Overcoming Barriers to Data Sharing in the United States

By Gillian Diebold | September 25, 2023

Both public and private sector actors face legal, social, technical, and economic barriers to data sharing in the United States, inhibiting much-needed innovation and discoveries. Overly restrictive data privacy laws and a lack of technical standards hinder sector-specific data sharing in fields such as education and health care, and the misfire of past experiments has led to both a lack of trust and data siloes. This report details the challenges associated with data sharing and the steps U.S. policymakers can take to overcome these barriers and bring the social and economic benefits of data to all Americans.

There are countless facets of the economy and society that could be improved with better data. Data enables people and organizations to better understand the world and use that understanding to make better decisions, large and small. Better data would help researchers understand how to best treat infectious diseases and which interventions are most likely to alleviate poverty. Better data would allow scientists to improve predictions about extreme weather events and natural disasters. And better data would enable educators to understand which pedagogical practices work best for which kinds of students.

But better data requires more data sharing, and getting the right data to the right place at the right time is not always easy. For example, one government agency might need data held by a different government agency or a firm in the private sector. Organizations may need to transfer, aggregate, or combine datasets before they can use or reuse data. However, legal, social, technical, and economic barriers may impede data sharing. When organizations cannot obtain data already collected by

another organization, they must either proceed without it (leading to suboptimal services) or collect it again (creating duplicative costs eventually passed on to consumers and taxpayers, as well as creating an onslaught of additional requests for personal information for individuals). Moreover, continued obstacles to data sharing can greatly inhibit the burgeoning AI economy. For example, the potential of large language models is only as great as their training data. Effective data-sharing mechanisms are therefore essential for individuals and organizations to overcome these barriers and obtain the social and economic benefits of data.

While many organizations in the United States do share data, whether it be internally, via set agreements with other parties, or even via data brokers, more is still needed, particularly in high-value areas. Certain parts of the economy, including health care, financial services, and education, share less data than they could despite the potential for data-driven innovation. This is due to a variety of challenges that come with data sharing. For example, privacy laws in some sectors, such as HIPAA (Health Insurance Portability and Accountability Act) in health care, tend to be more restrictive rather than enabling, leading organizations to shy away from sharing information to avoid the risk of penalties for noncompliance. Likewise, anti-data advocates have fueled fears and mistrust about data sharing, creating an environment wherein people are averse to data sharing. Moreover, data sharing can be costly to the participating actors and can require complex technical components that under-resourced areas are unlikely to prioritize.

Without policy change, the United States will continue trending toward data siloes—an inefficient world in which data is isolated, and its benefits are restricted. Data siloes are repositories of information that exist in a closed system, often sealed off from the rest of an organization or other organizations and incompatible with other datasets.[1] Data sharing spans a whole spectrum of possibilities: on one end are data siloes, where data remains isolated and unshared, and on the other end are data collaboratives, where data flows freely between organizations with no restrictions on use. The United States needs to move more toward data collaboratives, and doing so will require overcoming these legal, social, technical, and economic barriers. It will take coordinated government action to both enable data sharing by default and counter pervasive privacy fears. Specifically, policymakers should:

- reform existing data protection laws to reduce legal barriers to data sharing;
- direct key federal agencies to create model data-sharing contracts to simplify legal agreements;
- create data literacy initiatives to help communities understand the benefits of data and how data can be shared securely;

- enable consumers to easily donate their data, particularly in high-impact areas such as health care and education;
- develop data standards in high-impact areas; and
- identify and address instances where fragmented ownership of data prevents compiling valuable datasets.

BARRIERS TO SHARING DATA

The United States faces legal, social, technical, and economic barriers to more widespread data sharing. A patchwork of consumer data protection laws at the state level coupled with overly restrictive national sector-specific laws hinders data sharing necessary for data-driven innovation in in fields such as health care and education. Moreover, privacy fanatics have embedded a number of myths about data sharing into society, leading to social skepticism. Collective action problems also plague data sharing and require a new set of incentives for industry actors. Finally, not all organizations are technically equipped to share data effectively, and a lack of national standards for data and metadata formatting continues to inhibit sharing across organizations.

