NNIP'S RESOURCE GUIDE TO DATA GOVERNANCE AND SECURITY

SEPTEMBER 2018

Leah Hendey
Amanda Gold
Kathryn L.S. Pettit

Version 1.0

NATIONAL NEIGHBORHOOD INDICATORS PARTNERSHIP

ACKNOWLEDGMENTS

The authors would like to thank several Urban Institute staff members for their contributions to this guide: Ilana Brody and Matt Gerken for their assistance summarizing resources; Martha Galvez, Matthew Eldridge, and Rayanne Hawkins for their contributions and guidance on resources related to pay for success projects. The authors especially acknowledge, Madeline Brown, also at the Urban Institute, who provided valuable assistance organizing and deidentifying data-sharing agreements from NNIP partners. Thank you to all the NNIP partners who generously shared their resources and advice with us and the network, particularly Susan Millea of Children's Optimal Health and Amy Hawn Nelson at Actionable Intelligence for Social Policy and formerly at the University of North Carolina at Charlotte's Institute for Social Capital. Finally, thank you to Chris Kingsley at the Annie E. Casey foundation for his helpful review of the guide.

This guide was funded by the Annie E. Casey Foundation. We thank them for their support but acknowledge that the advice and conclusions presented in this report are those of the authors alone, and do not necessarily reflect the opinions of the Urban Institute or the Foundation.

ACRONYM LIST

AISP – Actionable Intelligence for Social Policy

API – application programming interface

CFR – Code of Federal Regulations

DCC – Digital Curation Centre

ED – US Department of Education

eMOU – enterprise memorandum of understanding

ETL – extract, transform, load

FERPA – Family Educational Rights and Privacy Act

FOIA – Freedom of Information Act

GIS – geographic information system

HHS – US Department of Health and Human Services

HIPAA – Health Insurance Portability and Accountability Act

HITECH – Health Information Technology for Economic and Clinical Health Act

ICPSR – Inter-university Consortium for Political and Social Research

IDS – integrated data system

IT – information technology

IRB – institutional review board

MOU – memorandum of understanding

NNIP – National Neighborhood Indicators Partnership

PHI – protected health information

PII – personally identifiable information

PTAC – Privacy Technical Assistance Center at the US Department of Education

SAMHSA – Substance Abuse and Mental Health Services Administration

Table of Contents

Chapter 1. Data Governance 2

Who should use this Guide? 2

What Is Data Goverance? 3

Who Is Accountable? 4

Chapter 2. Protecting Privacy and human subjects 7

Privacy versus confidentiality 7

Privacy Laws and Regulations 7

FERPA 9

HIPAA 10

Human Subjects' Protection 14

Chapter 3. Data Security Plan 18

Data Use and Handling Procedures 19

Secure Transfer 19

Access Control 20

Storage 20

Vulnerability Assessment and Risk Management 22

Data Security Audits 23

Threats 23

Data Security Incidents and Breaches 24

Procedures 25

Liability Insurance 26

Monitoring Compliance 27

Staff Training 27
Training Data Partners 28
Access Audits and Procedures 29

Chapter 4. Data Life cycle 31

Data Acquisition 32
Public Websites 32
Freedom of Information Requests 33
Negotiated Data Sharing 34
Elements of the data use agreement 35
Data Processing 37
Data Inventory 38
Metadata 38
Transformation 39
Data Quality 40
Linking 40
Data dissemination 42
Establishing Guidelines for Dissemination 42
Methods to Protect against Reidentification 43
Data Licenses 45
Data Disposition 45

Bibliography 47

Version updates 48

CHAPTER 1. DATA GOVERNANCE

Local data intermediaries in the National Neighborhood Indicators Partnership (NNIP) have a shared mission to help local stakeholders use neighborhood data for better decisionmaking and community building. To accomplish this mission, NNIP partners create useful indicators at relevant geographies for their communities from surveys and administrative data sources. Building and maintaining a reputation for handling data carefully, protecting confidentiality, and interpreting data thoughtfully and accurately are essential to ensuring a data intermediary's long-term sustainability.

This guide aims to provide resources and advice from the experiences of those in the NNIP network and other related organizations on developing a strong data governance program and protecting the security of confidential data.

Because data science and technology are rapidly evolving and new resources appear frequently, we will update this guide periodically. If you have additional helpful resources to include or other significant comments, please send them to nnip@urban.org.

WHO SHOULD USE THIS GUIDE?

Though developed for the organizations participating in NNIP, known as data intermediaries, this guide is useful to any organization that uses secondary data, especially those who have sensitive or confidential information about people or businesses. The rest of this chapter explains data governance processes, roles and accountability, and who is accountable at an organization for data governance. The remaining chapters cover three important parts of a data governance program: protecting privacy and human subjects, data security, and the data life cycle.

In each chapter, we provide introductory comments that highlight key resources or experiences from NNIP and then provide a curated set of resources on the topic. We tried to focus the resource descriptions on the elements most relevant for the topic. Some resources are cited several times in different topic areas. Throughout, we have highlighted resources that might be relevant to organizations that need to share individual-level data across sectors, such as for integrated data systems (IDS) or for pay for success projects (see the callout boxes).

This guide does not provide guidance for organizations on how to collect primary data, though it does have information on human subjects' protections, and the need for good data security and data management processes is still relevant.

PAY FOR SUCCESS

Pay for success (PFS) projects tie payments for services to the achievement of predetermined, measurable outcomes. These outcomes are typically measured using administrative data that are often housed in different government agencies. This guide is useful for jurisdictions considering PFS because it offers advice on how to create a data governance plan that will enable the data sharing needed for measuring the outcomes necessary to determine payment in PFS projects. For more information, visit the Urban Institute's Pay for Success Initiative website at https://pfs.urban.org/.

DEFINING INTEGRATED DATA SYSTEMS

An integrated data system (IDS) links individual-level administrative records from multiple sources on a periodic basis. IDS can operate at the city, county, or state level. An IDS might link data on education, juvenile justice, child welfare, and social assistance. An IDS can be used for policy analysis, program planning, and evaluation. For more information on and resources related to IDS, visit the NNIP website at https://www.neighborhoodindicators.org/issue-area/292 and the Actionable Intelligence for Social Policy website at https://www.aisp.upenn.edu/.

WHAT IS DATA GOVERANCE?

The US Department of Education's (ED) Privacy Technical Assistance Center (PTAC) defines data governance as ***an approach to data management, which is formalized into a set of policies and procedures that address the full life cycle of data, from acquisition to disposal*** (PTAC 2015a). Having formal data governance is recommended for all data intermediaries, even if your organization is working only with public records and national survey data. Data governance ensures the quality, accuracy, and usability of data, in addition to being critical for ensuring ethical use, protecting privacy, and confidentiality.

In this guide, we have broken out what should be included as part of a data governance program into three parts:

1. Protecting privacy and human subjects
2. Implementing a data security plan

3. Establishing policies and processes for data governance at all stages of the data life cycle: acquisition, processing, dissemination, and destruction

Protecting privacy and human subjects involves protecting data with personally identifiable information (PII) or protected health information (PHI under the Health Insurance Portability and Accountability Act, or HIPAA) and ensuring compliance with the terms of data-sharing agreements and local, state, and national privacy laws that govern access and use of data. When appropriate, it also involves taking additional precautions to minimize risks and protect vulnerable populations and their sensitive information when they are the subject of research. Stiles and Boothroyd (2015) provide an overview of ethical issues and considerations, including discussions of appropriate use, associated with administrative data.

PERSONALLY IDENTIFIABLE INFORMATION:

Data that alone or in combination with other information can be used to identify or trace a specific person.

Establishing data security is a prerequisite to acquiring confidential data and must be maintained to protect against misuse and unauthorized data breaches. A data security plan defines who has access to data, how data will be held confidential and secured during transfer and storage, how vulnerability to attacks will be assessed, and what protocols are established in the event of data misuse or breach, and to monitor compliance and train staff.

The plan should define and document the policies and procedures that guide a data governance program so that staff can comply with them. PTAC recommends that procedures be proactive (managing all foreseeable activities related to data governance), ongoing (providing sustained guidance and support), and reactive (generating additional guidance around areas of confusion or as new challenges arise) (PTAC 2015a).

WHO IS ACCOUNTABLE?

Everyone in an organization shares responsibility for data governance, including users, managers, data owners, and executives. Defining clear lines of authority with respect to roles for data protection and accountability within an organization, and between entities when data are shared, is important. Good data governance requires identifying who holds decisionmaking authority and accountability. This authority should oversee the design and implementation of data governance and compliance with the plan. Checks and balances and independence across roles are also important. For instance, a data security officer should not be supervised by a data user. Clear lines of accountability help build trust for data sharing. Several factors will influence how your organization sets up this authority, including

- the size of your organization (staff and funding),
- organizational form (e.g., a nonprofit or a university),
- the complexity of the data and data sharing (e.g., a single source of confidential data or data sharing across a dozen partners for a Promise Neighborhood), and
- the population's vulnerability (that might signal a need for an advisory committee to govern how data are used).

If your organization is small or handles only a few confidential data sources, you might decide that a single decisionmaker is sufficient for good data governance. Your organization should designate a data manager or "chief data security officer." This person is internally accountable for data governance and can be the source for questions from outside your organization about data use and confidentiality (Strive Together 2015, 9). Even if a single person holds the authority, PTAC suggests identifying data stakeholders—data owners and users—who can provide feedback on governance goals. Soliciting their feedback early on can also establish support for data governance and sharing moving forward. For example, a data intermediary, through its chief data security officer, could build a relationship with the school district and develop a data use agreement to obtain student records to help them map where students live and estimate future enrollment patterns. This chief data security officer creates a point of access for the school district and for parents to ask questions about how the data are used and secured. This person would be responsible for establishing safeguards for the data and responding to incidents of misuse.

If your organization is large, handles highly sensitive data, or does complex data sharing and linking, in addition to data governance roles people in your organization play, you might establish a data governance committee.

Responsibility for good data governance is shared and requires the participation of everyone in the organization that manages data. In addition to the data manager or data governance committee, data stewards can implement the data governance program and oversee management of one or more data sources. Other people, such as analysts, researchers, and users, are also responsible for learning data governance and security procedures and following established protocols.

The remainder of this guide outlines and describes resources for the elements of a good data governance program.

DATA GOVERNANCE COMMITTEE

A data governance committee is a good idea if you are hosting integrated data systems (IDS) or linking records across several organizations, such as providing backbone support for a Promise Neighborhood or collective impact initiative. Such committees should contain members who represent the data source owners but can also include funders, community organizations, legal counsel, and privacy and security experts, as well as the data system administrator, data managers, and users. The committee should be formalized with an agreement that defines the mission, goals for data sharing, and who the data sharing should benefit. Gibbs et al. (2017) is an excellent resource for thinking about governance committees and stakeholders, both generally and for IDS. Such a committee can drive the agenda for an IDS—what additional data sources to pursue to bring into the system, which policy or program questions are most urgent to use the data to answer, and who can access or request deidentified or aggregated data from the IDS.

CHAPTER 2. PROTECTING PRIVACY AND HUMAN SUBJECTS

Survey and administrative data sources might contain sensitive information about people and their families. Safeguards need to be in place when these data are shared to protect people's privacy, provide confidentiality, and ensure that any research project has protocols to protect people from harm. [i] Statues and regulations at all levels of government protect people's privacy and have implications for how data are used and shared.

Protecting people's or organizations' privacy does not mean data cannot be shared. To protect privacy, data providers and users must ensure the data can be kept secure and released without risk of reidentification. Many of the regulations discussed in this chapter are meant to provide a clear path for securing data and protecting privacy. These regulations reduce the fear of and risks associated with data sharing and can increase appropriate access to the data.

PRIVACY VERSUS CONFIDENTIALITY

Privacy and confidentiality are not the same thing, though they are commonly used interchangeably. *Privacy* relates to an individual, and in the context of data, it is a person's freedom to choose how and under what circumstances information about themselves can be shared. *Confidentiality* has to do with protecting information. It refers to the promise the data collector makes to a data provider about what information about the data provider will be shared. Confidentiality can be applied to data about people or about organizations.

