfire Camera Privacy Issues - *CC BY 4.0*
  - 7.7.3: Impact of Potential AI Regulations - *CC BY 4.0*
- 7.8: Glossary of Terms - *CC BY 4.0*

- 8: COVID-19 Contact Tracing App - *CC BY 4.0*
  - 8.1: Description of the problem and potential solutions - *CC BY 4.0*
  - 8.2: Project Management Considerations - *CC BY 4.0*
  - 8.3: Legal Considerations - *CC BY 4.0*
  - 8.4: Glossary - *CC BY 4.0*
- Back Matter - *Undeclared*
  - Index - *Public Domain*
  - Glossary - *CC BY 4.0*
  - Detailed Licensing - *Public Domain*
  - Detailed Licensing - *Undeclared*

This page titled Managing Technology Projects is shared under a license and was authored, remixed, and/or curated by Robert Slate and Daniel Slate.

DETAILED LICENSING

Overview

Title: Managing Technology Projects

Webpages: 196

Applicable Restrictions: Noncommercial

All licenses found:

- CC BY 4.0: 82.7% (162 pages)
- CC BY-NC 4.0: 10.2% (20 pages)
- Public Domain: 3.6% (7 pages)
- Undeclared: 1.5% (3 pages)
- Unknown License: 1% (2 pages)
- CC BY-NC-SA 4.0: 1% (2 pages)
- Unknown License: 0% (0 page)

By Page

- Managing Technology Projects - *Unknown License*
  - Front Matter - *Unknown License*
    - TitlePage - *CC BY 4.0*
    - InfoPage - *Public Domain*
    - Table of Contents - *Undeclared*
    - Licensing - *Public Domain*
    - Program Page - *Public Domain*
    - AI Use - *CC BY 4.0*
    - Accessibility Statement - *Public Domain*
    - About the Authors - *CC BY 4.0*
  - 1: Introduction - *CC BY 4.0*
  - 2: Operation Warp Speed and COVID-19 Vaccines - *CC BY 4.0*
    - 2.1: Introduction - *CC BY 4.0*
    - 2.2: Operation Warp Speed - *CC BY 4.0*
    - 2.3: COVID-19 Vaccine Development - *CC BY 4.0*
    - 2.4: Drug Development for Traditional Drugs and Biologics - *CC BY 4.0*
    - 2.5: Immune System and Vaccines - *CC BY 4.0*
      - 2.5.1: Section Goals - *CC BY 4.0*
      - 2.5.2: Review of the Anatomy of the Immune System - *CC BY 4.0*
      - 2.5.3: Vaccines - *CC BY 4.0*
      - 2.5.4: References - *CC BY 4.0*
      - 2.5.5: Glossary - *CC BY 4.0*
    - 2.6: Introduction to Pharmacology - *CC BY-NC-SA 4.0*
    - 2.7: FDA Vaccine Development 101 - *Public Domain*
    - 2.8: mRNA Vaccines - *CC BY 4.0*
    - 2.9: Investigational New Drug (IND) Application - *CC BY 4.0*
    - 2.10: Biologics License Application (BLA) - *CC BY 4.0*
    - 2.11: Emergency Use Authorization (EUA) - *CC BY 4.0*
    - 2.12: Current Good Manufacturing Practice (CGMP) - *CC BY 4.0*
    - 2.13: CDC Advisory Committee on Immunization Practices (ACIP) - *CC BY 4.0*
    - 2.14: Project Management Considerations - *CC BY 4.0*
      - 2.14.1: Project Schedule Preparation - *CC BY 4.0*
      - 2.14.2: Project Acceleration Techniques - *CC BY 4.0*
      - 2.14.3: Risk Management - *CC BY 4.0*
    - 2.15: Legal Considerations - *CC BY 4.0*
      - 2.15.1: National Emergency Powers - *CC BY 4.0*
      - 2.15.2: Contracts - *CC BY 4.0*
        - 2.15.2.1: Selecting the Type of Contract - *CC BY-NC-SA 4.0*
        - 2.15.2.2: BARDA Contracts - *CC BY 4.0*
        - 2.15.2.3: BARDA contract with Moderna - *CC BY 4.0*
        - 2.15.2.4: BARDA Contract with SiO2 Medical Products, Inc. for Vials - *CC BY 4.0*
      - 2.15.3: Risk Mitigation and Liability - *CC BY 4.0*
    - 2.16: Glossary - *CC BY 4.0*
  - 3: EU GDPR Compliance - *CC BY 4.0*
    - 3.1: Introduction - *CC BY 4.0*
    - 3.2: The Regulation - *CC BY 4.0*