Legal Barriers

Protecting the privacy of personal information has long been one of the primary motivations for laws about data in the United States and around the world. As a result, most legal barriers relating to data sharing revolve around privacy. Some legal barriers prevent government data sharing, while others prevent data sharing in fields such as health care or education, or data about children. To be clear, data collected in these areas is often highly sensitive in nature and does require adequate protection. But sensitive data also often has the greatest potential to create widespread social and economic benefits.

Legal barriers often create jurisdictional siloes, which can make it difficult to combine and aggregate data across governmental agencies and develop targeted programs. For example, the Department of Housing and Urban Development (HUD) collects a large amount of data for all Continuum of Care (COC) programs nationwide.[2] But each COC program collects its own data in a homeless management information system. COCs often cannot share that data with HUD due to jurisdictional rules and anonymization challenges.[3]

Moreover, data sharing in the United States often requires consent from individuals, the legal requirements of which can either be extensive or, at times, largely unclear. The United States lacks a uniform definition of consent, with a patchwork of different federal and state laws using varying definitions. Both scenarios can inhibit data sharing due to compliance costs or liability concerns.[4] Both government and private organizations

must obtain direct consent from individuals and, in some instances, get renewed consent, which can be a resource-intensive process.

Table 1 demonstrates the key provisions restricting data sharing in U.S. federal law. In particular, federal privacy laws mostly focus on protecting data collected by certain types of organizations, including schools, financial institutions, health care providers, and government agencies. For example, the Family Educational Rights and Privacy Act (FERPA) provides rules for institutions sharing educational data with third parties in order to protect student privacy.[5]

Table 1: Federal laws restricting data sharing

Law	Focus	Provision
Privacy Act of 1974	Inter-agency data sharing	"No agency shall disclose any record which is contained in a system of records by any means of communication to any person, or to another agency, except pursuant to a written request by, or with the prior written consent of, the individual to whom the record pertains, unless disclosure of the record would be..."[6]
Gramm-Leach-Bliley Act (GLBA)	Financial services	"A financial institution may not, directly or through any affiliate, disclose to a nonaffiliated third party any nonpublic personal information, unless such financial institution provides or has provided to the consumer a notice that complies with section 503..."[7]
Health Insurance Portability and Accountability Act (HIPAA)	Health care	"A covered entity or business associate may not use or disclose protected health information except as permitted or required by this subpart or by subpart C of part 160 of this subchapter."[8]
Family Educational Rights and Privacy Act (FERPA)	Education	"An educational agency or institution may disclose personally identifiable information from an education record only on the condition that the party to whom the information is disclosed will not disclose the information to any other party without the prior consent of the parent or eligible student."[9]

Law	Focus	Provision
Children's Online Privacy Protection Rule (COPPA)	Online services directed at children under 13	"(1) An operator is required to obtain verifiable parental consent before any collection, use, or disclosure of personal information from children, including consent to any material change in the collection, use, or disclosure practices to which the parent has previously consented…(2) An operator must give the parent the option to consent to the collection and use of the child's personal information without consenting to disclosure of his or her personal information to third parties."[10]

As demonstrated in Table 1, there are a number of U.S. federal laws that restrict data sharing. While these laws all include exceptions, they differ in the exact extent and stringency of restrictions. For example, the Privacy Act of 1974 restricts data sharing between government agencies without written consent from an individual or in the case of 12 exceptions. These exceptions include the following:

- Need to know within agency
- Required FOIA disclosure
- Routine uses
- Bureau of the Census
- Statistical research
- National Archives
- Law enforcement request
- Health or safety of an individual
- Congress
- Government Accountability Office
- Court order
- Debt collection act