Besides privacy protections, there is also a need to ensure that people are protected and not harmed, even through the use of secondary data. Though protection of human subjects in research arose largely in response to atrocities in clinical research, protection is still relevant for research and analyses conducted for social science purposes. Research using survey or administrative data carries risks to people if data are improperly handled and information disclosed. Government agencies, universities, and other research institutions established institutional review boards (IRBs) to govern such research.

This chapter is divided into two sections: resources on privacy laws and regulations and resources on protecting human subjects and IRBs. Chapter 3 describes resources on data security plans, key mechanisms for complying with these laws, and protections.

PRIVACY LAWS AND REGULATIONS

If your organization is using data covered by privacy laws and regulations, your data governance procedures must comply with these laws. The laws have provisions that define what data are protected, what data are PII or PHI, who can access the data, how people consent to release data, what data can be shared, and how data must be secured. Often, data providers

or their lawyers mistakenly claim that these laws forbid sharing data with external organizations. The laws do not forbid data sharing but define how to do so in ways that protect privacy and confidentiality. There are many examples of responsible data sharing that comply with the terms and conditions in the regulations mentioned below.

Among the most commonly known federal privacy laws are the Family Educational Rights and Privacy Act (FERPA), the Health Insurance Portability and Accountability Act (HIPAA), and the Health Information Technology for Economic and Clinical Health Act (HITECH). They outline federal protections and disclosure requirements for student data (FERPA), for patient data (HIPAA), and for electronic health data security (HITECH). Other federal laws and regulations cover other types of data, including federal administrative data (the Privacy Act of 1974), alcohol and substance abuse treatment records, homeless management information systems, and child abuse records. State laws usually address privacy issues related to criminal and juvenile justice records, mental health records, and data on HIV/AIDS status. Federal laws like FERPA and HIPAA permit states to adopt tougher standards; in those cases, the state standards would apply.

Below are several online resources that provide guidance and training on FERPA and HIPAA. The following two resources are excellent general ones on several common data sources, especially for those interested in building IDS.

Legal Issues in the Use of Electronic Data Systems for Social Science Research
John Petrila. (2015).

In addition to FERPA and HIPAA, this paper, published by Actionable Intelligence for Social Policy (AISP), covers the Privacy Act of 1974, federal regulations governing the confidentiality of alcohol and substance abuse treatment records, homeless management information systems, the Child Abuse Prevention and Treatment Act, and information about institutional review boards.

https://www.aisp.upenn.edu/wp-content/uploads/2015/09/0033_12_SP2_Legal_Issues_Data_Systems_000.pdf

Legal Issues for IDS Use: Finding a Way Forward
John Petrila, Barbara Cohn, Wendell Pritchett, Paul Stiles, Victoria Stodden, Jeffrey Vagle, Mark Humowiecki, and Natassia Rozario. (2017).

This expert panel report, published by AISP, provides information and resources for agencies and organizations looking to build a legal framework integrating data across administrative sources. It contains information on common legal concerns, elements of data-sharing agreements, specific laws that apply, and draft templates and sample memoranda of understanding.

https://www.aisp.upenn.edu/wp-content/uploads/2016/07/Legal-Issues.pdf

FERPA

FERPA applies to a student's education records that schools or local education agencies maintain. FERPA specifies for parents the right to consent to the disclosure of student records (except as provided by law), the right to access their children's records, and the right to request that records be amended. Records covered by the law include information about academic performance and attendance, as well as identifying information about students and their families, such as the student's name, home address, and other personal and indirect identifiers. FERPA does permit data sharing and has specific exemptions from the required written consent. The Protecting Student Privacy website (https://studentprivacy.ed.gov), a service of PTAC and the Family Policy Compliance Office at ED, have numerous guidance documents, videos, webinars, and trainings on FERPA, data sharing, and related privacy issues. We have listed a few of them here.

FERPA AND INTEGRATED DATA SYSTEMS

PTAC prepared detailed guidance on how educational authorities can comply with FERPA and share data with an integrated data system in "**Integrated Data Systems and Student Privacy**." Individual student records with PII can be shared only with the IDS lead agency with parental consent or the data sharing meets one of the FERPA exceptions to consent. The data governance framework for the IDS should meet all FERPA's legal requirements. The second link below is for a webinar that accompanied the guidance.

https://studentprivacy.ed.gov/sites/default/files/resource_document/file/IDS-Final_0.pdf
https://studentprivacy.ed.gov/training/integrated-data-systems-and-student-privacy

Training Modules on FERPA
Privacy Technical Assistance Center. (n.d.).

There are several training modules regarding FERPA and data sharing on the Protecting Student Privacy website, including "FERPA 101: For Local Education Agencies," "FERPA 101: For Colleges and Universities," and "FERPA 201: Data Sharing under FERPA." There are also webinars and videos on best practices in data security, data-sharing agreements, and disclosure elsewhere on the site.

https://studentprivacy.ed.gov/content/online-training-modules

FERPA Exceptions Summary
Privacy Technical Assistance Center. (2014).

This two-page handout summarizes the four main exceptions to the requirement of having written parental consent for data sharing: directory information, school official, studies, and audit or evaluation.

https://studentprivacy.ed.gov/resources/ferpa-exceptions-summary-apr-2014-2-page-standard-size

Measuring Performance: A Guidance Document for Promise Neighborhoods on Collecting Data and Reporting Results

Jennifer Comey, Peter A. Tatian, Lesley Freiman, Mary K. Winkler, Christopher R. Hayes, Kaitlin Franks, and Reed Jordan. (2013).

The Urban Institute's guidance for Promise Neighborhoods summarizes FERPA and HIPAA requirements in appendix 6.1. It discusses the type of data protected and entities covered by FERPA and the scope of protections, with a focus on situations in which FERPA restrictions are most likely to come up for Promise Neighborhoods, including (1) instances where school districts want to share personal and educational information on students with Promise Neighborhoods and (2) instances where Promise Neighborhoods need to report individual-level data to third parties, such as external evaluators or ED.

https://www.urban.org/research/publication/measuring-performance-guidance-document-promise-neighborhoods-collecting-data-and-reporting-results

Joint Guidance on the Application of the Family Educational Rights and Privacy Act (FERPA) and the Health Insurance Portability and Accountability Act of 1996 (HIPAA) to Student Health Records

US Department of Health and Human Services and US Department of Education. (2008).

This document addresses confusion surrounding the parts of student records covered by FERPA and covered by HIPAA. It gives an overview of potential areas of overlap and answers to 16 frequently asked questions, including, "Does FERPA or HIPAA apply to records on students at health clinics run by postsecondary institutions?" and "How does FERPA apply to health records on students maintained by elementary or secondary schools?"

https://www2.ed.gov/policy/gen/guid/fpco/doc/ferpa-hipaa-guidance.pdf

Legal Issues for IDS Use: Finding a Way Forward

John Petrila, Barbara Cohn, Wendell Pritchett, Paul Stiles, Victoria Stodden, Jeffrey Vagle, Mark Humowiecki, and Natassia Rozario. (2017).

This report, published by AISP, provides a concise and up-to-date overview of FERPA, highlighting a recent ED rule expanding access to student data for research. The document also identifies seven additional resources for FERPA, including guidance published by ED, The National Center for Education Statistics, and PTAC.

https://www.aisp.upenn.edu/wp-content/uploads/2016/07/Legal-Issues.pdf

HIPAA

HIPAA protections cover health plans, health care clearinghouses, and health care providers, as well as their business associates, who transmit health information *in electronic form* in connection with a covered transaction. Information held by a covered entity or its business associate is protected under HIPAA and is referred to as protected health information, or PHI. HIPAA has been operationalized into regulations that cover both privacy ("HIPAA Privacy Rule") and security ("HIPAA Security Rule").

There are several important terms to know related to HIPAA:

> ***Protected heath information.*** *(PHI)* Individually identifiable health information, including demographic information, which relates to these factors:
>
> - The person's past, present, or future physical or mental health or condition.

- The provision of health care to the person.
- The past, present, or future payment for the provision of health care to the person and information that identifies the person or for which there is a reasonable basis to believe it can be used to identify the person. Protected health information includes many common identifiers when they can be associated with the health information listed above.

https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html#protected

Covered entity. Health care providers, such as doctors, clinics, and pharmacies; health plans such as insurance companies; and health care clearinghouses. The US Department of Health and Human Services (HHS) website also has a decision tool to figure out if your organization is a covered entity. Some organizations, such as public health departments, are hybrid entities with some functions that are covered by HIPAA and some that are not covered and has designated them as such.

https://www.hhs.gov/hipaa/for-professionals/covered-entities/index.html

Business associate. A person or entity that performs certain functions or activities that involve the use or disclosure of protected health information on behalf of, or provides services to, a covered entity but is not an employee of the entity.

https://www.hhs.gov/hipaa/for-professionals/privacy/guidance/business-associates/index.html

The HHS website offers detailed information on HIPAA and training resources. State laws might have tighter standards than the federal regulations. For example, in Texas, the definition of covered entity is more expansive and specifies ongoing employee training and increased penalties.[ii]

Summary of the HIPAA Privacy Rule
US Department of Health and Human Services. (2013).

HHS summarizes the key elements of the HIPAA Privacy Rule on its website. The guidance is extensive and details the type of information protected, uses and disclosures of data, and the required safeguards for protecting electronic health information. The website provides links to relevant regulations and resources, including a tool for determining who is covered by HIPAA, published by the Centers for Medicare and Medicaid Services.

https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html

Summary of the HIPAA Security Rule
US Department of Health and Human Services. (2013).

HHS summarizes the key elements of the HIPAA Security Rule on its website. The security rule "established a national set of security standards for protecting certain health information that is held

or transferred in electronic form." It covers both technical and nontechnical safeguards that covered entities must use to secure protected health information.

https://www.hhs.gov/hipaa/for-professionals/security/laws-regulations/index.html

HIPAA Security Rule and HITECH
US Department of Health and Human Services. (2013).

The HITECH Act now applies certain HIPAA provisions directly to business associates. Formerly, privacy and security requirements were imposed on business associates via contractual agreements with covered entities. Under the HITECH Act, business associates are now required to comply with the safeguards contained in the HIPAA Security Rule.

https://www.hhs.gov/hipaa/for-professionals/security/laws-regulations/index.html
Additional guidance for business associates:
https://www.hhs.gov/hipaa/for-professionals/privacy/guidance/business-associates/index.html

Training Materials
US Department of Health and Human Services. (2018).

The HHS website has links to a training from HealthIT.gov and for state attorneys general.

https://www.hhs.gov/hipaa/for-professionals/training/index.html

Additional HIPAA resources:

Protecting Personal Health Information in Research: Understanding the HIPAA Privacy Rule
US Department of Health and Human Services, National Institutes of Health. (2003).

This document, designed for researchers, covers the purpose of HIPAA and its importance to researchers, how it interacts with other privacy protections, who the rule applies to, and the type of health information that is protected.

https://privacyruleandresearch.nih.gov/pdf/hipaa_privacy_rule_booklet.pdf

Health Services Research and the HIPAA Privacy Rule
US Department of Health and Human Services, National Institutes of Health. (2005).

HHS's overview of HIPAA discusses how the Privacy Rule allows certain health care providers, health plans, and other entities covered by the Privacy Rule to use and disclose personal health information for health services research. The document includes commonly asked questions and answers about the Privacy Rule and health services research. It also includes links to additional references.

https://privacyruleandresearch.nih.gov/pdf/healthservicesresearchhipaaprivacyrule.pdf

Health Information Data Sharing: HIPAA Facts and Fallacies
The Network for Public Health Law. (2017).

This webinar, designed for public health practitioners, aims to explain how data can be shared under HIPAA and covers HIPAA basics, such as what it does and what it covers, as well as information about breaches and enforcement.

https://www.networkforphl.org/webinars/2017/08/30/914/health_information_data_sharing_hipaa_facts_and_fallacies

Data Sharing within Cross-Sector Collaborations: Challenges and Opportunities
The BUILD Health Challenge. (2018).

