MANAGING DATA *for*
PATRON PRIVACY

Comprehensive Strategies for Libraries

KRISTIN BRINEY *and* BECKY YOOSE

ALA Editions
CHICAGO | 2022

KRISTIN BRINEY is the Biology & Biological Engineering Librarian at the California Institute of Technology. She has a PhD in physical chemistry and an MLIS, both from the University of Wisconsin–Madison. In addition to being a liaison librarian, Kristin teaches and consults on research data management for both researchers and librarians. She is the author of the book *Data Management for Researchers: Organize, Maintain and Share Your Data for Research Success* and conducts research on student privacy in academic libraries.

BECKY YOOSE is the founder of and Library Data Privacy Consultant for LDH Consulting Services. She received her MA-LIS from the University of Wisconsin–Madison and is a certified privacy professional by the International Association of Privacy Professionals. With well over a decade of experience in wrangling library data in academic and public libraries, Becky helps libraries and library vendors navigate the intersection of library data and privacy.

ISBNs
978-0-8389-3828-7 (paper)
978-0-8389-3734-1 (pdf)

Library of Congress Cataloging-in-Publication Data

Names: Briney, Kristin, author. | Yoose, Becky, author.
Title: Managing data for patron privacy : comprehensive strategies for libraries / Kristin Briney and Becky Yoose.
Description: Chicago : ALA Editions, 2022. | Includes bibliographical references and index. | Summary: "This book will provide a deeper understanding of how to protect patron data and how to implement strategies to help you improve patron data privacy at your library"—Provided by publisher.
Identifiers: LCCN 2021062319 (print) | LCCN 2021062320 (ebook) | ISBN 9780838938287 (paperback) | ISBN 9780838937341 (pdf)
Subjects: LCSH: Library legislation—United States. | Data protection—Law and legislation—United States.
Classification: LCC KF4319.P75 B75 2022 (print) | LCC KF4319.P75 (ebook) | DDC 342.7308/58—dc23/eng/20220204
LC record available at https://lccn.loc.gov/2021062319
LC ebook record available at https://lccn.loc.gov/2021062320

Book design by Kim Hudgins in the Chaparral, Tisa, and Proxima Nova typefaces.
Cover image © kitka/Adobe Stock

♾ This paper meets the requirements of ANSI/NISO Z39.48–1992 (Permanence of Paper).

Printed in the United States of America

26 25 24 23 22 5 4 3 2 1

Contents

Acknowledgments *ix*

Chapter 1 The Value of Data and Privacy 1

Chapter 2 The Data Landscape. 11

Chapter 3 Data Inventory 27

Chapter 4 Risk Assessment 41

Chapter 5 Library Data Privacy Policy and Operations. 61

Chapter 6 Security 101. 83

Chapter 7 Vendor Relations 105

Chapter 8 Library Assessment 121

Chapter 9 Library Worker Privacy Training 135

Chapter 10 Developing and Sustaining a Culture of Privacy . . . 149

Index *157*

Chapter 1

THE VALUE OF DATA AND PRIVACY

THE AVERAGE DATA BREACH IS ESTIMATED TO COST $4.24 MILLION, or $161 per record, according to the 2021 Cost of a Data Breach Report from IBM Security.[1] The study found that over a third of this cost comes from lost business and that costs occur for several years after the breach. Note that this is the global average—the average cost of a data breach in the United States in 2021 was even higher at $9.05 million. It's very clear from the report that data breaches are expensive and damaging.

Libraries experience data breaches, even while the numbers used in the *Cost of a Data Breach Report* better represent businesses and the medical sector. Several public library systems have dealt with data breaches in the past few years:

- Contra Costa County Library in California had its data compromised in a ransomware attack in early 2020; luckily, the library had stopped collecting driver's license numbers the year before and had deleted older license numbers, mitigating the damage from the attack.[2]
- The Columbus Metropolitan Library in Ohio experienced a breach in 2018 that was most likely the result of spear phishing.[3]
- The Indianhead Federated Library System in Wisconsin experienced a breach in 2017 that resulted in the loss of data ranging from names and addresses to birth dates and driver's license numbers.[4]
- The Alameda County Library in California experienced a data breach in 2017, though it was not clear how many patron records were exposed.[5]

These are just breaches that have been reported in the news and don't represent all of the public library breaches that have occurred.