Most importantly, the Privacy Act restricts "matching programs," or cross-comparison of agency databases in order to determine eligibility for federal

benefit programs. Agencies may only match records after notification to an individual and give them the opportunity to contest the accuracy of the information being used.[11] Of the exemptions to the Privacy Act, agencies frequently use the "routine use" exemption to share information not only between executive agencies but also with law enforcement.[12]

GLBA, HIPAA, FERPA, and COPPA all focus on specific areas for data collection. Like the Privacy Act, they all generally adhere to limited disclosures by default, but include some exemptions that allow data sharing in limited circumstances. For example, HIPAA allows the use of data for 12 national priority purposes, such as public health activities.[13] Additionally, most privacy laws include some form of exemption for law enforcement access to data.

At the same time, these laws can have unintended consequences. For example, HIPAA includes a "minimum necessary" requirement, meaning health care professionals should not share any more health data that is necessary for a specific purpose or function.[14] This type of minimization clause can inadvertently end up hindering important research in areas such as genomics, where certain races and ethnicities are already underrepresented in key databases.[15]

Likewise, restrictions within FERPA limit the access of researchers, including government officials, without prior consent to educational records that contain health information.[16] This can include mental health evaluations, which in turn can restrict research and understanding about things such as autism spectrum disorders and Attention-deficit/hyperactivity disorder (ADHD). With COPPA, these obligations can be restrictive even in the case of necessary or highly beneficial applications, such as remote learning with edtech.[17]

While the United States still lacks a comprehensive federal privacy law, legislative proposals such as the American Data Privacy and Protection Act (ADPPA) still pose a barrier to data sharing. First, most legislative proposals would add additional data privacy rules for consumer data, but they would not reduce or harmonize existing sector-based data protection rules such as HIPAA or FERPA. Second, they would expand some of the existing restrictions on data sharing to more types of data. For example, the Congressional Research Service found that ADPPA would include a "minimum necessary" clause for data sharing, similar to that of the HIPAA privacy rule.[18] Stipulations such as opt-in consent, data minimization, and purpose specification requirements are designed to limit, rather than facilitate, the use and sharing of data. These types of requirements protect individual interests in data but do nothing to advance societal ones. Despite these drawbacks, one comprehensive piece of legislation for privacy would alleviate the challenges associated with a patchwork of state privacy laws and give one clear set of rules for public and private organizations to adhere to. But if it is not designed to enable appropriate data sharing, it will hinder the emergence of a smart society.

Social Barriers

Social support and opposition play a major role in policy. There has been growing mistrust of data collection and hostility to technology as a whole, which affects legislation and, in turn, the ability of organizations to embark on data-sharing ventures.[19] This lack of trust stems from a variety of sources, including antidata coalitions in in the United States and abroad, as well as from a lack of transparency in historical data-sharing projects.

Opposition to the "datafication" of society has spread throughout the United States, where growing animus toward Big Tech has also led to increased calls for restrictive privacy legislation.[20] Importantly, people who distrust any organization that handles their data (i.e., privacy fundamentalists) are taking advantage of the latest antitech narrative to bolster their opposition and create a political climate that is hostile to organizations using data even for positive purposes, such as for addressing pandemics or finding missing children.[21] Some members of Congress, such as Senator Ed Markey (D-MA), seek to ban certain data collection activities, such as government use of biometric technologies on the basis of surveillance capacity and potential for discrimination. Overall, these pervasive attitudes can have far-reaching effects, despite the fact that most Americans are "privacy pragmatists" who are willing to make trade-offs between privacy and personal benefit.[22]

The potential for social backlash might inhibit companies from sharing the data they have, even if it is useful to others. For example, a large credit card company's dataset of user transactions can be immensely useful to the U.S. government for understanding the American economy at a certain point in time. But the threat of privacy backlash might deter that credit card company from engaging in a collaborative effort because it could be accused of surveilling its users. As the case of InBloom in box 1 demonstrates, the loudest voices in a debate often prevail.