This report summarizes challenges and opportunities to data sharing for cross-sector initiatives involving health data. It addresses challenges to sharing data covered by HIPAA and provides best practices and advice from BUILD partner sites to facilitate data sharing.

http://bit.ly/DataSharingReport

Measuring Performance: A Guidance Document for Promise Neighborhoods on Collecting Data and Reporting Results
Jennifer Comey, Peter A. Tatian, Lesley Freiman, Mary K. Winkler, Christopher R. Hayes, Kaitlin Franks, and Reed Jordan. (2013).

The Urban Institute's guide summarizes HIPAA requirements in appendix 6.1. The authors briefly discuss the type of data protected and entities covered by HIPAA and the scope of protections. This guide also has a concise summary of the key differences between FERPA and HIPAA, including differences in covered institutions or entities, protected information, exceptions, PII, deidentified records, data sharing without consent or authorization, and consent or authorization requirements.

https://www.urban.org/research/publication/measuring-performance-guidance-document-promise-neighborhoods-collecting-data-and-reporting-results

Joint Guidance on the Application of the Family Educational Rights and Privacy Act (FERPA) and the Health Insurance Portability and Accountability Act of 1996 (HIPAA) to Student Health Records
US Department of Health and Human Services and US Department of Education. (2008).

This document addresses confusion surrounding the parts of student records covered by FERPA and covered by HIPAA. In addition to providing an overview on potential areas of overlap, the document also provides answers to 16 frequently asked questions, including, "Does the HIPAA Privacy Rule apply to an elementary or secondary school?" and "Can a postsecondary institution be a 'hybrid entity' under the HIPAA Privacy Rule?"

https://www2.ed.gov/policy/gen/guid/fpco/doc/ferpa-hipaa-guidance.pdf

Legal Issues for IDS Use: Finding a Way Forward
John Petrila, Barbara Cohn, Wendell Pritchett, Paul Stiles, Victoria Stodden, Jeffrey Vagle, Mark Humowiecki, and Natassia Rozario. (2017).

This report, published by AISP, provides a concise and up-to-date overview of HIPAA, highlighting the three key points pertaining to the minimum standard of protection established by HIPAA, covered entities, and considerations surrounding the importance of identifying individuals in an IDS. The document identifies nine additional resources for HIPAA, including guidance published by HHS, National Institutes of Health, Centers for Medicare and Medicaid Services, North Carolina's Department of Health and Human Services, and the University of Buffalo.

https://www.aisp.upenn.edu/wp-content/uploads/2016/07/Legal-Issues.pdf

PROTECTING PRIVACY OF PEOPLE WITH SUBSTANCE USE DISORDERS

The Substance Abuse and Mental Health Services Administration (SAMHSA) is responsible for maintaining the regulation 42 C.F.R. part 2—known as Part 2—which protects people with substance use disorders, in addition to HIPAA protections. The law requires that people give consent to share their substance use and treatment history before their health care or treatment provider releases the information. The regulations were enacted so that people could seek treatment without fear of negative consequences such as arrest, job loss, or loss of custody of their children. Any dataset that contains information protected by these regulations would need to meet the guidelines for consent (specific or general) and protect the information accordingly. SAMHSA maintains a list of frequently asked questions about Part 2 and a fact sheet on how to exchange Part 2–protected data. The Legal Action Center has helpful resources on Part 2 for patients, families, and practitioners.

HUMAN SUBJECTS' PROTECTION

People often think about the need for human subjects' protections for medical research, drug trials, or psychological experiments. But participating in social science research also involves direct or indirect risks with how secondary data get used and shared. For example, asking people questions about their experiences might bring up traumatic memories and cause emotional harm. Risk can be physical, psychological, social, economic, or legal. Mapping community disease burden, for example, can put privacy and confidentiality at risk if small numbers of people are being represented. Beyond the risks to individuals, communities can be placed at risk when data are mapped, defining neighborhoods as 'high crime' or with 'failing schools' without sufficient attention to the local assets, historical context, and potential steps to address disparities. Such labeling can do economic harm to these communities and residents. These potential harms and risks to privacy and confidentiality should be mitigated. These protections should do the following:[iii]

- **Inform** subjects about the purpose of the research, what data will be collected and how they will be used, and what methods will be used to safeguard data
- **Protect** privacy and maintain confidentiality
- **Minimize** risks to subjects
- **Balance** associated risks with anticipated benefits of research for individuals and society
- **Fairly distribute** burdens and benefits of research, paying attention to vulnerable populations

In some cases, the potential risks to participants are serious enough to require informed consent from participants. Informed consent means that the participants have either verbally or in writing been given a clear, accessible explanation of the research and of participation risks. Any primary data collection should have an informed consent process. Some regulations might also require informed consent for use of secondary data for research.

COMMUNITY-BASED IRBS

In "Community IRBs and Research Review Boards: Shaping the Future of Community-Engaged Research," the Albert Einstein College of Medicine, the Bronx Health Link, and Community-Campus Partnerships for Health define community-based IRBs as those run by community-based organizations or coalitions. These IRBs make sure the community benefits from research being conducted in the area, protect communities from harm, and give community members a voice in the research in which they are the subjects. Those registered with HHS can be found here.

IRBs review the procedures for gathering information to determine if the protections for a project are sufficient. Even if your organization does not have an IRB or is not conducting federally funded research, it is good practice to review the above goals and protect the people participating in your research. If you are required to complete an IRB and are not in an organization that has one, you can borrow the services of the IRB of a local partner or university or contract with a commercial IRB.

A federal policy known as the ***Common Rule*** (45 C.F.R. part 46) governs human subjects' protections for any federally funded research. These regulations help determine what research activities must be reviewed by an IRB.[iv] Over time, additional protections have been added for pregnant women, people in prison, and children.

Decision charts are available on the HHS website to help users figure out[v]

- whether an activity **is research** that must be reviewed by an IRB,
- whether the review can be performed by **expedited procedures**, and
- whether **informed consent** or its documentation can be waived.

The Common Rule was revised on January 19, 2017, and was supposed to take effect on January 19, 2018, but, as of April 2018, its implementation has been delayed at least 12 months.[vi] Not all the proposed changes to the Common Rule were incorporated into the final rule. Full text is available here: https://www.hhs.gov/ohrp/regulations-and-policy/regulations/finalized-revisions-common-rule/index.html.

This news release from HHS summarizes changes to the rule, including changes to consent and new example categories of research based on the levels of risk to participants.

http://wayback.archive-it.org/3926/20170127095200/https://www.hhs.gov/about/news/2017/01/18/final-rule-enhances-protections-research-participants-modernizes-oversight-system.html

INFORMED CONSENT

In some cases, secondary data with PII can be obtained only with the person's informed consent. For example, student education data might require consent if the exceptions under FERPA are not met. Collective impact initiatives, such as Promise Neighborhoods, need to have participants' consent to share data among service providers. Informed consent should also be gathered for any primary data collection effort, including interviews, focus groups, or surveys.

Measuring Performance: A Guidance Document for Promise Neighborhoods on Collecting Data and Reporting Results provides guidance on informed consent on page 143. NNIP partner CI:Now shared lessons (minute 0:56) on developing a consent to support the Eastside Promise Neighborhood in San Antonio. The Believe to Become Master Data Sharing Agreement also includes the consent and parental consent forms used.

Institutional Review Boards and the HIPAA Privacy Rule
US Department of Health and Human Services, National Institutes of Health. (2004).

This document provides an overview of the IRB and its role under HIPAA. IRB approval is required for a covered entity to release protected health information (PHI) for research purposes without first obtaining signed permission from the participant(s). The document also discusses instances when a waiver—excusing researchers from obtaining written permission from people to use PHI—can be approved and discusses the evaluation criteria.

https://privacyruleandresearch.nih.gov/irbandprivacyrule.asp

Promise Neighborhoods Technical Assistance on Institutional Review Board (IRB) Approval
Urban Institute, prepared for the US Department of Education. (2013).

This article discusses two options for organizations requiring IRB approval that do not have an IRB and cannot set up their own (because of time sensitivity or burden): (1) borrow the IRB of a local partner or university or (2) contract with a commercial IRB to review proposals within a set time frame for a fee. Your organization can search for registered IRBs through the Office of Human Research Protections database. Fees for commercial IRBs can range from $500 for an expedited review to $1,200 for a full review; the differences between both types of reviews are also discussed.

https://www.neighborhoodindicators.org/library/catalog/promise-neighborhoods-technical-assistance-institutional-review-board-irb

Protecting Human Research Participants
National Institutes of Health, Office of Extramural Research. (n.d.).

All key staff involved in the human subjects' component of research (including new staff) must have human subjects' education. The National Institutes of Health offers a free, online training consisting of seven modules that address the principles defining ethical research using human subjects, as well as

the regulations, guidance, and policy related to the implementation of such research. The entire course (including quizzes) takes about three hours.

https://phrp.nihtraining.com/users/login.php

Human Subjects' Research

Citi Program. (n.d.)

Citi Program provides training for anyone involved in human subjects' research directly or for overseeing such research (including IRBs). Content is organized into two courses: biomedical and social-behavioral-educational. Citi Program also includes modules on other topics, including a course designed for IRB chairs. Most courses are available in both English and Korean. Fees apply.

https://about.citiprogram.org/en/series/human-subjects-research-hsr/

Notes

[i] For detailed definitions of privacy and confidentiality see National Research Council (1993, 22).

[ii] Peg D. Hall and Matt Nickel, "New Medical Privacy Law in Texas: What You Need to Know," Dallas Bar Association, July 24, 2015, http://www.dallasbar.org/book-page/new-medical-privacy-law-texas-what-you-need-know.

[iii] See the Urban Institute's Institutional Review Board document:
https://www.reginfo.gov/public/do/DownloadDocument?objectID=34696401.

[iv] "Federal Policy for the Protection of Human Subjects ('Common Rule')," US Department of Health and Human Services, last updated March 18, 2015, https://www.hhs.gov/ohrp/regulations-and-policy/regulations/common-rule/index.html; "45 CRF 46," US Department of Health and Human Services, last updated February 16, 2016, https://www.hhs.gov/ohrp/regulations-and-policy/regulations/45-cfr-46/index.html.

[v] "Human Subject Regulations Decision Charts," US Department of Health and Human Services, last updated May 10, 2016, https://www.hhs.gov/ohrp/regulations-and-policy/decision-charts/index.html.

[vi] "Federal Policy for the Protection of Human Subjects: Proposed Six-Month Delay of the General Compliance Date While Allowing the Use of Three Burden-Reducing Provisions during the Delay Period," 83 Fed. Reg., 17595, https://www.gpo.gov/fdsys/pkg/FR-2018-04-20/pdf/2018-08231.pdf.

CHAPTER 3. DATA SECURITY PLAN

Data security plans are important for creating and maintaining successful data security. PTAC's Data Security Checklist summarizes the essential components of a data security plan for protecting education records and is a helpful starting point for thinking about the best way to provide clear guidance on data security (PTAC 2015b). A data security plan should address the following elements: (1) data use and handling procedures, (2) vulnerability assessment and risk management, and (3) data security incidents and breaches. We have also included resources for monitoring compliance with data security plans. Additional references on these fundamental components of a data security plan are discussed below.

Data security plans are usually a required part of a data-sharing agreement. NNIP partners have noted that updating the data use agreement every time the data security plan changes can be burdensome. Instead, partners recommend including a reference to the data security plan in the data-sharing agreement. That way, the data security plan can be updated without having to update the entire agreement. It also keeps organizations from revealing all elements of their data security. Although transparency is important for building trust with the data provider, exposing all the details of a data security plan can increase the risk of a data security breach. If an entity contracts with a technology center for web-based data storage and online access, it is important that the requirements for compliance with data security standards and auditing for compliance be part of the contracted agreement with the service provider. It is then important to conduct and document a periodic compliance audit, which can be done via interview. The documents below provide general guidance on the policies and procedures that are important for successful data security.

Measuring Performance: A Guidance Document for Promise Neighborhoods on Collecting Data and Reporting Results

Jennifer Comey, Peter A. Tatian, Lesley Freiman, Mary K. Winkler, Christopher R. Hayes, Kaitlin Franks, and Reed Jordan. (2013).