Academic libraries are not immune from breaches. These breaches usually happen at the university level and often focus on credential theft (which can then be used to access library resources). In 2019 institutions ranging from large public universities (e.g., Arizona State University and Oregon State University) to community colleges (e.g., City College of San Francisco and Community College of Allegheny County) to small private colleges (e.g., Oberlin College and Hamilton College) experienced data breaches.[6] Universities were also specifically targeted by the Silent Librarian phishing campaign since 2013, in which people received e-mails stating that their library account had expired and asking them to follow a fraudulent link to reactivate it. The scope of the attack was large. The Silent Librarian campaign is estimated to have affected 144 universities in the United States and another 176 universities worldwide, resulting in almost 8,000 compromised accounts and causing the loss of over 30 terabytes of stolen data and intellectual property (valued at over $3 billion).[7] In 2018, the US Justice Department announced charges against nine Iranians behind the attacks.[8] Finally, K–12 education and public school systems also had a large number of data breaches in 2019, which may have affected school libraries.[9]

Library vendors have been breached as well. The video-streaming service Kanopy exposed API and website access logs in 2019.[10] Scholarly publisher Elsevier also had a breach in 2019 that exposed users' e-mails and passwords online.[11] In 2014 Adobe Digital Editions was found to be collecting detailed logs of library users' e-book reading and sending this data unencrypted back to Adobe, a fact of which libraries were not made aware.[12] Again, it is very likely that other libraries and vendors have had breaches that haven't been reported in the news or even been discovered yet.

All of these stories show that data has value. Part of that value is monetary, as exemplified by the attention-grabbing numbers at the very beginning of this chapter, but part of data's value is inherent to what it represents to the person to whom the data belongs. There is value in the trust a patron loses when the library is unable to properly care for their data and in the privacy lost should a patron's data be shared beyond the library. The value of privacy and trust in the library are central to this book's focus on how to handle patron data responsibly.

Valuing privacy and trust means that libraries should be just as concerned about small data leaks as they are about large data breaches. A leak happens when circulation staff give a woman's account information to her abusive ex-husband or give parents access to their teen's separate library account to find checkouts relating to questioning one's gender or sexuality. Library staff can also improperly access and

use patron data, such as when contacting patrons for a nonlibrary purpose. Accidents also happen, for example, when sending patron data to the wrong e-mail address or when decommissioning outdated computers or photocopiers without properly destroying data on hard drives. Libraries also share patron data intentionally, such as for assessment (or, as higher education calls it, learning analytics) and marketing segmentation research; this can be considered as an ethical breach because patrons often do not know that their data is being used in these projects. These small data leaks are much more frequent than large breaches and require every library worker to help prevent them.

Data is a major part of running a library, and the way the library cares for that data has a major impact on our patrons' privacy and trust in the library (and, yes, good data management is also good for the library's budget by preventing a data breach). Only by protecting data responsibly does the library demonstrate that it values patrons' privacy and trust. This book explores the many ways to care for our patrons' data through the lens of patron privacy, a core tenet of librarianship.

Patron Privacy

This book starts from the default position that we are all working toward improved patron privacy. All actions and strategies recommended within these pages are offered in pursuit of that goal. This is because privacy is a general good and core to the values of the library and society at large. Privacy supports personal autonomy, as free inquiry and unmonitored communication allow for self-development. Privacy also allows for anonymous speech and free association between people, which are valuable practices within a democracy. The right to act as one's true self without being monitored is an important human right.

It is beyond the scope of this book to provide a definitive meaning to the term *privacy*. Scholars regularly debate different privacy definitions and frameworks, so we will refer to several existing sources rather than offering a new definition here. A relevant definition comes from the American Library Association (ALA), which places privacy in the library context. In its document "An Interpretation of the Library Bill of Rights," ALA states that privacy "includes the right to open inquiry without having the subject of one's interest examined or scrutinized by others, in person or online."[13] (The document also notes that privacy is different from confidentiality, which "exists when a library is in possession of personally identifiable information about its users and keeps that information private on their behalf.") This framing is based in information privacy, which allows for unrestricted

consultation of information resources. Privacy, however, is a broader concept than this, as seen in the simple definition of privacy offered in 1890 by Warren and Brandeis: the "right to be let alone."[14] Both definitions are useful for thinking about patron privacy in libraries.

Privacy can also be defined with respect to how information flows. Helen Nissenbaum describes privacy through the framework of contextual integrity, wherein information is expected to be shared within particular contexts but not within others.[15] For example, sharing medical information with one's doctor is expected, but it breaks contextual integrity for your doctor to share your medical information with a stranger; the second case invades your privacy because you don't expect your medical information to be shared in this context. Similarly, Daniel Solove defines privacy through the types of activities that threaten to violate it: information collection, information processing, information dissemination, and invasion.[16] We find framing privacy through controlling information flows to be useful, as much of the rest of the book focuses on managing the flows of patron data.

A final model that is useful for thinking about privacy in a library context is "physical-equivalent privacy."[17] Dorothea Salo's model challenges us to translate the use of electronic resources and services into a comparable, physical equivalent to evaluate privacy threats. Under the heuristic, libraries should avoid electronic resource services, vendors, and standards that afford less privacy than their physical equivalents. For example, libraries should avoid detailed website tracking because its physical equivalent, following patrons throughout their time at the physical library, would be a privacy violation. Libraries have long-established library privacy practices around physical resources and services, so physical-equivalent privacy is a useful starting point for evaluating the privacy implications of virtual services and resources.