Box 1: InBloom's enduring legacy

In 2011, the education technology space explored the possibility of a large-scale collaborative platform to aggregate educational data across the United States. Spearheaded by the Bill and Melinda Gates Foundation, the education data trust project known as InBloom had over $100 million in funding and support, yet the project shut down within a year of its launch.[23] What went wrong?

The InBloom platform was intended to be a centralized platform for data sharing and curricula that addressed the challenges of data siloes that prevented the interoperability of school datasets. The platform would create shared data standards. It would also create more opportunities for new vendors to enter the edtech space with the ultimate goal of improving learning outcomes for students nationwide. However, a number of factors led to public backlash and the project's

failure, namely entrenched privacy concerns and hostility toward data-driven education.[24] The project had a glaring lack of community dialogue, which only exacerbated concerns about transparency and accountability. Moreover, opposition groups reacted harshly to the swift pace of the project, raising concerns about student data use, including the potential sale of such data to third parties for targeted advertising. The truth of the matter became irrelevant, and InBloom's key stakeholders pulled out within the year.

InBloom provides an important insight into data sharing in the education space today. As a result of the project's implosion, edtech has trended toward a patchwork of closed, proprietary data systems. Privacy advocates continue to use InBloom as a reason to avoid collective data-sharing models. Yet, the closed systems of education data today perpetuate the same lack of transparency supposedly promoted by InBloom.

Technical Barriers

Sometimes, data sharing doesn't occur due to technical barriers that transcend any social obstacle or legal restriction. Data sharing often requires infrastructure that can receive, aggregate, and analyze data from a variety of sources in multiple formats. A lack of universal standards for data formatting can hinder the interoperability and useability of shared data. Moreover, issues with data quality can exacerbate interoperability challenges; and a lack of standards for metadata also contributes to the technical challenges of data sharing.

Data interoperability ensures that different services can exchange and use information together.[25] It requires that services sharing data understand and cooperate with each other. For example, interoperability means a patient can move between health care systems that may have different technical infrastructures without losing access to their electronic health record.[26] As such, data sharing requires compatible standards for data formatting beyond a baseline of machine readability. A lack of common standards is a major obstacle to data aggregation and the creation of longitudinal datasets.[27] For example, that lack can get in the way of understanding communities' resilience to natural disasters when different agencies measure variables such as debris flow, smoke, or drought differently.[28] This standards issue also applies to metadata, or the description of the dataset that includes the content of the data, its origin, and collection methods. Inconsistent metadata formatting can also limit secondary use and the interoperability of the dataset.

When data standards are in place, data-driven technologies can easily be transferred and replicated at different scales. For example, Google partnered with the Tri-County Metropolitan Transportation District of Oregon in 2005 to develop the General Transit Feed Specification (GTFS).[29] GTFS is a common format for public transportation schedules that

integrates with third parties such as Google Maps to help cities integrate their transportation data with services that are popular with consumers.

Economic Barriers

Economic barriers to data sharing include questions of cost and ownership. Different actors interpret the advantages and disadvantages of data sharing differently and act accordingly. An organization may not be ideologically opposed to data sharing so much as it perceives withholding data to be in its best interest, whether it be for profitability or intellectual property, or any other reason.

Data sharing poses a collective action problem. Collective action problems exist in economics when individuals pursue their self-interests at the expense of the interests of others (and ultimately their own).[30] Such a problem appears in a variety of policy areas; the "tragedy of the commons" frequently appears as a well-known collective action problem in which individual actors overuse limited resources, such as grazing pastures or fisheries, which gives them short-term gains but leads to long-term depletion of the resource and social harm.[31] Along those lines, data sharing represents the tragedy of the anticommons, in which actors under-use or under-contribute to a limitless resource due to poor incentives to share.[32] In the case of data, too many individual data holders can create a coordination problem that makes it difficult to pool data, preventing the creation of valuable data resources that could have enormous potential for innovation and profitability.