- 3.2.1: Article 1 Subject-matter and objectives - *CC BY 4.0*
- 3.2.2: Article 3 Territorial Scope - *CC BY 4.0*
- 3.2.3: Article 4 Definitions - *CC BY 4.0*
- 3.2.4: Article 5 Principles related to processing of data - *CC BY 4.0*
- 3.2.5: Article 6 Lawfulness of processing - *CC BY 4.0*
- 3.2.6: Article 7 Conditions for consent - *CC BY 4.0*
- 3.2.7: Article 9 Processing of special categories of personal data - *CC BY 4.0*
- 3.2.8: Article 13 Information to be provided where personal data are collected from the data subject - *CC BY 4.0*
- 3.2.9: Article 15 Right of access by the data subject - *CC BY 4.0*
- 3.2.10: Article 16 Right to rectification - *CC BY 4.0*
- 3.2.11: Article 17 Right to erasure ('right to be forgotten') - *CC BY 4.0*
- 3.2.12: Article 18 Right to restriction of processing - *CC BY 4.0*
- 3.2.13: Article 20 Right to data portability - *CC BY 4.0*
- 3.2.14: Article 21 Right to object - *CC BY 4.0*
- 3.2.15: Article 22 Automated individual decision-making, including profiling - *CC BY 4.0*
- 3.2.16: Article 24 Responsibility of the controller - *CC BY 4.0*
- 3.2.17: Article 25 Data protection by design and by default - *CC BY 4.0*
- 3.2.18: Article 28 Processor - *CC BY 4.0*
- 3.2.19: Article 32 Security of processing - *CC BY 4.0*
- 3.2.20: Article 33 Notification of a personal data breach to the supervisory authority - *CC BY 4.0*
- 3.2.21: Article 34 Communication of a personal data breach to the data subject - *CC BY 4.0*
- 3.2.22: Articles 35-36 Data Protection Impact Assessments - *CC BY 4.0*
- 3.2.23: Article 37 Designation of the data protection officer - *CC BY 4.0*
- 3.2.24: Article 38 Position of the data protection officer - *CC BY 4.0*
- 3.2.25: Article 39 Tasks of the data protection officer - *CC BY 4.0*
- 3.2.26: Article 42 Certification - *CC BY 4.0*
- 3.2.27: Article 83 General conditions for imposing administrative fines - *CC BY 4.0*

- 3.3: Implementing GDPR Compliance - *CC BY 4.0*
  - 3.3.1: Implementation Considerations - *CC BY 4.0*
  - 3.4: Project Management Considerations - *CC BY 4.0*
  - 3.5: Legal Considerations - *CC BY 4.0*
  - 3.6: Glossary - *CC BY 4.0*