No matter the exact definition, protecting patron privacy is core to the library profession. Privacy is encapsulated in our codes of ethics, from both ALA—"we protect each library user's right to privacy and confidentiality with respect to information sought or received and resources consulted, borrowed, acquired or transmitted"—and IFLA (International Federation of Library Associations and Institutions)—"librarians and other information workers respect personal privacy, and the protection of personal data, necessarily shared between individuals and institutions."[18] Article VII of ALA's Library Bill of Rights also states that "all people, regardless of origin, age, background, or views, possess a right to privacy and confidentiality in their library use. Libraries should advocate for, educate about, and

protect people's privacy, safeguarding all library use data, including personally identifiable information."[19] The second sentence specifically calls out protecting library data, including personally identifiable information, which is the major theme of this book.

Why Data Management?

Our modern economy runs on data. It is the backbone of our financial systems, social media, educational systems, retail, and government. Similarly, the modern library generates and utilizes a large amount of data. Consider all of the data involved when a patron checks out a book: the book's data appears in the catalog; the patron's data is found in the integrated library system (ILS); and when the patron checks out the book, the system connects the patron to the book in the form of new data with a time stamp; there may also be an e-mail associated with this transaction—new data—to tell the patron the book is available and/or coming due. Patrons generate data as they access resources via proxy servers or amass fines for overdue items. Data is generated via building security, such as from security cameras or card access swipes in libraries with restricted access and from paper sign-in sheets for library events. These represent just some of the types of data that libraries regularly collect and use.

The simple truth is that by using a modern library, our patrons leave data trails behind them. This is not a problem, in itself, as some of this data is necessary for running a library. We need to track who has which books checked out, for example, so we can get them back or monitor computer session time if there are time limits. The problems occur, however, when data is breached, spreads, is improperly accessed, or succumbs to any of the other myriad scenarios that happen with data. These issues quickly move from being only data problems to being privacy problems as well, as much of the data libraries hold is about patrons and their activities.

While many resources covering patron privacy exist, this book focuses on data management as a means to protect patron privacy. This is because the data itself is what encapsulates the patron information that must be kept private. When data is breached or leaked, patron privacy is lost. Good ethical choices provide the framework, but good data management actually determines the degree to which patron privacy gets implemented in a modern library. The goal of this book is for libraries of all types to better manage their data to ensure it is being handled responsibly, efficiently, and in a manner that protects patron privacy.

Defining Data

As this is a book about data, it is pertinent to also define what *data* means. Many definitions of data exist in every context that data is used, from personally identifiable information (PII)—more on this soon—to student data, from research data to health data. This book specifically focuses on digital patron data, though it touches on other types of data. We define library patron data as any digital information directly about patrons, or information that records interactions between patrons and the library or its resources, that is held by the library itself, its vendors, or other partners.

Patron data comes in many forms, from numeric to text to images and more. Consider the range of data types that may be found in a library: patron name and e-mail (text), checkout counts (numbers), pictures from library events (images), recordings from security cameras (video), and voice messages left with the library (sound recordings). Many people associate data with only numbers or information in spreadsheets, but it is important to recognize that patron data may exist in formats that have not historically been considered as data. This book considers any type of information about patrons and their library activities to be patron data, no matter the format.

PERSONALLY IDENTIFIABLE INFORMATION

Within all types of data, PII is especially important to define so as to understand how individual patrons may be identified from their data. Broadly, PII is any information about a person that can be used, either as a single data point or with other data points, to identify who that person is.

In the United States, we can look to the National Institute of Standards and Technology (NIST) for guidance on what constitutes PII. They use this definition of PII:

> Any information about an individual maintained by an agency, including (1) any information that can be used to distinguish or trace an individual's identity, such as name, social security number, date and place of birth, mother's maiden name, or biometric records; and (2) any other information that is linked or linkable to an individual, such as medical, educational, financial, and employment information.[20]

It is worth noting that NIST differentiates between PII that can distinguish an individual from others; trace an individual to their activities or status; and link together information that is logically associated but is not held in the same dataset, so as to identify the individual.

Other countries have slightly different standards for the definition of PII. Of note is the recent General Data Protection Regulation (GDPR) from Europe that provides this definition of *personal data*:

> 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.[21]

The GDPR definition covers a wide range of personal information, from identification numbers to genetic information, under its definition of PII.