Additionally, the cost of making data public can exceed the value a firm gets from the open data, despite the societal benefits. In the short term, organizations need to account for costs such as data storage, processing, infrastructure, and regulatory compliance.[33] Setting up the actual infrastructure for data sharing can require a large investment in things such as servers, network equipment, and cloud services. Likewise, data must be prepared before it can be shared, which opens another vein of costs. Anonymizing data to protect individual privacy and maintain its utility can be a complex process and often requires advanced techniques that demand their own software and expertise. And missteps in this process can expose a firm to liability and reputational damage.[34] The exact cost of data sharing depends on firm size, sector, and the actual sharing mechanisms involved.

RECOMMENDATIONS

Policymakers should take steps to address the legal, social, technical, and economic challenges that limit data sharing in the United States.

Reform Existing Data Protection Laws to Reduce Legal Barriers to Data Sharing

Many federal data privacy laws in the United States inhibit data sharing, even when it benefits the data subject. Although much political focus has been on strengthening privacy, data protection laws also need to enable greater use of data. Congress should create a bipartisan, bicameral taskforce to identify improvements to existing sectoral data protection laws that would enable greater data sharing. For example, the taskforce should examine how to revise the HIPAA Privacy Rule to improve health data sharing, how to reform FERPA to improve data sharing for educational research, and how to update the Privacy Act of 1974 to allow more government data sharing.

The taskforce should also identify opportunities to streamline and harmonize data-sharing provisions within existing sectoral privacy laws, such as those giving individuals the right to access their data or move it to other services. Privacy legislation should not restrict users from moving and using their own data, and the taskforce should also take note of whether existing legislation includes unnecessarily restrictive requirements such as data retention limits, data minimization requirements, or purpose specification requirements. Given that enabling secondary use is one of the benefits of data sharing, requirements such as purpose specification ultimately restrict that type of innovation. In addition, the taskforce should evaluate the impact of state-level data protection laws on data sharing. Many states have passed additional privacy laws that limit data sharing for particular types of data, such as biometric data, or in particular sectors, such as health care and education.[35] The taskforce should propose legislation to preempt state privacy laws to ensure harmonization among states.

Direct Agencies to Create Model Data-Sharing Contracts

To alleviate some of the legal barriers to data sharing, policymakers in federal agencies with stringent data protection laws such as the Department of Education (DOE) and the Department of Health and Human Services (HHS) should develop model contracts organizations such as health companies or edtech firms can adopt for data sharing within the confines of their respective privacy laws. Given that these laws can create inefficiencies for organizations and even deter data-sharing activities, offering templates for data sharing within a specific sector can help ease the contract and negotiations process and enhance collaboration. For example, an organization interested in sharing educational data would be able to adopt the template provided by DOE and customize as needed, accounting for things such as terms of retention, while ensuring compliance with laws such as FERPA and COPPA.

Create Data Literacy Initiatives

Communities may resist data-sharing initiatives if they do not understand how greater use of data benefits them or how organizations protect data. Federal agencies engaging in high-impact data projects should create data literacy initiatives to answer these questions for impacted communities. Ideally, these agencies should work with local organizations, businesses, and educational institutions to provide data literacy programs tailored to the needs of impacted communities. By tailoring such programs, agencies can ensure that data literacy programs effectively raise public awareness and promote positive perceptions of data and data sharing as a whole. For example, DOE should address data literacy topics relevant to parents and educators, whereas the Department of Veterans Affairs should address topics relevant to those who served in the armed forces. Improving data literacy will play an important role in building a positive culture around data.

Enable Data Donation

Most data protection laws and regulations focus on reducing data sharing, such as requiring consumers to opt in before data collection can occur or providing consumers with details on how to opt out of data sharing. But few, if any, data protection laws and regulations focus on increasing data sharing, such as by encouraging consumers to donate their data for beneficial purposes. In effect, these policies encourage Americans to be selfish with their data. Policymakers should provide an opportunity for individuals to be altruistic with their data, particularly in high-value areas such as health care and education.