Urban's guidance summarizes the elements of a data security plan (page 153): What data are being collected and why? How are data transferred and stored? How long are data kept? Who has access to the data? And who is accountable for overseeing compliance with the plan? The guide provides a sample plan in appendix 6.6.

https://www.urban.org/research/publication/measuring-performance-guidance-document-promise-neighborhoods-collecting-data-and-reporting-results

NeighborhooodInfo DC Data Security Guidelines
NeighborhoodInfo DC. (2010).

Urban's data security guidelines document the policies and procedures established to safeguard NeighborhoodInfo DC's confidential data. The document identifies a data security officer, who enforces data security procedures, which include limiting access, physical security, and transferring confidential data. The plan also specifies staff responsibilities, including password management and file permissions, labeling confidential storage media, maintaining access logs, and disposing of electronic and physical data.

https://www.neighborhoodindicators.org/sites/default/files/publications/sample_data_security.pdf

Information Security Policy Templates
The SANS Institute. (2018).

The SANS Institute, an information security training organization, offers free information security policy templates on its website. It offers templates for 27 security requirements that cover topics related to network security (e.g., acquisition assessment and remote access), server security (e.g., database credentials and software installation), web application security, and general information security (e.g., encryption and data breach response).

https://www.sans.org/security-resources/policies

DATA USE AND HANDLING PROCEDURES

Securing data is not only a technical problem. Many data security breaches are failures (intentional or otherwise) to follow established security procedures and protect confidentiality when transferring, handling, and storing data. In this section, we have included advice and resources for establishing these necessary procedures.

Secure Transfer

Confidential data are vulnerable to unauthorized breaches or disclosures when they are transferred between two entities. Procedures should be established to ensure that confidential data are transferred securely and only to authorized users. Depending on the transfer procedure, it might be necessary to encrypt and password-protect data (e.g., for in-person transfers or transfers by mail). To control access, data transfer should be coordinated with the sending entity, and data could be transferred in person, by mail (if certified and tracked), or electronically using Secure File Transfer Protocol. Email is not a secure means of data transfer.

Sharing Sensitive Data within the Government
Sunlight Foundation. (2015).

Data sharing within the government is governed by both formal and informal processes. This post discusses the data-sharing mechanisms frequently used when sharing data within the government and how the government shares data with external research organizations. It includes examples of data-sharing partnerships, including a description of California's partnership with research organizations around county correctional data.

https://sunlightfoundation.com/2015/02/11/sharing-sensitive-data-within-government/

Access Control

Limiting the number of people who can access the data can reduce the risk of data breach or disclosure. Access control procedures determine who has access to read, write, or modify data and information. One way to control access is to physically limit who has access to the data or computers—this can be through secured locks to offices or the server room. There are several ways to limit and monitor electronic access.

Role-based access means access to data is determined by a person's job responsibilities. Staff who need access to the data must verify their identity, such as a user ID and password, and their access would be tracked. Workstations for these staff should be set up to automatically log off after a certain period of inactivity. For some staff, aggregate or deidentified data will meet their information needs, and they should never have access to PII. These procedures ensure that only staff who need to use the data for a legitimate purpose have access.

Authentication technologies provide assurance that the person is authorized to access network assets, services, and information. Authentication can fall into three categories: something you know, something you have, or something you are. At a minimum, authentication includes having passwords (something you know) that are never shared, with a minimum length and level of complexity. Users should avoid easily guessed passwords and ones that include dictionary words or names. Combinations of uppercase and lowercase letters, numbers, and special characters is recommended. Long passwords are more secure than short (even random) passwords. We recommend using a random password generator and changing your password every 90 days. You can test a password's security here.

Two-factor authentication combines two of the authentication elements mentioned above. For example, a user would need to access the data with a user-generated password (something you know) and a series of random numbers generated by an application common to the authentication system and the user. Two-factor authentication costs more, but the additional security might be necessary and should be considered based on the data's sensitivity. Two-factor authentication might be more important for situations where there are remote users or "super users," who have greater access to the data (PTAC, n.d.).

Storage

Physical data, such as flash drives or CDs, should be stored in a locked filing cabinet or in a locked office. Best practices for storing electronic data include storing it in password-protected, encrypted storage devices. A strong encryption standard of 256 bits is recommended. Staff accessing confidential data electronically should have a hardened workstation (all the necessary hardware is installed on the computer to prevent unauthorized access). This includes

applying all security patches, removing unnecessary software, and maintaining necessary firewall, antivirus, antimalware configurations.

Increasingly, cloud storage is becoming an option for storing confidential data, although there are many debates about whether cloud storage is appropriate for sensitive data. Many industries, including health care and banking, already use cloud storage.[vii] Government agencies might also have requirements about the physical location of the cloud servers or place restrictions on the use of cloud storage that get incorporated into data-sharing agreements. The Coleridge Initiative developed a cloud-based administrative data research facility that contains confidential data. This facility has received approval from the Federal Risk and Authorization Management Program, or FedRAMP.[viii] Austin-based Children's Optimal Health, an NNIP partner, uses a cloud storage system designed to be secure. The system is both HIPAA and FERPA compliant and has passed multiple security audits, including a HIPAA HITECH audit. Several NNIP partners noted that neither cloud-based storage nor storage on a protected network survey or hardened workstation are inherently secure. They recommend that no matter what storage options you use, you need to have the data encrypted in transit or at rest and that staff are trained on security procedures and have the proper procedures in place to make sure you can prevent and address emerging threats.

Cloud Computing Frequently Asked Questions
Privacy Technical Assistance Center. (2015).

The US Department of Education has a list of qualifications to check before transferring to cloud services. The department recommends evaluating your organization's protection and security capabilities before moving to a cloud system. It also offers best practices when moving to the cloud. The department cannot recommend whether your organization should move to the cloud. Additional resources are provided regarding federal regulations and guidance about security issues.

https://studentprivacy.ed.gov/sites/default/files/resource_document/file/FAQ_Cloud_Computing_0.pdf

How Nonprofits Can Ensure Security and Compliance of Sensitive Data in the Cloud
Matt Goche. (2014)

The article mentions how the cloud—which you use when working with Gmail, Google Docs, Microsoft Office Live Workspace, Salesforce, Yahoo! mail, and other applications—is advantageous for nonprofits from a cost, flexibility, and security standpoint. Security and compliance are still important considerations for nonprofits worried about protecting their data or meeting compliance regulations. The article presents the following best practices for ensuring data security and meeting certification obligations: focus on security basics such as antivirus protection or administrative passwords, know and adhere to industry-specific rules and regulations, test annually to make sure safety processes are consistent, and use trusted providers.

https://www.nten.org/article/how-nonprofits-can-ensure-security-and-compliance-of-sensitive-data-in-the-cloud/

VULNERABILITY ASSESSMENT AND RISK MANAGEMENT

One strategy for improving data security is to conduct a vulnerability assessment and to develop a plan to address the risks your organization is mostly likely to face. The Nonprofit Technology Network suggests that organizations begin this process by creating an inventory of all data by location. Once the data have been inventoried, they can be categorized into data that cannot be lost, data that cannot be exposed, and data that are nonessential. For data that cannot be exposed or lost, you should consider the risks of exposure (e.g., physical theft, Ransomware, or natural disaster) and the relative likelihood of those risks. The resources below can help your organization assess its vulnerability. We also provide more specific resources in this section on data security audits and threats.

NIST Cybersecurity Framework
US Department of Commerce, National Institute of Standards and Technology. (2018).

The National Institute of Standards and Technology, or NIST, provides guidance and best practices on its website for managing cybersecurity risks. Its framework describes five main functions of cybersecurity—identify, protect, detect, respond, and recover—which are applicable to cybersecurity risk management and risk management at large. The institute provides training modules and frequently asked questions on its website that can be accessed at any time. It also posts information about upcoming events.

https://www.nist.gov/cyberframework

Risk Assessment Template
RoundTable Technology. (2015).

RoundTable Technology's risk assessment template helps organizations identify potential risks and safeguards for their data. The template is organized by data source. For each source, the template records basic information (i.e., name of the information, description, and location). Your organization must then judge the data's level of confidentiality, integrity, and availability (high, medium, or low for each). The risks, safeguards, and recommendations are considered through that lens.

https://docs.google.com/spreadsheets/d/1L1FP-ePpPLcrkYKKQkuLdFHV6xj9Y-k6z4jaBQKxgKE/edit#gid=0

Information Security Resources
Educause. (2017).

Educause, a nonprofit focused on advancing education through information technology, has several resources on its website to help organizations gauge their need for cyber liability insurance. These include a tool for evaluating the maturity of your organization's information security program, which is intended for a chief information security officer, and a tool for evaluating the strength of third-party cloud service providers, which is designed for a general audience.

https://library.educause.edu/resources/2016/10/higher-education-cloud-vendor-assessment-tool

Data Security Audits

Data security audits are tools that organizations can use proactively to assess their preparedness on data security. Certain government agencies also require proof of a security audit before they will agree to share data or will request an audit to monitor compliance with agreed-upon procedures.

Audit Protocol–Updated August 2018
US Department of Health and Human Services. (2018).

The website details the audit protocol the Phase 2 HIPAA Audit Program uses to review the policies and procedures of entities and business associates to ensure they meet standards and implementation specifications stated in the privacy, security, and breach notification rules. The audit protocol provides the rule and regulatory provision for each key activity, as well as the audit type, which includes privacy, security, and breach. Key activities include elements that are to be checked against the standards and implementation specifications during the audit.

https://www.hhs.gov/hipaa/for-professionals/compliance-enforcement/audit/protocol/index.html

Responding to IT Security Audits: Improving Data Security Practices
Privacy Technical Assistance Center. (2011).

The paper explains what audits are with regards to improving data security and recommends steps for creating a point-by-point plan in response to any audit findings. Audit response recommendations include meeting with auditors to understand audit findings, prioritizing tasks according to the severity of findings, identifying corrective actions and solutions, developing an implementation schedule for effective remediation in a timely manner, confirming plans with senior management and regulatory authorities, and delivering on the plan. The source also includes examples of responses to audit findings at the end in the appendix.

https://studentprivacy.ed.gov/resources/responding-it-security-audits-improving-data-security-practices

Protection of Personally Identifiable Information in Statewide Longitudinal Data Systems
Daniel P. Schultz. (2017).

The source provides examples of audits in Indiana and Oregon, which determined whether the state's department of education had sufficient controls to prevent, detect, report, and respond to unauthorized access and the disclosure of personally identifiable information in the state's Statewide Longitudinal Data System. These memos discuss the findings of the audits, which determined the systems did not meet minimum security requirements, and include recommendations to address the findings.

https://www2.ed.gov/about/offices/list/oig/auditreports/fy2017/a06q0001.pdf (Indiana)
https://www2.ed.gov/about/offices/list/oig/auditreports/fy2016/a02p0007.pdf (Oregon)

Threats

There are both technical and nontechnical threats to data security. Technical threats include nonexistent security architecture, unpatched client-side software and applications, "phishing" and targeted attacks ("spear phishing"), websites, poor configuration management, mobile devices, cloud computing, removable media, botnets, and zero-day attacks. Nontechnical

threats include insiders, poor passwords, physical security, insufficient backup and recovery, improper destruction, social media, and social engineering. Failure to properly and regularly apply security patches, changes to the computer program that update or fix vulnerabilities, are also a threat to data security.

Data Security: Top Threats to Data Protection
Privacy Technical Assistance Council. (2015).

The paper discusses technical and nontechnical threats to education data and information systems, as well as suggestions for risk mitigation measures for each threat. The paper provides additional resources on data security at the end.

https://studentprivacy.ed.gov/resources/issue-brief-data-security-top-threats-data-protection

Data Security Threats: Education Systems in the Crosshairs
Mike Tassey. (2012).

The presentation highlights the basics of data security threats, including what constitutes a threat, the actors and types of threats, the information attackers are looking for, the methods attackers use to access information, and the ways we can mitigate threats. The source also details how online and internet advancements have made it easier for attackers to access information. The presentation ends with additional Privacy Technical Assistance Center resources and information on various assistance that PTAC provides, including site visits and rapid response.

http://ptac.ed.gov/sites/default/files/Data_Security_Threats.pdf

Phishing Resources
Cofense. (2018).