- 4: Network Switch Development - *CC BY 4.0*
  - 4.1: Introduction - *CC BY 4.0*
  - 4.2: Types of Network Switches - *CC BY 4.0*
  - 4.3: System Scope - *CC BY 4.0*
  - 4.4: Hardware System Design - *CC BY 4.0*
    - 4.4.1: Network ASICs - *CC BY 4.0*
    - 4.4.2: Chassis Design - *CC BY 4.0*
    - 4.4.3: Board Design - *CC BY 4.0*
    - 4.4.4: Signal integrity - *CC BY 4.0*
    - 4.4.5: Electro-magnetic Regulatory Compliance - *CC BY 4.0*
    - 4.4.6: Testing (HALT) - *CC BY 4.0*
  - 4.5: Software System Design - *CC BY 4.0*
  - 4.6: Project Management Considerations - *CC BY 4.0*
    - 4.6.1: Procurement - *CC BY 4.0*
      - 4.6.1.1: Bid Documents (RFI, RFP, etc.) - *CC BY 4.0*
      - 4.6.1.2: PR to PO procedure - *CC BY 4.0*
    - 4.6.2: Team Management - *CC BY 4.0*
      - 4.6.2.1: Content Theories of Motivation - *CC BY 4.0*
      - 4.6.2.2: Leadership - *CC BY 4.0*
        - 4.6.2.2.1: Introduction - *CC BY 4.0*
        - 4.6.2.2.2: The Nature of Leadership - *CC BY 4.0*
        - 4.6.2.2.3: The Leadership Process - *CC BY 4.0*
        - 4.6.2.2.4: Types of Leaders and Leader Emergence - *CC BY 4.0*
        - 4.6.2.2.5: The Trait Approach to Leadership - *CC BY 4.0*
        - 4.6.2.2.6: Behavioral Approaches to Leadership - *CC BY 4.0*
        - 4.6.2.2.7: Situational (Contingency) Approaches to Leadership - *CC BY 4.0*
        - 4.6.2.2.8: Substitutes for and Neutralizers of Leadership - *CC BY 4.0*
        - 4.6.2.2.9: Transformational, Visionary, and Charismatic Leadership - *CC BY 4.0*
        - 4.6.2.2.10: Leadership Needs in the 21st Century - *CC BY 4.0*
        - 4.6.2.2.11: Summary - *CC BY 4.0*
  - 4.7: Legal Considerations - *CC BY 4.0*
    - 4.7.1: Patents - *CC BY 4.0*
      - 4.7.1.1: The Patenting Process - *CC BY 4.0*
      - 4.7.1.2: Patent Disclosures - *CC BY 4.0*
      - 4.7.1.3: Patent cross-licensing - *CC BY*

4.0
- 4.7.2: Trade Secrets - *CC BY 4.0*
  - 4.7.2.1: Trade Secret Protections - *CC BY 4.0*
  - 4.7.2.2: The Foundations of Trade Secrets Law - *CC BY 4.0*
  - 4.7.2.3: Elements of a Trade Secret - *CC BY 4.0*
  - 4.7.2.4: The Secrecy Requirement - *CC BY 4.0*
  - 4.7.2.5: Misappropriation of Trade Secrets - *CC BY 4.0*
  - 4.7.2.6: Remedies Available for the Misappropriation of Trade Secrets - *CC BY 4.0*
  - 4.7.2.7: Assessment Questions - *CC BY 4.0*
- 4.7.3: Trademark Basics - *CC BY 4.0*
  - 4.7.3.1: Core Concepts - *CC BY 4.0*
  - 4.7.3.2: Early Trademark Systems - *CC BY 4.0*
  - 4.7.3.3: U.S. Trademark Law - *CC BY 4.0*
  - 4.7.3.4: Four Types of Trademarks - *CC BY 4.0*
  - 4.7.3.5: Subject Matter of Trademarks - *CC BY 4.0*
  - 4.7.3.6: The Spectrum of Distinctiveness - *CC BY 4.0*
  - 4.7.3.7: Bars to Trademark - *CC BY 4.0*
  - 4.7.3.8: Establishing Trademark Protection - *CC BY 4.0*
  - 4.7.3.9: Trademark Infringement - *CC BY 4.0*
  - 4.7.3.10: Trademark Remedies - *CC BY 4.0*
  - 4.7.3.11: Fair Use of Trademarks - *CC BY 4.0*
  - 4.7.3.12: Assessment Questions - *CC BY 4.0*
- 4.8: Glossary - *CC BY 4.0*

◦ 5: EU RoHS Compliance - *CC BY 4.0*
- 5.1: Introduction - *CC BY 4.0*
- 5.2: The Directive - *CC BY 4.0*
- 5.3: Implementing the Directive - *CC BY 4.0*
- 5.4: Project Management Considerations - *CC BY 4.0*
- 5.5: Legal Considerations - *CC BY 4.0*
- 5.6: Glossary - *CC BY 4.0*