For example, under the Gramm-Leach-Bliley Act (GLBA), financial institutions must provide annual privacy notices to their customers outlining how their data is shared with third parties and provide consumers an opportunity to opt out of the sharing.[36] But there are no similar policy mechanisms directing financial institutions to alert consumers of how they can increase data sharing. The Consumer Financial Protection Bureau should recommend an amendment to GLBA that includes a positive opportunity for consumers to donate their data, instead of only instructing them on how to opt out. For example, researchers could use this data to better analyze income levels for specific groups of college graduates. Likewise, HHS should require all certifiable electronic health record systems to give patients the option to donate their data to third-party medical research. Congress should direct key federal agencies to create data donation policies that allow individuals to voluntarily contribute their personal data to third parties.

Develop Data Standards in High-Impact Areas

Data standards are essential to facilitate data sharing. For some types of data, there are mature and well-developed standards; however, there are many areas where data standards do not exist, resulting in different

organizations collecting similar data they cannot easily share due to technical limitations or the costs necessary to clean and integrate different datasets. To encourage more economic efficiency, the National Institute of Standards and Technology (NIST) should offer guidance or training to organizations interested in launching new data standards in high-impact areas. In some cases, when there is no organization available to take on the responsibility of coordinating between different stakeholders, NIST should also consider providing grants for a new or existing organization to take on that activity. By creating data standards, NIST can facilitate data sharing between different organizations.

Identify and Address Instances of Data Ownership Fragmentation

As discussed previously, when too many people own part of something, they can prohibit others from using it, creating the "tragedy of the anti-commons." In the case of data, when there are too many stakeholders that own part of a potential dataset, they can prevent that dataset from being created. Economic theory posits that unification of fragmented ownership can solve this problem, often through a central authority, including through regulation, nationalization, or eminent domain.[37] For example, in the case of data, in the United States public health laws require certain facilities to report data on infectious diseases (rather than leaving it up to each facility to decide whether they wish to voluntarily contribute this information).[38]

The European Union's Data Act aims to ameliorate the data fragmentation issue in cases of "exceptional need" by overriding individual interests and requiring companies to turn over their data to public institutions, with some compensation.[39] While the EU's actions may be seen as an extreme measure, U.S. policymakers should seek to identify instances where data fragmentation is occurring and investigate potential solutions. For example, HHS might identify more instances where healthcare providers should allow third-party access to electronic health records for high-impact medical research without obtaining prior authorization. In addition, federal agencies should explore how they can work with the private sector to create more data-sharing consortia (where participants in the consortia share their data in exchange for access to data from the rest of the consortia). Data-sharing consortia create an incentive mechanism wherein, if enough participants join, it is better to be part of the group than not be a part of it.

CONCLUSION

Data-driven innovation offers enormous potential in many sectors, including agriculture, education, energy, health care, public safety, transportation, and so much more. There is more data collected now than ever before, yet much of this information is not put to productive use in the United States because of legal, social, technical, and economic barriers.

Enabling data sharing is critical to building a smart society in America. But a plethora of barriers stand in the way and require widespread, complex considerations. Without definitive action to amend privacy laws, overcome social opposition, and address economic and technical barriers to foster data sharing across government and industry, the United States will remain far behind its potential in using data for social and economic benefit, and many initiatives to use data for productive purposes will fall short.