Cofense helps organizations prepare employees to be more resilient and vigilant against targeted cyberattacks. Its services include simulated phishing attacks to condition employees to resist phishing attempts and to identify and report attacks in real time. It also helps organizations identify and respond to attacks efficiently.

https://cofense.com/

DATA SECURITY INCIDENTS AND BREACHES

PTAC defines a data breach as a situation when there is an unauthorized release or access of PII or other information that is not suitable for public release (PTAC 2012). Examples of breaches include hackers gaining access to data through a malicious attack; lost, stolen, or temporarily misplaced equipment (e.g., laptops, mobile phones, or portable thumb drives); employee negligence (e.g., leaving a password list in a publicly accessible location or technical staff misconfiguring a security service or device); and policy or system failure (e.g., a policy that does not require multiple overlapping security measures—if backup security measures are absent, failure of a single protective system can leave data vulnerable). A data security incident can occur without malicious intent, such as an employee improperly emailing sensitive data. These incidents constitute a misuse of data but might not result in a breach.

Data security incidents and breaches can happen to any organization. The procedures your organization puts in place are important for properly managing the incident and breach after it occurs. Liability insurance is also a consideration for organizations wanting to insure against the risk of a breach. An emerging trend is a requirement by HIPAA for covered entities that those entities with whom they share data must demonstrate that they carry sufficient data breach liability insurance. The following section identifies resources related to data security procedures and liability insurance.

Procedures

Procedures for managing a security incident or breach should be easy to follow so staff have clear instructions in the event of an incident. Incidents need to be reported internally at minimum but might not require further action besides additional training and review of procedures. Data breaches are more serious; the procedures to address them typically include information about how breaches are detected and documented, who is responsible for responding, communication with the rest of the organization or constituents, and recovering files. It is also helpful to outline how the response team should respond in different scenarios.

Security Breach Notification Laws
National Conference of State Legislatures. (2018).

Security breach notification laws specify when private or governmental entities are required to inform people of a data security breach involving personally identifiable information. Information about security breach notification laws in all 50 states, the District of Columbia, Guam, Puerto Rico, and the Virgin Islands can be found on the National Conference of State Legislature's website.

http://www.ncsl.org/research/telecommunications-and-information-technology/security-breach-notification-laws.aspx

Data Breach Response Checklist
Privacy Technical Assistance Center. (2012).

This checklist is designed to assist educational organizations building a comprehensive data breach response capability. It contains information on procedures to establish a breach protocol, prevent a breach, and respond to one. PTAC provides additional resources at the end of the document.

https://studentprivacy.ed.gov/resources/data-breach-response-checklist

Best Practices for Victim Response and Reporting of Cyber Incidents
US Department of Justice, Criminal Division, Computer Crime and Intellectual Property Section, Cybersecurity Unit. (2015).

This document identifies four steps your organization should take in the event of a cyber incident: (1) perform an initial assessment, (2) implement measures to minimize continuing damage, (3) record and collect information, and (4) notify key people and organizations. This document also discusses the precautions organizations should take before a cyber incident and what no to do during the aftermath.

https://www.justice.gov/sites/default/files/opa/speeches/attachments/2015/04/29/criminal_division_guidance_on_best_practices_for_victim_response_and_reporting_cyber_incidents.pdf

ISC Information Security Incident Protocol
University of North Carolina at Charlotte's Urban Institute (n.d.).

Charlotte's Incident Protocol is a step-by-step guide for how to proceed following a data security incident. Steps include notifying the appropriate executive leadership and legal counsel, developing an information security incident response team, developing a communication strategy, and discussing the incident as a team.

https://www.neighborhoodindicators.org/library/catalog/isc-information-security-incident-protocol

WPRDC Breach Plan
Western Pennsylvania Regional Data Center. (2018).

The WPRDC breach response plan outlines the steps that staff should take in the event a breach involving the open data portal, including making the dataset private, notifying the publishing organization, and documenting the incident and response to improve data management practices.

https://www.neighborhoodindicators.org/library/catalog/wprdc-breach-plan

Guide to Protecting the Confidentiality of Personally Identifiable Information (PII)
Erika McCallister, Tim Grance, and Karen Scarfone. (2010).

The document describes a risk-based approach to protecting the confidentiality of PII, intended mostly for US federal government agencies and those that conduct business with these agencies. Section 5 includes details on developing an incident response plan to deal with PII breaches. It also includes useful guidance on determining the impact and sensitivity of PII and safeguards to protect confidentiality.

http://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-122.pdf

Example Data Security Plan
National Neighborhood Indicators Partnership. (2018).

This example plan was shared by an NNIP partner for their data system that allows for data sharing and access with multiple entities. It discusses the procedures that comply with regulations like HIPAA; administrative, physical, and technical safeguards; and data breach notification procedures.

https://www.neighborhoodindicators.org/library/catalog/example-data-security-plan

Liability Insurance

Liability insurance for data security breaches looks different across different types of organizations. NNIP partners working at small nonprofit organizations have shared that it can be difficult to afford liability insurance and that sufficient coverage insures between $1 million and $6 million in damages. Cloud storage might be a good option for smaller organizations that cannot afford insurance. An agreement with a cloud service provider can specify liability in the event of a breach. Organizations storing data in house are taking on the liability of a breach and might need insurance. NNIP partner staff who are part of a center in a university have mentioned that they are covered by the university's insurance policy.

Frequently Asked Questions about Cyber Insurance
University Risk Management and Insurance Association and Educause. (2017).

In response to growing interest in cyber liability insurance, this resource guides organizations considering liability insurance. Beginning with a brief discussion of cyber risk, aspects of coverage, and the types of data and losses covered by insurance, this document provides key considerations for anyone interested in cyber insurance. Covered topics include special concerns for public institutions, the process for purchasing cyber insurance, what happens when an event occurs that might be covered by insurance, and what institutions need to know about the claims process. The document also discusses the trade-offs of self-insuring compared with buying cyber insurance and the implications of cyber insurance on your organization's computer systems and processes.

https://library.educause.edu/~/media/files/library/2017/10/urmiafaq.pdf

MONITORING COMPLIANCE

Monitoring compliance is important for identifying and minimizing potential security risks. An important first step is requiring that staff read data security plans and sign confidentiality pledges and nondisclosure agreements before accessing confidential data. These documents ensure that staff understand the confidential nature of the data and that they must comply with all the data security procedures (see example). In addition, your organization's staff and data partners should receive training on internal data security procedures and expectations. We also recommend that your organization conduct data access audits to ensure compliance with security policies.

Staff Training

Most data breaches are caused by human error. Your organization should train staff on internal data security procedures and expectations. All staff working with confidential data should be given access to and must read the data security plan. Trainings should focus on the day-to-day elements of data security and include coverage of rare issues, such as data breaches. Day-to-day elements include guidance on password security, remote access to data, data backup and recovery, how to securely transfer and store physical and electronic data, and handling common threats, such as how to identify and protect against phishing scams.

Rice Cyber Security Awareness Training
Rice University. (2016).

Rice University's information technology department worked with Securing the Human (from SANS Security Awareness Training) to develop modules on cybersecurity. Research staff are required to take mandatory modules, including privacy, phishing, passwords, encryption, and protecting your personal computer. They also have access to additional recommended material, including modules on cloud services, PII, physical security, and working remotely. Securing the Human provides additional resources on its website, including trainings on how to build a successful cyber security awareness program and how to promote lasting awareness.

https://www.sans.org/security-awareness-training

Trusted CI Cyber Security Trainings
Trusted CI, the NSF Cybersecurity Center of Excellence. (n.d.).

Supported by the National Science Foundation (NSF), Trusted CI provides information about upcoming trainings, webinars, and conferences related to cyber security. Content is targeted toward the NSF community and the broader community of scientists and scholars to increase people's understanding of cybersecurity and explain how to maintain effective cybersecurity programs. Trusted CI also includes archived slides and videos from past events on its website.

https://trustedci.org/

Data Security and Management Training: Best Practice Considerations
Privacy Technical Assistance Center. (2015).

PTAC's guide to providing effective training in data security and management highlights best practices for trainings. The document recommends content for trainings and a different training delivery methods. On-demand training, allowing participants to progress through the material at their own pace, can be a good option for reaching all employees within your organization. Virtual trainings, delivering specific material at the same time, are a good option for allowing participants to have questions answered in real time. On-site trainings can be more in depth, tailoring the training to fit the roles of the participants in the room.

https://studentprivacy.ed.gov/sites/default/files/resource_document/file/Data%20Security%20and%20Management%20Training_1.pdf

Training Data Partners

Data security breaches can also occur when data are received by an external partner. An example of this would be receiving data from a partner through an email instead of a Secure File Transfer Protocol. Organizations that plan to receive data from external partners should train their partners on how to securely store and transfer confidential data.

Data Privacy and Security Tips
Amy Hawn Nelson. (2017).

If you plan to receive confidential data from other organizations, you need to train your data partners in data security procedures. Hawn Nelson's email, sent in honor of National Data Privacy Day, is a reminder to partners about common data privacy risks relating to improper storage and transfer of data and allowing unlimited data access. The email also has tips to improve data security, including providing several layers of security and creating a culture of data integrity.

https://www.neighborhoodindicators.org/library/catalog/example-email-data-partners-data-privacy-and-security-tips

Capacity Building Training Series
Seeding Success Partnership. (2015).

The Seeding Success Partnership in Memphis, Tennessee, a member of the StriveTogether network, provided an 11-hour curriculum that provides training for organizations in the partnership on data sharing, FERPA, and using data for continuous improvement. Descriptions of the training modules are available in the resources section of Principle 7 of the StriveTogether Data Drives School-Community Collaboration playbook.

http://circ.communityreport.org/data-sharing-playbook/reports/73/data-drives-school-community-collaboration#page-9

Access Audits and Procedures

Access audits assess how well staff follow data access procedures and can identify security risks. Strategies for access audits include comparing the list of staff who have access to the data with the list of staff who need to have access to the data and ensuring that all staff have signed confidentiality pledges and received the required trainings. Information that is routinely collected about your information technology system can be used to improve system security. Logs track information about specific security events. Information collected in the log can be analyzed to better understand security incidents, identify vulnerabilities, and improve overall network security.

Data Stewardship: Managing Personally Identifiable Information in Electronic Student Education Records

US Department of Education, Institute for Education Sciences, National Center for Education Statistics. (2010).

Access audits can help you monitor compliance with data access procedures. On pages 17 and 18 of this document, the National Center for Education Statistics identifies the three steps for conducting a successful access audit: (1) review the list of staff with access to ensure that no one has access who does not need it, (2) ensure that all staff with access have completed all required trainings and background investigations and have signed confidentiality pledges, and (3) review the level of data access each staff member has and ensure that no staff member has greater access than is needed. Greater detail on these steps and other types of audits are provided in the document.

https://nces.ed.gov/pubs2011/2011602.pdf

Guide to Computer Security Log Management

US Department of Commerce, National Institute of Standards and Technology. (2006).

This document seeks to provide your organization an understanding of computer security log management, discussing the development, implementation, and maintenance of log management practices throughout an organization. Topics covered include establishing log management infrastructures and developing and performing robust log management processes throughout an organization. This is not a step-by-step guide to implementing or using logging technologies.

Guide: https://csrc.nist.gov/publications/detail/sp/800-92/final
Bulletin: https://csrc.nist.gov/CSRC/media/Publications/Shared/documents/itl-bulletin/itlbul2006-10.pdf

[vii] Jennifer Bresnick, "82% of Hospitals Using Cloud Services for HIE, Data Storage," *HealthITAnalytics,* June 24, 2014, https://healthitanalytics.com/news/82-of-hospitals-using-cloud-services-for-hie-data-storage; Penny Crosman, "Banks Ramp Up Cloud Adoption; Holdouts Cite Hands-On Control," *American Banker,* March 6, 2015, https://www.americanbanker.com/news/banks-ramp-up-cloud-adoption-holdouts-cite-hands-on-control; David F Carr, "Larry Ellison Says Security Can Be Better in the Cloud," accessed July 31, 2018, https://www.oracle.com/corporate/features/openworld-2015-larry-ellison-demo-keynote.html.