◦ 6: Boeing 737 Max Quality Issue Recovery 2018 - *CC BY 4.0*
- 6.1: Introduction - *CC BY 4.0*
- 6.2: Project Scope- Steps to Recertification - *CC BY 4.0*
- 6.3: Project Management Considerations - *CC BY 4.0*
- 6.4: Managing Quality - *CC BY-NC 4.0*
  - 6.4.1: Introduction to Managing Quality - *CC BY-NC 4.0*
  - 6.4.2: Gurus of Quality - *CC BY-NC 4.0*
    - 6.4.2.1: Walter Shewhart (1891-1967) - *CC BY-NC 4.0*
    - 6.4.2.2: Edwards Deming (1900-1993) - *CC BY-NC 4.0*
    - 6.4.2.3: Joseph Juran (1904-2008) - *CC BY-NC 4.0*
    - 6.4.2.4: Philip Crosby (1926-2001) and Armand Fiegenbaum (1920-2014) - *CC BY-NC 4.0*
  - 6.4.3: Cost of Quality - *CC BY-NC 4.0*
  - 6.4.4: Quality Systems - *CC BY-NC 4.0*
    - 6.4.4.1: Total Quality Management (TQM) - *CC BY-NC 4.0*
    - 6.4.4.2: ISO - *CC BY-NC 4.0*
    - 6.4.4.3: Six Sigma - *CC BY-NC 4.0*
  - 6.4.5: Tools for Quality Improvement - *CC BY-NC 4.0*
    - 6.4.5.1: Check Sheets - *CC BY-NC 4.0*
    - 6.4.5.2: Histograms - *CC BY-NC 4.0*
    - 6.4.5.3: Control Charts - *CC BY-NC 4.0*
    - 6.4.5.4: Pareto Charts - *CC BY-NC 4.0*
    - 6.4.5.5: Scatter Diagrams - *CC BY-NC 4.0*
    - 6.4.5.6: Cause and Effect Diagrams - *CC BY-NC 4.0*
  - 6.4.6: References - *CC BY-NC 4.0*
- 6.5: Legal Considerations - *CC BY 4.0*
- 6.6: Glossary - *CC BY 4.0*

◦ 7: Artificial Intelligence Usage by Utilities - *CC BY 4.0*
- 7.1: Introduction - *CC BY 4.0*
- 7.2: Specific Areas of Application of Artificial Intelligence - *CC BY 4.0*
- 7.3: Artificial Intelligence Use by a California Utility (PGandE) - *CC BY 4.0*
- 7.4: AI Initiatives by other California Utilities - *CC BY 4.0*
  - 7.4.1: Southern California Edison - *CC BY 4.0*
  - 7.4.2: San Diego Gas and Electric - *CC BY 4.0*
- 7.5: Applying Artificial Intelligence to Wildfire Detection and Management - *CC BY 4.0*
  - 7.5.1: Vendor Solutions - *CC BY 4.0*
    - 7.5.1.1: N5 Sensors - *CC BY 4.0*
    - 7.5.1.2: Ladris - *CC BY 4.0*
    - 7.5.1.3: Pano.ai - *CC BY 4.0*
- 7.6: Project Management Considerations - *CC BY 4.0*
- 7.7: Legal Considerations - *CC BY 4.0*
  - 7.7.1: CPUC Governance - *CC BY 4.0*
  - 7.7.2: Wildfire Camera Privacy Issues - *CC BY 4.0*
  - 7.7.3: Impact of Potential AI Regulations - *CC BY 4.0*
- 7.8: Glossary of Terms - *CC BY 4.0*

- 8: COVID-19 Contact Tracing App - *CC BY 4.0*
  - 8.1: Description of the problem and potential solutions - *CC BY 4.0*
  - 8.2: Project Management Considerations - *CC BY 4.0*
  - 8.3: Legal Considerations - *CC BY 4.0*
  - 8.4: Glossary - *CC BY 4.0*
- Back Matter - *Undeclared*
  - Index - *Public Domain*
  - Glossary - *CC BY 4.0*
  - Detailed Licensing - *Public Domain*
  - Detailed Licensing - *Undeclared*

This page titled Managing Technology Projects is shared under a license and was authored, remixed, and/or curated by Robert Slate and Daniel Slate.