As outlined in both definitions, PII comes in different types. The most identifiable data is called a *direct identifier*, as a single data point can directly reveal a person's identity. Direct identifiers are what are most commonly thought of as PII, as there is a very clear relationship between this data and a person's identity. The second type of PII is the "indirect identifier," which often describes a person's attributes. Indirect identifiers don't reveal a person's identity with one data point but are, instead, extremely powerful in combination; for example, the combination of birth date, zip code, and gender is estimated to uniquely identify 87 percent of the US population.[22] A third type of PII, behavioral data, describes a person's actions. While a single action may not identify an individual, actions viewed collectively demonstrate patterns in a person's behavior which can then be used to identify that person. The different types of PII are discussed more in chapter 4.

As you read through this book, it is important to keep all of these types of PII in mind as they represent the wide range of data that libraries hold about patrons and their behaviors. The full spectrum of patron information must be managed properly to protect patron privacy, though the strategies for managing different types of PII covered in this book vary.

DATA EXHAUST

One more pernicious form of patron data bears calling out, and that is *data exhaust*.[23] As our digital systems universally grow more complex, many record every interaction an individual has with that system. For example, online retail sites record every product we view, building security systems track each keycard access and time, and university learning management systems log every activity a student does within a course. In libraries, proxy servers record every web page accessed by a patron with a time stamp and the ILS can track patrons' entire checkout history. Moving

through modern life leaves a trail of data exhaust within every system with which we interact. In thinking about where patron data exists within library systems in order to protect patron privacy, libraries must be cognizant that these digital bread crumbs create a link between an individual's identity and their activities.

As your library works to improve data practices, it will be important to consider privacy issues for both the patron data that the library consciously collects and the data exhaust that is simply created by our systems.

Case Studies

Throughout this book, we examine a case study that covers both academic and public libraries to show you how the discussed concepts apply in a library setting. The case study follows two friends, Leilani and Quinn, who went to library school together and are now meeting up at the ALA Annual Conference, where they coincidentally attend the same session on data and privacy. Leilani is a systems librarian for Malibu Beach Public Library and Quinn is the science librarian at Seattle State University. As the two catch up they realize that they're both dealing with data and privacy issues at their respective libraries. Leilani recently had to clean up after a data breach that included the loss of hundreds of library patrons' addresses and phone numbers and wants to prevent future breaches. Quinn just started working on a learning analytics project to correlate student outcomes with library usage and has privacy concerns about doing this type of analysis. After the session ends, the two continue discussing data and privacy over bad convention hall coffee and decide to keep in touch after the Annual Conference to share information and strategies on the topic. The rest of the book follows their developments.

How to Work Through This Book

Although reading this entire book will provide a deeper understanding of how to protect patron data and keep it private, this book can also be read modularly. Chapters were written to be fairly independent and provide a broader understanding of each specific topic without requiring significant background from previous chapters. The end-of-chapter case studies, however, build across chapters and may make more sense when read in order. We would certainly love for you to read and learn from the entire book, but we also recognize the value in diving into a specific chapter to make targeted improvements at your library. No matter how you choose to read the book, we hope that you are able to learn and implement the strategies within these pages to improve the privacy of patron data at your library.

NOTES

1. IBM Security, *2021 Cost of a Data Breach Report*, 2021, www.ibm.com/security/data-breach (registration required).
2. CBS SF Bay Area, "Cyber Attack Snarls CoCo County Library System," January 3, 2020, https://sanfrancisco.cbslocal.com/2020/01/03/cyber-attack-snarls-contra-costa-county-library-system/.
3. Dean Narciso, "Columbus Library Data Breach May Have Been Caused by Phishing Link," *The Columbus Dispatch*, January 17, 2020, www.dispatch.com/story/news/local/2020/01/17/columbus-library-data-breach-may/1871534007/.
4. "Library Patron Records Breached," *The Chronotype*, October 12, 2017, www.apg-wi.com/rice_lake_chronotype/free/library-patron-records-breached/article_07ba335a-af59-11e7-a018-530242c62ab7.html.
5. Joseph Geha, "Alameda County Library Still Doesn't Know How Many Patrons Were Hacked," *East Bay Times* (blog), December 1, 2017, www.eastbaytimes.com/2017/12/01/library-still-doesnt-know-how-how-many-patrons-were-hacked/.
6. Identity Theft Resource Center, *2019 End-of-Year Data Breach Report*, January 8, 2020, www.idtheftcenter.org/wp-content/uploads/2020/01/01.28.2020_ITRC_2019-End-of-Year-Data-Breach-Report_FINAL_Highres-Appendix.pdf.
7. The Proofpoint Threat Insight Team, "Threat Actor Profile: TA407, the Silent Librarian," *Proofpoint* (blog), October 9, 2019, www.proofpoint.com/us/threat-insight/post/threat-actor-profile-ta407-silent-librarian.
8. Jessica Ellis, "Silent Librarian: More to the Story of the Iranian Mabna Institute Indictment," *The PhishLabs Blog*, March 26, 2018, https://info.phishlabs.com/blog/silent-librarian-more-to-the-story-of-the-iranian-mabna-institute-indictment.
9. Identity Theft Resource Center, *2019 Data Breach Report*.
10. xxdesmus, "Kanopy.com Leaking API and Website Access Logs," Rainbowtabl.es, March 21, 2019, https://rainbowtabl.es/2019/03/21/kanopy-data-leak/.
11. Abeerah Hashim, "Elsevier Exposed User Credentials Publicly through Misconfigured Server," *Latest Hacking News*, March 25, 2019, https://latesthackingnews.com/2019/03/25/elsevier-exposed-user-credentials-publicly-through-misconfigured-server/.
12. Sean Gallagher, "Adobe's E-book Reader Sends Your Reading Logs Back to Adobe—in Plain Text [Updated]," *Ars Technica,* October 7, 2014, https://arstechnica.com/information-technology/2014/10/adobes-e-book-reader-sends-your-reading-logs-back-to-adobe-in-plain-text/.
13. ALA, "Interpretations of the Library Bill of Rights," Advocacy, Legislation and Issues, July 30, 2007, www.ala.org/advocacy/intfreedom/librarybill/interpretations.
14. Samuel D. Warren and Louis D. Brandeis, "The Right to Privacy," *Harvard Law Review* 4, no. 5 (1890): 193–220, https://doi.org/10.2307/1321160.
15. Helen Nissenbaum, *Privacy in Context: Technology, Policy, and the Integrity of Social Life* (Stanford, CA: Stanford Law Books, 2010).
16. Daniel J. Solove, "A Taxonomy of Privacy," *University of Pennsylvania Law Review* 154, no. 3 (January 1, 2006): 477, https://doi.org/10.2307/40041279.