[viii] For details on the security of the Coleridge administrative data research facility, see Coleridge Initiative (n.d.). Learn more about FedRAMP at "About Us," FedRAMP, accessed September 7, 2018, https://www.fedramp.gov/about/.

CHAPTER 4. DATA LIFE CYCLE

The data governance and security plans described in the previous chapters will guide the policies and procedures for the entire life cycle of a dataset. This chapter provides resources connected to common tasks within specific stages of the data life cycle: acquisition, processing, dissemination, and destruction. We do not provide resources specific to *analyzing* data, but the overall access and privacy protections would apply.

For specific projects or data sources, you can develop a *data management plan* that covers the intentions for the organizational actions throughout the entire life cycle and is sometimes required by government agencies during the proposal process. The University of Minnesota has compiled a list of federal agency mandates related to data management plans, with links to regulations, guides, and portals. Some of the plan elements refer to the organization's overall data governance and security policies, but others will vary depending on the type and intended application of the data.

Guidelines for Effective Data Management Plans
Inter-university Consortium for Political and Social Research. (2012).

This online guide lists the elements of a data management plan, each with a mapping to the National Science Foundation standards. It also offers a framework for creating a data management plan, diving into more detail on the importance and content for each element with examples along the way.

https://www.icpsr.umich.edu/icpsrweb/content/datamanagement/dmp/index.html

Data Management Plans
Digital Curation Centre. (n.d.).

This site was developed with United Kingdom standards in mind but contains resources relevant for any country. The tool aids researchers by providing guidance and best practices via links to Digital Curation Centre resources and external advice. Its 2011 guide on developing and sharing a data management plan is supplemented by a quick checklist of the components of a plan and examples from various fields and applications. Finally, it includes a web-based tool to help users create plans personalized to their context or funder.

http://www.dcc.ac.uk/resources/data-management-plans

Ten Simple Rules for Digital Data Storage
Edmund M. Hart, Pauline Barmby, David LeBauer, François Michonneau, Sarah Mount, Patrick Mulrooney, Timothée Poisot, et al. (2016).

This article in the *PLOS Computational Biology* journal describes 10 simple rules for digital data storage based on the experiences of instructors for the software and data carpentry initiatives. The rules span the entire data life cycle and encourage data stewards to think about such factors as the ultimate uses, documentation, and ease of analysis. It also provides a glossary of common vocabulary related

to file formats, programming and algorithms, and persistent identifiers for readers who are not familiar with these terms.

http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005097

DATA ACQUISITION

Your organization can acquire secondary data in several different ways. State and local governments often publish nonconfidential data through public websites, in which the terms of use are set globally by each agency. They can also share data in response to Freedom of Information Act (FOIA) requests. Most often, to acquire confidential or sensitive data, you will need to negotiate with the source agency to receive data and develop a written agreement that stipulates how the data will be transferred, stored, managed, and used. The sections below discuss each of these paths for obtaining data.

Public Websites

Federal, state, and local governments are publishing more data than ever before through open data portals. The data on these portals will not have any private or confidential information but may offer aggregate indicators based on confidential data. Open data portals allow you to search through the catalog and download data in various file formats or access the data through an application programming interface (API) that will update data automatically. Open data portals sometimes have data licenses that specify the allowed use of the data (e.g., noncommercial uses), disclaimers, and requirements or requests for citations. For example, the City of Chicago's data portal includes indicators on natality, mortality, infectious disease, and lead poisoning for the city's 77 community areas. The Western Pennsylvania Regional Data Center has indicators related to cardiovascular disease risk for census tracts in Allegheny County.

The federal government maintains the cross-department open data portal at www.data.gov. Many agencies still have dedicated websites to access data collections, such as HUD User or the Centers for Disease Control and Prevention's WONDER website, or specific datasets, such as the Department of Education's Common Core of Data. NNIP also maintains a list of nationally available data sources with data below the city level, most of which are from federal agencies.

Now, all states also have open data portals, though the usability and extent of holdings vary. For cities, the Sunlight Foundation and Open Knowledge Foundation maintain the US City Open Data Census for more than 260 cities, cataloging the presence of 20 types of data sources, such as crime reports, property assessments, service requests, or business licenses.

Many nongovernmental organizations also maintain data portals. These might be national in scope—such as the Urban Institute's Education Data portal, which has repackaged federal education data for easier use—or the Data Archive from the Inter-university Consortium for Political and Social Research (ICPSR), which is a common archive for researchers to deposit their data. Local organizations, such as Data Driven Detroit or the Boston Area Research Initiative, also host portals with data related to various topics. All may specify terms of how the data can be used or requirements on citation.

Freedom of Information Requests

Freedom of information laws allow access to government records, with the premise that they are obliged to be transparent about the information related to decisions and operations of public agencies. The laws establish a formal process by which you can request government-held information, to be received freely or at minimal cost. The laws also specify standard exceptions to the requirement for release. For example, the federal FOIA law provides for public access to federal executive branch agency records but exempts classified documents, those related to internal personnel rules and practices, and documents protected by other statutes. In addition, every state has laws governing FOIA for state agencies, often called open records or open meetings laws. These laws may list exemptions beyond those in the federal regulations or cover additional types of government units, such as commissions. All 50 states have public records laws that govern how people can obtain documents and other information from state and local government entities.

Freedom of Information Act (Federal) Research Guide
Georgetown Law Library. (2013).

This guide provides an overview of the federal Freedom of Information Act, including a list of the nine categories of documents that are exempt from disclosure. Next, it offers advice on making a FOIA request. In addition, it links to several handbooks on FOIA and provides brief descriptions of related administrative materials, regulations, and advocacy groups.

http://guides.ll.georgetown.edu/c.php?g=320807&p=2146490

State FOI
National Freedom of Information Coalition. (n.d.).

The National Freedom of Information Coalition works nationally and with its state organization members to promote laws, policies, and practices that ensure expeditious access to state and local public records, proceedings, and officials. Its website compiles links to all state freedom of information and open records laws. It also provides sample FOIA request letters for all states and FOI hotlines across the country.

https://www.nfoic.org/organizations

Negotiated Data Sharing

Federal and local government agencies often share data with other public or private organizations for specific purposes. The agency might have a formal process for requesting data, but the process is frequently ad hoc. NNIP's Guide to Data Sharing offers advice for organizations negotiating with government staff about sharing data. It suggests steps for your organization to get organized before the request, including developing data governance policies and procedures that ensure protection of confidential data. Based on the experiences of dozens of local data intermediaries, the guide lists five common reasons data providers refuse requests, such as poor-quality data or staff shortages, and possible responses you can give. The AISP expert panel report by Petrila and coauthors (2017) addresses similar concerns, specifically for integrated data systems.

It is best practice to develop a memorandum of understanding (MOU), a written agreement to govern the transfer, use, and disposition of a dataset. Any agreement should have more conditions around security for confidential data than for data that are nonsensitive but just not in the public domain. Parties should develop an agreement through an iterative process that includes the agency's legal counsel to address concerns on all sides (Carlson et al. 2011) and write the language in plain language that all parties can understand (Petrila et al. 2017).

In its simplest form, an MOU could be a bilateral agreement between two entities for a time-limited project. If you anticipate that more than two organizations need access to the data, you should think through the conditions at the beginning of the negotiations. Multiparty agreements are feasible, but negotiations will likely be more complex and time consuming with navigation of the politics and legal requirements of multiple organizations.

CORPORATE DATA SHARING

Most resources in this guide focus on government data, but using corporate data for public good also has great promise. Programs like GovLab's Data Collaboratives offer guidance and real-world examples for using proprietary data for research and policymaking. The Urban Institute has also written about the idea of data philanthropy for the private sector. The motivations and incentives for private companies to share data differ from those of government agencies, as do the internal policies and laws that govern the use of their data. But our guidance on the components of data agreements apply to all data-sharing efforts, regardless of the source.

For example, in a pay for success (PFS) initiative, the agreement should include access for organizations doing the evaluation or validation of findings and acknowledge the evaluator's independence. As illustrated in the Massachusetts PFS agreement on job and English language training, other characteristics specific to PFS arrangements include the expectation of public

scrutiny, a clearly defined schedule of reporting data, and the link between outcome data and payments.

As described in Petrila et al. (2017), integrated data systems will be governed by multiple legal agreements. The MOU stipulates the terms among the agency managing the linked data system and the agencies contributing data. An IDS MOU is an example of a multiparty agreement, specifying multiple named agencies. It might also take the form of an enterprise MOU (or eMOU), to which agencies can sign on to the data-sharing terms of the IDS partnership, without having to re-sign as entities are added or dropped.

After the MOU is established, the IDS manager distributes data to users through *data use agreements* that might have to be approved for each individual project. These agreements fulfill the source agency's obligation from HIPAA and FERPA to have the data continuously under their control. These data licenses for open and confidential data are described below from the perspective of the data distributor and more generally in Petrila et al. (2017).

ELEMENTS OF THE DATA-SHARING AGREEMENT

1. Purpose and intended use of data sharing
2. Period of agreement
3. Description of data
4. Timing and frequency of updates
5. Custodial responsibility and data stewardship
6. Roles and responsibilities
7. Permissible data use, linking, and sharing (in compliance with requirements of source data)
8. Resources and costs of data sharing and data management
9. No warranty for data or linkage quality
10. Indemnification
11. Publication and dissemination of results
12. Termination and modification of this agreement

Several resources provide sample agreement documents for different kinds of data.

The Urban Institute discusses developing data-sharing agreements in *Measuring Performance: A Guidance Document for Promise Neighborhoods on Collecting Data and Reporting Results*. While focusing on data about children, their advice and description of the general components of agreements applies as well to all types of data.

This guide also recommends creating a modular and generalized document. This structure allows for groups to update attachments (e.g., the organization's security procedures) by mutual consent as procedures are revised or improved over time without modifying the primary legal document. Groups may

identify a narrow scope of the inquiry to facilitate data sharing, particularly if it is a new relationship (Carlson et al. 2011). By having the statement of allowable uses or scope of analysis as an attachment, groups can add approved uses as the relationship evolves.

The agreement should define who has access to the dataset, or fields within the dataset, and by which modes. A limited number of staff might be able to use record-level data with personally identifiable information and only on a computer not connected to the internet. A wider group of analysts could have remote access to the deidentified version of the file.

Legal Issues for IDS Use: Finding a Way Forward
John Petrila, Barbara Cohn, Wendell Pritchett, Paul Stiles, Victoria Stodden, Jeffrey Vagle, Mark Humowiecki, and Natassia Rozario. (2017).

This report, published by AISP, provides information and resources for agencies and organizations looking to build the legal framework integrating data across administrative sources. It contains information on common legal concerns for agencies considering sharing data, elements of data-sharing agreements, specific laws that apply, and templates and sample memoranda of understanding and data use licenses.

https://www.aisp.upenn.edu/wp-content/uploads/2016/07/Legal-Issues.pdf

Measuring Performance: A Guidance Document for Promise Neighborhoods on Collecting Data and Reporting Results
Jennifer Comey, Peter A. Tatian, Lesley Freiman, Mary K. Winkler, Christopher R. Hayes, Kaitlin Franks, and Reed Jordan. (2013).

Chapter 6 of the Urban Institute's guidance discusses data sharing in general—writing and negotiating data-sharing agreements with service providers and with researchers and evaluators—and describes the elements of a typical data-sharing agreement. A model master data-sharing agreement is provided in appendix 6.3 A.

https://www.urban.org/research/publication/measuring-performance-guidance-document-promise-neighborhoods-collecting-data-and-reporting-results

Developing a Master Data Sharing Agreement: Seeking Student-Level Evidence to Support a Collaborative Community Effort in Education
Neil E. Carlson, Edwin Hernández, Chaná Edmond-Verley, Gustavo Rotondaro, and Eleibny Feliz-Santana. (2011).