REFERENCES

1. "Data Silo," TechTarget, accessed August 2023, https://www.techtarget.com/searchdatamanagement/definition/data-silo.
2. "Continuum of Care Program," U.S. Department of Housing and Urban Development, last modified July 2023, https://www.hud.gov/program_offices/comm_planning/coc.
3. Tiffany Fishman et al., "Disrupting housing insecurity and homelessness' (Deloitte, January 2022), https://www.deloitte.com/global/en/our-thinking/insights/industry/government-public-services/health-human-services-innovations-reform/homelessness-housing-insecurity-challenges.html.
4. Deja Kemp, Amy Nelson, and Della Jenkins, "Yes, No, Maybe? Legal & Ethical Considerations for Informed Consent in Data Sharing and Integration' (Actionable Intelligence for Social Policy, University of Pennsylvania, 2023), https://aisp.upenn.edu/wp-content/uploads/2023/05/Consent-Brief-Appx.pdf.
5. "Data Sharing," Protecting Student Privacy, January 2020, https://studentprivacy.ed.gov/content/data-sharing?page=1.
6. Privacy Act of 1974, 5 U.S.C. § 552a, https://www.govinfo.gov/content/pkg/USCODE-2018-title5/pdf/USCODE-2018-title5-partI-chap5-subchapII-sec552a.pdf.
7. Gramm-Leach-Bliley Act, S.900, 106th Congress, (1999) https://www.congress.gov/bill/106th-congress/senate-bill/900.
8. Health Insurance Portability and Accountability Act (HIPAA), H.R. 3103 (1996), https://www.hhs.gov/sites/default/files/ocr/privacy/hipaa/administrative/combined/hipaa-simplification-201303.pdf.
9. Family Educational Rights and Privacy Act, (1974), https://www.law.cornell.edu/cfr/text/34/99.30.
10. Children's Online Privacy Protection Rule, 5 U.S.C. 6501, (1998), https://www.ecfr.gov/current/title-16/chapter-I/subchapter-C/part-312#312.3.
11. Natalie Ortiz, "Computer Matching and Privacy Protection Act: Data Integration and Individual Rights' (Washington, D.C.: Congressional Research Service, December 2022), https://crsreports.congress.gov/product/pdf/R/R47325.
12. "Overview of the Privacy Act: 2020 Edition," Office of Privacy and Civil Liberties, U.S. Department of Justice, last modified December 2022, https://www.justice.gov/opcl/overview-privacy-act-1974-2020-edition/disclosures-third-parties#exceptions.
13. Health Insurance Portability and Accountability Act (HIPAA), H.R. 3103 (1996), https://www.hhs.gov/sites/default/files/ocr/privacy/hipaa/administrative/combined/hipaa-simplification-201303.pdf.
14. "Minimum Necessary Requirement," U.S. Department of Health and Human Services, last modified April 4, 2003, https://www.hhs.gov/hipaa/for-professionals/privacy/guidance/minimum-necessary-requirement/index.html.

15. Amy R. Bentley, Shawneequa Callier, and Charles N. Rotimi, "Diversity and Inclusion in Genomic Research: Why the Uneven Progress?" *Journal of Community Genetics* (2017), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5614884/.

16. Daniel Wartenberg and W. Douglas Thompson, "Privacy Versus Public Health: The Impact of Current Confidentiality Rules," *American Journal of Public Health*, https://ajph.aphapublications.org/doi/full/10.2105/AJPH.2009.166249.

17. Christopher Savage, "FTC Warns EdTech Vendors About COPPA Compliance," June 2, 2022, https://www.dwt.com/blogs/privacy--security-law-blog/2022/06/ftc-coppa-policy-edtech-student-data.

18. Jonathan Gaffney, Chris Linebaugh, and Eric Holmes, "Overview of the American Data Privacy and Protection Act, H.R. 8152' (Washington, D.C.: Congressional Research Service, August 2022), https://crsreports.congress.gov/product/pdf/LSB/LSB10776.

19. Element AI, "Data Trusts: A new tool for data governance' (ElementAI), https://hello.elementai.com/rs/024-OAQ-547/images/Data_Trusts_EN_201914.pdf.

20. Robert D. Atkinson et al., "A Policymaker's Guide to the 'Techlash'—What It Is and Why It's a Threat to Growth and Progress' (ITIF, October 2019), https://itif.org/publications/2019/10/28/policymakers-guide-techlash/.