17. Dorothea Salo, "Physical-Equivalent Privacy," *The Serials Librarian*, February 22, 2021, 1–15. https://doi.org/10.1080/0361526X.2021.1875962.
18. ALA, "Professional Ethics," Tools, Publications and Resources, May 19, 2017, www.ala.org/tools/ethics; IFLA, "Professional Codes of Ethics for Librarians," www.ifla.org/faife/professional-codes-of-ethics-for-librarians.
19. ALA, "Library Bill of Rights," Advocacy, Legislation and Issues, June 30, 2006, www.ala.org/advocacy/intfreedom/librarybill.
20. Erika McCallister, Tim Grance, and Karen Scarfone, *Guide to Protecting the Confidentiality of Personally Identifiable Information (PII),* NIST Special Publication 800-122 (Gaithersburg, MD: National Institute of Standards and Technology, 2010), https://doi.org/10.6028/NIST.SP.800-122.
21. "Art. 4 GDPR—Definitions," General Data Protection Regulation (GDPR) (website), https://gdpr-info.eu/art-4-gdpr/.
22. Latanya Sweeney, "Simple Demographics Often Identify People Uniquely" (Data Privacy Working Paper 3, Carnegie Mellon University, Pittsburgh, PA, 2000), https://dataprivacylab.org/projects/identifiability/paper1.pdf.
23. Shoshana Zuboff, *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power* (New York: PublicAffairs, Hachette Book Group, 2019).

Index

A

access and access controls, 29, 53–54, 92–94
accessibility, 125, 144
acronyms, limiting use of, 69
active learning, 143
addenda, contract, 111–112, 113
administrative controls, 88
Adobe Digital Editions, 2
aggregate data, 44, 45, 54–56, 115
ALA. *See* American Library Association (ALA)
Alameda County Library, California, 1
American Brook Lamprey, 4–5
American Library Association (ALA), 3, 18, 20–21, 141, 145
antivirus software, 94
API Security Project, 97
application programming interfaces (APIs), 97
assessments, library, 121–132
auditors, third-party, 114–115, 117–118
availability, 87

B

behavioral data, 7, 42, 43–44, 45
best practices, current, 20–21
Brandeis, Louis D., 4
business operation continuity planning, 116
buy-in, 149–151

C

California Consumer Privacy Act (CCPA), 15
cap values, 55
case studies
  on assessments, 130–132
  conference presentation based on, 155
  on data inventories, 38–40
  on data landscape, 21–23
  on information security, 101–102
  overview of, 8
  on policy and operations, 80–81
  on risk assessments, 58–60
  on training, 145–146
  on vendors, 117–118
CCPA. *See* California Consumer Privacy Act (CCPA)
checklists, 138
Children's Internet Protection Act (CIPA), 12–13
Children's Online Privacy Protection Act (COPPA), 13
Chisholm, Alex, 141
Choose Privacy Every Day website, 20, 141, 145, 146
Chronology of Data Breaches (Privacy Rights Clearinghouse), 98
CIPA. *See* Children's Internet Protection Act (CIPA)
Code of Ethics (CoE; ALA), 4, 18–19