The article in *Foundation Review* details the development of a data-sharing agreement for the Believe 2 Become initiative, in which the NNIP partner in Grand Rapids, Community Research Institute at Grand Valley State University, played a role. Its data-sharing negotiations involved a small working group that communicated with a larger network of organizations and leaders. The authors also provide the details of their master data-sharing agreement. The main document includes terms, roles, responsibilities, names of key people, definitions of permissible data use, and statements regarding the school district's right of review and other warranties. The data-sharing agreement also includes 10 attachments, which consist of a clarification of which people can see what data at the organizational level, an agreement not to disclose confidential data, an up-to-date listing of current service providers and research partners, a description of protocols and procedures for linked primary datasets, a consent form, a research request form, a FERPA confidentiality-protection agreement, and a parental consent form.

Journal article: http://scholarworks.gvsu.edu/cgi/viewcontent.cgi?article=1043&context=tfr
Data-sharing agreement: http://www.cridata.org/b2bmdsa/

Data Sharing: Creating Agreements in Support of Community-Academic Partnerships
Paige Backlund Jarquín. (2012).

This guide from the Colorado Clinical and Translational Sciences Institute and the Rocky Mountain Prevention Research Center focuses on community-academic partnerships, with advice on how to build trust and engage the community at different stages. It offers generalized steps for data sharing and typical content of an agreement. Within each element, it offers prompting questions for the parties to consider. With the final section on resources, the sections on data management and partnership are remain relevant. (The references in the agreement template section are out of date).

http://www.ucdenver.edu/research/CCTSI/community-engagement/resources/Documents/DataSharingCreatingAgreements.pdf

Written Agreement Checklist
Privacy Technical Assistance Center. (2015).

The document summarizes the requirements for written agreements for data sharing under the studies exception and the audit or evaluation exception as specified in FERPA. The document also notes that state privacy or procurement laws could contain more stringent requirements regarding data-sharing written agreements. The document also includes best practices for written agreements.

https://studentprivacy.ed.gov/sites/default/files/resource_document/file/Written_Agreement_Checklist.pdf

Legal Guide to Administrative Data Sharing for Economic and Workforce Development
Center for Regional Economic Competitiveness. (2018).

This guide is intended for state government executives seeking to responsibly share data for evidence-based policymaking related to economic and workforce development. It is structured as a series of frequently asked questions, including what data sources are likely to be relevant, what the applicable laws and regulations are, and how agencies overcome cultural barriers to data sharing. Throughout, the document provides sample language for memoranda of agreement to address specific concerns, such as authorizing data sharing or protecting confidentiality. It is part of the State Data Sharing Initiative's Data Sharing Toolkit, which includes related webinars on state-level data sharing and a landscape analysis of which states have data management bodies or data-sharing legislation.

http://www.statedatasharing.org/data-sharing/2018-03_-_SDS_Legal_Guide_to_Administrative_Data_Sharing_for_Economic_and_Workforce_Development.pdf

NNIP's Collection of Example Data-Sharing Agreements
National Neighborhood Indicators Partnership. (2018).

The NNIP website provides various data-sharing resources from the local partner organizations in NNIP, which include nonprofits, university centers, and public agencies. There are a number of data-sharing agreements and MOUs, organized by topical area (e.g., health, education, and housing). We have also included links to examples and templates from organizations outside of NNIP that have been shared with us.

https://www.neighborhoodindicators.org/issue-area/295

DATA PROCESSING

As part of a data governance plan, your organization should have standard written policies and procedures for processing data. This refers to such activities as storing, cataloging, and producing metadata; assessing data quality; cleaning the data; and linking the data. You should periodically review these management practices to ensure they are efficient, produce high-quality data, and meet your organization's needs. This guide does not comprehensively cover all the elements of data processing but compiles tips and resources for a few elements of these activities, particularly those related to privacy issues.

Data Inventory

As recommended in the Vulnerability Assessment and Risk Management section in chapter 3, your data security procedures should include maintaining an organization-wide data inventory. A centralized listing of all data will help your staff find data they need, manage agreements, and ensure compliance with any terms of use. Not every derivative dataset should be included in the inventory; you will need to define the scope of data files to be included. Although an initial inventory might be time consuming, staff can incrementally update the catalog as part of their regular workflow when new data are acquired.

Any inventory should include one or more fields that indicate the access to the data (public or restricted), whether the file contains PII, and the reason for the restrictions, such as the set contains PII or is proprietary information. The Nonprofit Technology Network recommends that nonprofits create a data inventory and links to a template for an inventory as part of a risk analysis. The Digital Civil Society Lab at the Stanford Center on Philanthropy and Civil Society has produced a simple worksheet as part of its digital toolkit to help you get started.

Metadata

You should have standard procedures for producing metadata that documents datasets and fields within those sets. Data already published on the web from government agency sites or open data portals will likely have a minimum set of documentation noting the date published, labeling the fields, and the like. Often, administrative data obtained directly from government agencies has no or poor documentation about the data file overall or fields within it. The resources below describe best practices for developing metadata and other types of documentation.

The federal government has specified a set of fields and related values to document datasets through the Project Open Data Metadata Schema, based on the World Wide Web Consortium (W3C) Data Catalog Vocabulary. Although the project focuses on publishing metadata through

JSON files, the list offers a complete menu of potential fields that one could subset and adapt for other formats.

Guide to Social Science Data Preparation and Archiving
Inter-university Consortium for Political and Social Research. (2012).

This ICPSR guide's section on Data Collection and File Creation (Phase 3) describes best practices in building the data and documentation components. It covers widely accepted norms for creation and documentation of quantitative, GIS, qualitative, and other types of data in the social sciences. It also suggests best practices for creating metadata and a list of metadata elements compliant with the Dublin Core metadata standard.

PDF: https://www.icpsr.umich.edu/files/deposit/dataprep.pdf
Online guide: https://www.icpsr.umich.edu/icpsrweb/content/depcosit/guide/chapter3docs.html

Data Management Best Practices: Documentation
University of Pennsylvania Library. (n.d.).

The documentation section of the online guide describes best practices for documenting files and has links to resources to create readme files, codebooks, and data dictionaries.

https://guides.library.upenn.edu/datamgmt/documentation

Transformation

After the initial cataloging and documentation, analysts will transform the data to prepare the file for analysis. This includes assessing data quality, linking records to other data sources, imputing data to address quality issues, standardizing fields, or making other desired alterations. Analysts should consider privacy and confidentiality concerns as they undertake each of these steps. Standard procedures could include separating the data file to store the PII in a different location than the rest of the data.

NNIP's Lessons on Data Management Practices for Local Data Intermediaries
Rob Pitingolo. (2017).

Based on the experiences of the National Neighborhood Indicators Partnership, this brief is designed to help organizations develop and improve part of the data management process known as ETL (extract, transform, load). It provides advice on reading in data, cleaning and processing data, and publishing data. In addition, case studies from five organizations are included to illustrate the ETL process in practice.

https://www.neighborhoodindicators.org/library/catalog/lessons-data-management-practices-local-data-intermediaries

STANDARD DATA REPURPOSING FOR IDS

AISP's expert panel report "**Establishing a Standard Data Model for Large-Scale IDS Use**" has a section describing the processing needed to prepare source data for import into an IDS. It includes discussion of data structure and quality of the source data, transformation of the data through merging and record linking, and data cleaning.

https://www.aisp.upenn.edu/wp-content/uploads/2016/07/Data-Standards.pdf

Data Quality

Assessing the quality of administrative data is essential. These data are typically collected to run government operations, not for statistical analyses. The method of data entry, the purpose for collecting the data elements, and source of information are ways that can affect data consistency or completeness. Your data management plan should include procedures for assessing and documenting data quality.

Data Quality Assessment Tool for Administrative Data
William Iwig, Michael Berning, Paul Marck, and Mark Prell. (2013).

This tool, published by the US Bureau of Labor Statistics, can help you better understand the attributes of your data file and promotes practices to improve data quality. The tool provides questions that help you consider six dimensions of data quality: relevance, accessibility, coherence, interpretability, accuracy, and institutional environment. Questions are also organized by the phase of the data-sharing process: discovery, initial acquisition, and repeated acquisition.

https://www.bls.gov/osmr/datatool.pdf

Linking

Data managers can link individual records from different administrative data sources (as with integrated data systems) or to enhance or validate survey data. Researchers should first investigate whether the linkage of records they wish to implement requires informed consent. When informed consent is required, differences in who consents or not will affect data quality. Some of the resources below relate to linkage methodology and tools, including a listing of public and commercial linkage software. Others cover issues connected to individual consent, including data quality, ethics, and metadata.

Introduction to Data Linkage
Katie Harron. (2016).

This guide from the Administrative Data Research Network in the United Kingdom gives researchers a practical introduction to data linkage. It covers data preparation, deterministic and probabilistic linkage methods, analysis of linked data, and how to evaluate and report data linkage quality. It also includes examples relevant to health and other administrative data sources.

https://www.adrn.ac.uk/media/174220/data_linkage_final.pdf

Linking Data for Health Services Research: A Framework and Instructional Guide
Stacie B. Dusetzina, Seth Tyree, Anne-Marie Meyer, Adrian Meyer, and Laura Green. (2014).

This online book, published by the Agency for Healthcare Research and Quality, begins with an overview of linking data and background on the research context. It continues with two chapters with guidance on how to judge the feasibility of linking and linking methods, including a listing of public and commercial linkage software. It concludes with an evaluation of methods linking health registry data to insurance claims in scenarios of varying available information.

https://www.ncbi.nlm.nih.gov/books/NBK253312/

A Cross-Disciplinary Review of Record Linkage Methodologies
Jana L. Asher. (2017).

This presentation from the 2017 Joint Statistical Meetings in Baltimore provides a brief literature review of record linkage across disciplines, such as medical, economic, history, and survey enhancement. The author provides references related to different aspects of data linkage, including deterministic record linkage, string comparators, privacy-preserving record linkage, informed consent, temporal record linkage, differential record linkage, bayesian methodologies, accuracy, and computational intensity.

https://ww2.amstat.org/meetings/jsm/2017/onlineprogram/AbstractDetails.cfm?abstractid=323931

Regulatory and Ethical Considerations for Linking Clinical and Administrative Databases
Rachel S. Dokholyan, Lawrence H. Muhlbaier, John M. Falletta, Jeffrey P. Jacobs, David Shahian, Constance K. Haan, and Eric D. Peterson. (2009).

This article in the *American Heart Journal* focuses on the regulatory and ethical considerations that arise from the use of clinical health registry data for research, including linkage of clinical and administrative datasets. It provides a series of screens for researchers to consider, such as the purpose of the study quality assessment and improvement, research, or both; whether institutional review board review is required; circumstances for waivers of informed consent and HIPAA authorization required; and the potential for a limited dataset that would not require review. The approaches outlined in this article represent a local interpretation of the regulations in the context of several clinical data registry projects and focuses on a case study of the Society of Thoracic Surgeons National Database.

https://www.ahjonline.com/article/S0002-8703(09)00268-3/pdf

A Novel Metadata Management Model to Capture Consent for Record Linkage in Longitudinal Research Studies
Christiana McMahon and Spiros Denaxas. (2017).

This article in the *Journal of Informatics for Health and Social Care* presents a structured approach to capturing consent-related metadata. Through a systematic literature review and qualitative analysis of consent forms, the authors explore the state of the art for recording consent and identify key elements of consent required for record linkage. Finally, it presents and evaluates a metadata management model to capture consent-related metadata to facilitate the harmonization and streamlining of linkage and analysis.

https://www.tandfonline.com/doi/full/10.1080/17538157.2017.1364251

DATA DISSEMINATION

Data dissemination ranges from sharing full data files to sharing derived indicators and analysis to publishing data visualizations. Organizations like the members of NNIP have long transformed confidential administrative data into aggregate indicators to share with the community for better decisionmaking. Researchers generally share insights from integrated data systems through descriptive statistics or modeling, but innovators in the field are considering how to publish summary indicators. Pay for success projects present an opportunity to disseminate information on program performance and outcomes. In all cases, organizations should safeguard any PII or other confidential data from improper release through organizational procedures, statistical methods, and licenses.