collection stage, 28
Columbus Metropolitan Library, Ohio, 1, 85
communications
  e-mail, 76, 84–85
  issues with, 78–79
  policies and, 67–70
communities of practice, 21
compounding threats, 51
computer operating system image management systems, 94
confidentiality, 3, 87–88
confidentiality, integrity, and availability (CIA) of data, 87–88
consent, 51, 56, 80–81, 127, 128–129
contextual integrity, 4
Contra Costa County Library, California, 1
contract addenda, 111–112, 113
contracts, 108–115, 110–111*fig*, 112*fig*, 154–155
COPPA. *See* Children's Online Privacy Protection Act (COPPA)
corrective controls, 88
credentials theft, 2, 48
critical thinking, 123–126
culture of privacy, 149–156

D

data
  aggregate, 44, 45, 54–56, 115
  defining, 6–8
  de-identified, 44–45, 54–56, 115, 128, 130
  deleting, 50, 51, 52–53, 90–95, 115, 127, 129, 130, 153
  life cycle of, 28–30, 28*fig*, 63, 65
  minimizing collection of, 52
  personal, 7, 42–46. *see also* identifiers
  at rest, 90–95
  separating, 53
  sharing, 29, 50
  in transit, 48, 127, 128–129
  value of, 2
data agreements, 116
data breaches
  academic libraries and, 2
  cost of, 1
  as malicious threat, 46
  overview of, 1–3
  types of, 98
data exhaust, 7–8, 105
data governance, 50–51
data inventories
  case studies on, 35–38
  conducting, 30–35
  data life cycle and, 28–30
  introduction to, 27
  organizing information from, 34–35
  steps for optimization, 30
  updating, 153
  using, 35–38
data landscape, 11–23
data leaks, 2–3
data management, role of, 5
Data Privacy and Cybersecurity Training for Libraries project, 141
Data Privacy Project, 141
data processing agreements, 54
data protection impact assessments, 27, 35
data risk, 41–46, 45*fig*
data security, defining, 83
data set, growth of, 51
data spread, 51
database backups, 75
database encryption, 91
decision tree format, 75, 76*fig*
de-identified data, 44–45, 54–56, 115, 128, 130
deleting data, 50, 51, 52–53, 90–95, 115, 127, 129, 130, 153
delta backups, 95
design issues, 70–71
detective controls, 88
diceware, 92
dictionary attacks, 92
Digital Library Federation (DLF), 21, 145
Digital Shred, 141
direct identifiers, 7, 42, 45, 50, 54, 127
disability justice movement, 124–125
disclosure exemptions, 72
disclosure threats, 48–49

disposal stage, 29
DLF. *See* Digital Library Federation (DLF)
documentation, 72

E
editorial control of policy documents, 65
EDUCAUSE, 108, 111, 114
EFF. *See* Electronic Frontier Foundation (EFF)
Electronic Frontier Foundation (EFF), 92, 141, 145, 154
Elsevier, 2
e-mail, 76, 84–85
empathy maps, 151
encryption, 91
E-Rate discounts, 12–13
ethical breaches, 51
ethics, 4, 17–19
event-based training, 137
external trainers, 140

F
Fair Information Practice Principles (FIPPs), 63
Family Educational Rights and Privacy Act (FERPA), 13–14
Family Policy Compliance Office (FPCO), 13
FBI. *See* Federal Bureau of Investigation (FBI)
Federal Bureau of Investigation (FBI), 48
federal regulations, 12–14
feedback on training, 143–144
FERPA. *See* Family Educational Rights and Privacy Act (FERPA)
file-level encryption, 91
filtering software, 12–13
FIPPs. *See* Fair Information Practice Principles (FIPPs)
Five Whys technique, 123–124
FOI laws. *See* freedom-of-information (FOI) laws
FOIA. *See* Freedom of Information Act (FOIA)
folder-level encryption, 91
FPCO. *See* Family Policy Compliance Office (FPCO)
FPF. *See* Future of Privacy Forum (FPF)
Freedom of Information Act (FOIA), 16
freedom-of-information (FOI) laws, 16
full disk encryption, 91
functional requirements, 107–108
Future of Privacy Forum (FPF), 145

G
GDPR. *See* General Data Protection Regulation (GDPR)
General Data Protection Regulation (GDPR), 7, 16, 35, 154
generalizing data, 55
governance threats, 50–51
group permissions, 93
guidelines, 72

H
hardware end-of-life, 53
harm reduction, 124–125
Hartman-Caverly, Sarah, 141
hash algorithms, 54–55
Health Information Technology for Economic and Clinical Health Act (HITECH), 12
Health Insurance Portability and Accountability Act of 1996 (HIPAA), 12
Higher Education Community Vendor Assessment Toolkit (HECVAT), 114
hypertext transfer protocol secure (HTTPS), 96–97

I
IAPP. *See* International Association of Privacy Professionals (IAPP)
identifiers
  direct, 7, 42, 45, 50, 54, 127
  indirect, 7, 42–43, 45, 50, 54, 130
"IFLA Code of Ethics for Librarians and Other Information Workers," 19
"IFLA Statement on Privacy in the Library Environment," 19

IMLS. *See* Institute of Museum and Library Services (IMLS)
immigration status, 48
incident response, 98–100, 100*t*
inclusion, 124–125
Indiana University Libraries Privacy Policy, 65
Indianhead Federated Library System, Wisconsin, 1
indirect identifiers, 7, 42–43, 45, 50, 54, 130
information privacy, 3–4
information security
  basics of, 87–89
  case studies on, 101–102
  incident response and, 98–100
  policies and procedures involving, 77–78
  resources on, 102–103
  at rest and in use, 90–95
  role of, 83
  threats and, 84–87
  in transit, 96–97
insider threats, 46–47
Institute of Museum and Library Services (IMLS), 21
institutional review boards (IRBs), 128
integrity, 87
intellectual property, 88
internal access, unauthorized, 46–47
International Association of Privacy Professionals (IAPP), 99, 145, 154
International Federation of Library Associations and Institutions (IFLA), 4, 19
International Organization for Standardization (ISO), 89
internet safety policies, 12–13
"Interpretation of the Library Bill of Rights, An" (ALA), 3
IRBs. *See* institutional review boards (IRBs)
ISO. *See* International Organization for Standardization (ISO)
ISO/IEC 27001, 89
ISO/IEC 27035 series, 99

J
jargon, limiting use of, 69
just-in-time notices, 71

K
k-anonymity, 54
Kanopy, 2
key distribution, 94

L
law enforcement requests, 72–75, 76*fig*, 79
layered format for notices, 71
learning analytics, 3, 116
legal disclosure, 48, 49
Let's Encrypt, 96
LFI. *See* Library Freedom Institute (LFI)
LFP. *See* Library Freedom Project (LFP)
LIBLICENSE, 111
Library Bill of Rights (ALA), 4–5, 18
library data privacy operations, 61–62
Library Freedom Institute (LFI), 21, 145
Library Freedom Project (LFP), 21
Library Privacy Checklists, 20
Library Privacy Field Guides, 20
Library Privacy Guidelines, 20
library services platform (LSP), 53
LinkedIn Learning, 113–114
local regulations, 15
locks and keycards, 94
LSP. *See* library services platform (LSP)

M
malicious threats, 46–48
malware, 84, 94, 154
marketing segmentation research, 3
MFA. *See* multifactor authentication (MFA)
minimization, 127–128
Minimum Security Standards tables (Stanford), 78
minoritized groups, 55, 123, 124–125, 129, 131
mitigation plans, 57–58
mobile devices, staff, 94–95
multifactor authentication (MFA), 93

N
National Information Standards Organization (NISO), 19–20
National Institute of Standards and Technology (NIST), 6
NDAs. *See* nondisclosure agreements (NDAs)
negotiations with vendors, 108–109
NISO. *See* National Information Standards Organization (NISO)
NISO Privacy Principles, 19–20
Nissenbaum, Helen, 4
NIST. *See* National Institute of Standards and Technology (NIST)
NIST Cybersecurity Framework, 89
NIST Special Publication 800-61r2, 99
nondisclosure agreements (NDAs), 112–113
NYC Digital Safety, 141

O
OECD. *See* Organisation for Economic Co-operation and Development (OECD) Privacy Principles
onboarding training, 137
online training, 142–143
open data mandates, 116
open records laws, 17
Open Web Application Security Project (OWASP), 97
Organisation for Economic Co-operation and Development (OECD) Privacy Principles, 63
organizational culture, 151
outliers, 55, 128, 130
OWASP. *See* Open Web Application Security Project (OWASP)

P
parameters of data inventory, setting, 31
parent organization policies and regulations, 17
passive learning, 143
passphrases, 92
password managers, 48, 92
passwords, 92
personal data
  categories of, 42–45
  definition of, 7
  risk level of, 45–46
  See also identifiers
Personal Information Protection and Electronic Documents Act (PIPEDA), 12
personally identifiable information (PII), 6–7, 42
PET. *See* Privacy and Ethics in Technology (PET) Working Group
phishing, 84–85
physical controls, 88, 93–94
physical-equivalent privacy, 4
PIA. *See* privacy impact assessments (PIA)
PIA reports, 37–38
PII. *See* personally identifiable information (PII)
PIPEDA. *See* Personal Information Protection and Electronic Documents Act (PIPEDA)
planning stage, 28
policies
  accessibility, 125
  case studies on, 80–81
  for data storage, 95
  definition of, 62
  privacy and confidentiality, 63–65, 64*t*
  privacy notices, 66–67, 68*t*
  privacy-adjacent, 65–66
  setting, 56
  training on, 138–139
  updating, 154–155
  writing and communicating about, 67–71
postmortems, 79–80
practice, 62, 78–80
preventative controls, 88
privacy
  definitions of, 3–4
  overview of, 3–5
  value of, 2
privacy and confidentiality policy, 63–65, 64*t*