Establishing Guidelines for Dissemination

An organization's data governance plan should state the review process for external release of any data file or analytic product. This review should include checking the terms of the data use agreement for any release or sharing conditions, such as the required approval by the data owner, the removal of PII fields, or a minimum cell size for tables or graphics. Initiatives with data governance committees also need to review the data or any derivative products before they are released.

The type and level of review should cover quality, security, and documentation and will depend on the product. Data stewards might share identifiable record-level data with internal partners according to prearranged terms, requiring signing of confidentiality statements by the receiving group. Organizations might also share deidentified data for research purposes, after evaluating the risk of identifying any person directly or indirectly. Or groups might publish aggregate analysis—charts, tables, and maps—following such guidelines as only publishing data greater than a minimum count.

Pay for success projects are often more highly scrutinized by the public than other research with administrative data because the payment mechanism is tied to results. As such, evaluation results tend to be widely publicized, with the potential to elevate the role of data and evidence in the public's consciousness. Schedule F in the Massachusetts PFS agreement on job and English language training specifies a randomization report within 60 days of each randomization cycle, evaluator interim reports annually, a final learning report, and a final outcomes report. Reports like these can spark new or contribute to ongoing conversations about evidence-based decisionmaking and data use. Other non-PFS funding models that also tie evidence of impact to payment might yield similar benefits.

DISSEMINATING EDUCATION DATA FROM IDS

In **"Integrated Data Systems and Student Privacy,"** PTAC describes the role of an IDS lead in managing access and use of identified and deidentified integrated data by external organizations and research partners. Disclosure of PII is allowed only if it is approved by the educational authority, the redisclosure is permissible under a FERPA exception, and it complies with applicable FERPA requirements and with other relevant confidentiality and privacy provisions. The governance model could not permit any redisclosures, could require a new written agreement for every redisclosure, or could allow redisclosure under the existing agreements an educational authority has with the IDS lead and other stakeholders.

https://studentprivacy.ed.gov/sites/default/files/resource_document/file/IDS-Final_0.pdf

Methods to Protect against Reidentification

Even when records have been stripped of direct identifiers or aggregated into groups, organizations must review the data to prevent the disclosure of personal identity through linkages with other data sources. The Future of Privacy Forum has a helpful visual guide to practical data deidentification to distinguish categories of data. Researchers can use several statistical methods to reduce the risk, including suppression, top-coding, collapsing variables, sampling, and swapping the data. The resources below relate to strategies for avoiding the reidentification of confidential data. Publishing individual-level geospatial data might require additional methods to reduce the risk of reidentification. In addition to the resources below, NASA's Socioeconomic Data and Applications Center compiled an extensive reference list through 2012 that relates to geospatial data and confidentiality.

Report on Statistical Disclosure Limitation Methodology
Federal Committee on Statistical Methodology. (2005).

The source, written by a committee within the Office of Information and Regulatory Affairs in the Office of Management and Budget, discusses methods of disclosure limitation used by federal agencies to protect tables and microdata, such as recoding into intervals, adding random noise,

data swapping, and targeted suppression. The source then covers the policies, practices, and procedures for statistical disclosure limitation of 14 federal agencies. The guide has detailed chapters related to confidentiality for frequency tables and for microdata files.

https://www.hhs.gov/sites/default/files/spwp22.pdf

The Modernization of Statistical Disclosure Limitation at the US Census Bureau
Aref N. Dajani, Amy D. Lauger, Phyllis E. Singer, Daniel Kifer, Jerome P. Reiter, Ashwin Machanavajjhala, Simson L. Garfinkel, et al. (2017).

The Census Bureau previously used statistical disclosure limitation (SDL) techniques, such as top-coding, bottom-coding, suppression, and rounding, but is transitioning to modern techniques based on formal privacy approaches. Formal privacy methods start with a mathematical definition of privacy and then publish queries based on the confidential data. Differential privacy is the most developed of formal privacy methods. The rest of the paper discusses SDL methods supporting the 2020 Census of Population and Housing, the American Community Survey, and the 2017 Economic Census. The paper also discusses challenges with differential privacy and discusses methods to establish data accuracy.

https://www2.census.gov/cac/sac/meetings/2017-09/statistical-disclosure-limitation.pdf

Ensuring Confidentiality of Geocoded Health Data: Assessing Geographic Masking Strategies for Individual-Level Data
Paul A. Zandbergen. (2014).

This article in the journal *Advances in Medicine* discusses address information, which is often considered confidential and is not released or shared. Publishing maps with the locations of individuals, however, might also breach confidentiality because addresses and associated identities can be discovered through reverse geocoding. One commonly used technique to protect confidentiality when releasing individual-level geocoded data is geographic masking, applying a certain amount of random perturbation in a systematic manner. This paper presents a review of the current state of the art in geographic masking techniques, summarizing the various methods and their strengths and weaknesses. Any researcher publishing such maps is advised to become familiar with the different masking techniques available and their associated reidentification risks.

https://www.hindawi.com/journals/amed/2014/567049/

Statistical Methods for Protecting Personally Identifiable Information in Aggregate Reporting
Marilyn Seastrom. (2010).

This brief from the Institute of Education Sciences at the National Center for Education Statistics discusses how the release of educational data or outcomes can protect student privacy while meeting reporting requirements. Even with minimum-group-size reporting rules and other disclosure prevention methods, the risk of identifying an individual remains when reporting detailed categories or underlying counts that could then be used to retrieve suppressed information. The article relates best practices for mitigating disclosure risk, such as suppression in publishing counts or reporting school background information without subgroup totals. The article ends with numerical recommendations for reporting rules.

https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2011603

Guidance Regarding Methods for Deidentification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule
US Department of Health and Human Services. (2015).

The source describes the two methods to satisfy the HIPAA's deidentification standard. The *expert determination* method relies on a person with knowledge of and experience with statistical and scientific principles documenting that there is not a reasonable basis to believe the information can be used to identify an individual. The *safe harbor* method identifies the identifiers that are to be removed, such as names, detailed geographies, and identification numbers. The implementation specifications also provide guidance on reidentification—that is, assigning a unique code to deidentified health information to permit the covered entity to reidentify the data.

https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html

Data Licenses

If your organization is publishing datasets, you should determine the intellectual property rights for any data you publish and consider whether you want to use a license to specify the conditions of use. The "Introduction to Intellectual Property Rights in Data Management" summarizes issues associated with managing intellectual property rights in data projects.

Conditions in a license include the requirements for citation, the ability to modify and republish the data, and any restrictions on commercial use. Licenses for sensitive or confidential data can add other conditions of use, such as agreeing not to attempt to identify any person in the record, maintaining safeguards to prevent the use or disclosure of the dataset, or reporting any breaches in confidentiality.

The organization Open Data Commons maintains a list of licenses for open data, which specify any requirements to provide attribution and the terms of reuse. These are intended for information published by individuals or nongovernmental entities. The Sunlight Foundation believes that data created by state and local governments should be released into the public domain and that any license presents a barrier to the reuse of public information.

DATA DISPOSITION

Disposition of a dataset at the end of a project should follow conditions of any data agreement, state or local regulations, and your organization's data retention policy. The National Conference of State Legislatures has compiled a list of state data disposal laws.

When data deletion or destruction is required, the resources below provide information about the procedures for different media formats. Or you will need to determine if your organization wants to archive the data internally or with an external repository. Factors for determining whether to archive data include legal requirements, the uniqueness and replicability of the

data, the relevance to organizational mission, future historical or scientific value, and the costs of retaining the data (Whyte and Wilson 2016).

Guidelines for Media Sanitization
Richard Kisel, Andrew Regenscheid, Matthew School, and Kevin Stine. (2014).

Media sanitization renders access to target data on the media infeasible for a given level of effort. This guide from the National Institute of Standards and Technology will assist organizations and system owners in making practical sanitization decisions based on the categorization of confidentiality of their information.

http://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-88r1.pdf

Best Practices for Data Destruction
Privacy Technical Assistance Center. (2016).

The Data Destruction Document offers best practices on properly destroying sensitive student data after it is no longer needed. It details the life cycle of data and discusses various legal requirements relating to the destruction of data under FERPA and examines various methods for properly destroying data. The guide also provides real-world examples of how to implement it within your organization.

https://studentprivacy.ed.gov/resources/best-practices-data-destruction

How to Appraise and Select Research Data for Curation
Angus Whyte and Andrew Wilson. (2016).

This Digital Curation Centre guide will help you develop a managed approach to appraising and selecting datasets for curation. It provides working knowledge of current approaches, issues, and challenges and of the roles of research groups and institutional data services in addressing these.

http://www.dcc.ac.uk/resources/how-guides/appraise-select-data

BIBLIOGRAPHY

Carlson, Neil E., Edwin Hernández, Chaná Edmond-Verley, Gustavo Rotondaro, Eleibny Feliz-Santana, and Susan Heynig. 2011. "Developing a Master Data-Sharing Agreement: Seeking Student-Level Evidence to Support a Collaborative Community Effort in Education." *Foundation Review* 3(4): 14–33.

Coleridge Initiative. n.d. "Security Model." New York: New York University Center for Urban Science and Progress, Coleridge Initiative. https://coleridgeinitiative.org/assets/docs/security_whitepaper.pdf.

Gibbs, Linda, Amy Hawn Nelson, Erin Dalton, Joel Cantor, Stephanie Shipp, and Della Jenkins. 2017. *IDS Governance: Setting Up for Ethical and Effective Use*. Philadelphia: University of Pennsylvania, Actionable Intelligence for Social Policy. https://www.aisp.upenn.edu/wp-content/uploads/2016/07/Governance.pdf.

National Research Council. 1993. *Private Lives and Public Policies: Confidentiality and Accessibility of Government Statistics*. Washington, DC: National Academies Press. https://doi.org/10.17226/2122.

Petrila, John, Barbara Cohn, Wendell Pritchett, Paul Stiles, Victoria Stodden, Jeffrey Vagle, Mark Humowiecki, et al. 2017. *Legal Issues for IDS Use: Finding a Way Forward*. Philadelphia: University of Pennsylvania, Actionable Intelligence for Social Policy.

PTAC (Privacy Technical Assistance Center). 2012. "Data Breach Response Checklist." Washington, DC: US Department of Education, PTAC. https://studentprivacy.ed.gov/sites/default/files/resource_document/file/checklist_data_breach_response_092012_0.pdf.

———. 2015a. "Data Governance and Stewardship." Washington, DC: US Department of Education, PTAC. https://studentprivacy.ed.gov/sites/default/files/resource_document/file/Data_Governance_and_Stewardship_0.pdf.

———. 2015b. "Data Security Checklist." Washington, DC: US Department of Education, PTAC. https://studentprivacy.ed.gov/sites/default/files/resource_document/file/Data%20Security%20Checklist_0.pdf.

———. n.d. "Data Security Checklist." Washington, DC: US Department of Education, PTAC. https://nces.ed.gov/programs/ptac/pdf/ptac-data-security-checklist.pdf.

Stiles, Paul G., and Roger A. Boothroyd. 2015. *Ethical Use of Administrative Data for Research Purposes*. Philadelphia: Actionable Intelligence for Social Policy. https://www.aisp.upenn.edu/wp-content/uploads/2015/09/0033_12_SP2_Ethical_Admin_Data_001.pdf. Strive Together. 2015. "Student Data Privacy Best Practices: Five Ways Community Organizations Can Ensure Effective and Responsible Data Use." Cincinnati: Strive Together.

Whyte, Angus, and Andrew Wilson. 2016. "How to Appraise and Select Research Data for Curation." Edinburgh: Digital Curation Centre. http://www.dcc.ac.uk/resources/how-guides.

VERSION UPDATES

Version 1.0: Initial publication.

NNIP is a collaboration between the Urban Institute and partner organizations in more than thirty American cities. NNIP partners democratize data: they make it accessible and easy to understand and then help local stakeholders apply it to solve problems in their communities.

For more information about NNIP, go to www.neighborhoodindicators.org or email nnip@urban.org.