Privacy and Ethics in Technology (PET) Working Group, 21
privacy audits, 27, 38, 153
Privacy by Design framework, 35, 126
privacy impact assessments (PIA), 27, 35–38
privacy notices, 63, 66–67, 68*t*
Privacy Rights Clearinghouse, 98
privacy-adjacent policies, 65–66
procedures, 62, 71–78, 138–139, 154–155
procurement practices, 106–108
pseudonyms, 54–55
public records regulations, 16

Q
qualitative assessments, 125–126
quantitative assessments, 125–126
quick access guides, 71

R
randomization, 54
ransomware, 84, 154
reboot-to-restore systems, 94
red flags in contracts, 111, 112*fig*
refresher training, 137
regulations
  changes in, 154
  contracts and, 155
  on data breaches, 99
  local regulations, 15
  other notable, 16–17
  policies and, 63
  privacy notices and, 67
  state, 14–15
  US federal, 12–14
re-identification threats/risk, 49–50, 127
renewals, contract, 113–114
requests for information (RFI), 106–107
requests for proposals (RFP), 106–108
resource gaps, 152
REST Security cheat sheet, 97
restricting access, 53–54
retention schedules, 15, 50, 52, 75–77, 127
RFI. *See* requests for information (RFI)
RFP. *See* requests for proposals (RFP)
risk assessments
  case studies on, 58–60
  data risk and, 41–46
  introduction to, 41
  putting together, 57–58
  risk mitigation and, 52–56
  threat types and, 46–51
risk mitigation, 52–58

S
Safe Data | Safe Families, 141
Salo, Dorothea, 4
Schneier, Bruce, 29
scope creep, 50–51, 127
Section 215 of PATRIOT Act, 14
Section 6267, Government Code (California), 14–15, 22
sectoral approach, 12, 67
security, enhancing, 53
Security Education Companion (EFF), 141
separating data, 53
separation from vendors, 115–116
Server Security Checklist (RIT), 78
sharing data, 29, 50
Silent Librarian phishing campaign, 2, 85
SIP. *See* Standard Interchange Protocol (SIP)
small numbers, 55, 128, 129, 130
social engineering, 85–86
Solove, Daniel, 4
spear phishing, 85
specificity, reducing, 128
SSH tunnels, 97
staff
  communications with, 69–70
  computer security and, 94–95
  e-mail and, 76
  training for, 56, 70, 75, 135–147, 153
  turnover of, 51
stakeholders
  data privacy programs and, 62
  identifying, 32–33
  interviewing, 33–34
stalking, 46–47

Standard Interchange Protocol (SIP), 97
state regulations, 14–15
storage, 29, 90–95, 127
subpopulations, 55
summarizing data, 55–56
sunshine laws, 16–17
system backups, 95
system data migrations, 51

T
technical controls, 88
technical threats, 49
technology
  privacy and, 66, 77–78
  rapid evolution of, 79
telecommuting, 66, 97
third-party services, 66, 105, 116–117. *See also* vendors
threat actors, 86–87
threat modeling, 86–87
threat types, 46–51, 47*t*
threats
  common, 84–86
  modeling for, 86–87
  to patron data, 84–87
threshold assessments, 36–37
training, 56, 70, 75, 129–130, 135–147, 153
transformation stage, 29
transit, data in, 48, 127, 128–129
translating documents, 69
transparency, 51, 58, 127
trust, 2

U
unauthorized disclosure, 49
updated data inventories, 153
URLs, 96–97
US federal regulations, 12–14
USA Freedom Act, 14
USA FREEDOM Reauthorization Act, 14
USA PATRIOT Act, 14, 48
use stage, 29
user roles, 93

V
vendors
  assessments of, 114–115
  case studies on, 117–118
  contracts with, 108–115, 154–155
  incident response and, 99–100
  limited choices for, 105
  overview of, 105–106
  phishing and, 85
  privacy notices from, 70
  selecting, 106–108
  separation from, 115–116
virtual private networks (VPNs), 97
Voluntary Product Accessibility Templates (VPATS), 114
volunteers, training for, 137–138
VPATS. *See* Voluntary Product Accessibility Templates (VPATS)
VPNs. *See* virtual private networks (VPNs

W
Warren, Samuel D., 4
whaling, 85
Wireshark, 96
workstations, 94–95
writing and communicating about, 67–71

Y
Yelton, Andromeda, 96