21. Daniel Castro, "Privacy Fundamentalists Don't Care About the Privacy Preferences of the Silent Majority" (ITIF, January 28, 2020), https://itif.org/publications/2020/01/28/privacy-fundamentalists-dont-care-about-privacy-preferences-silent-majority/.

22. Ibid.

23. Benjamin Herold, "inBloom to Shut Down Amid Growing Dat Privacy Concerns," *Education Week,* April 21, 2014, https://www.edweek.org/technology/inbloom-to-shut-down-amid-growing-data-privacy-concerns/2014/04.

24. Monica Bulger, Patrick McCormick, and Mikaela Pitcan, "The Legacy of InBloom' (working paper, Data & Society, February 2017), https://datasociety.net/pubs/ecl/InBloom_feb_2017.pdf.

25. Daniel Castro, "Improving Consumer Welfare with Data Portability' (Center for Data Innovation, November 2021), https://www2.datainnovation.org/2021-data-portability.pdf.

26. Shannon Flynn, "Allowing Solutions to Speak to One Another – The Importance of Interoperability," IEEE Computer Society, September 2, 2022,https://www.computer.org/publications/tech-news/trends/importance-of-interoperability.

27. "Risks and challenges of data access and sharing," in *Enhancing Access to and Sharing of Data* (OECD, November 2019), https://www.oecd-ilibrary.org/sites/15c62f9c-. en/index.html?itemId=/content/component/15c62f9c-en.

28. Shefali Jenja Lakhina, Melanie Gall, and Monica Sanders, "Increasing National Resilience Through an Open Disaster Data Initiative' (Federation of American Scientists, June 2023), https://fas.org/publication/increasing-national-resilience-through-an-open-disaster-data-initiative/.

29. "GTFS: Making Public Transit Data Universally Accessible," accessed July 2023, https://gtfs.org/.

30. Keith Dowding, "collective action problem," Britannica, https://www.britannica.com/topic/collective-action-problem-1917157.

31. Elinor Ostrom, *Governing the Commons: The Evolution of Institutions for Collective Action* (Cambridge University Press, 1990), https://wtf.tw/ref/ostrom_1990.pdf.

32. "Tragedy of the anti-commons," Open Data Handbook, accessed August 2023, https://opendatahandbook.org/glossary/en/terms/tragedy-of-the-anti-commons/.

33. Moritz Godel, Ryan Perkins, and Clio von Petersdorff, "Research into the cost considerations of data sharing' (London Economics, June 2022), https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1124593/London_Economics_-_cost_considerations_of_data_sharing_-_June_2022.pdf.

34. Michael Arrington, "AOL Proudly Releases Massive Amounts of Private Data," *Tech Crunch*, August 6, 2006," https://techcrunch.com/2006/08/06/aol-proudly-releases-massive-amounts-of-user-search-data/.

35. Ashley Johnson, "The Impending Patchwork of Privacy Is Bad for Business and Consumers" (ITIF, March 27, 2023), https://itif.org/publications/2023/03/27/the-impending-patchwork-of-privacy-is-bad-for-business-and-consumers/.

36. 12 CFR Part 1016 (Regulation P) Subpart A - Privacy and Opt Out Notices, §1016.5, https://www.consumerfinance.gov/rules-policy/regulations/1016/5/.

37. Michael Heller, "The Tragedy of the Anticommons" in *The Wealth of the Commons: A World Beyond Market and State* (essay, 2008), https://wealthofthecommons.org/essay/tragedy-anticommons

38. "How We Conduct Case Surveillance," Centers for Disease Control and Prevention, National Notifiable Diseases Surveillance System (NNDSS), last modified May 2022, https://www.cdc.gov/nndss/about/conduct.html.

39. "Data Act," European Commission, last modified December 2022, https://digital-strategy.ec.europa.eu/en/library/data-act-proposal-regulation-harmonised-rules-fair-access-and-use-data.

ABOUT THE AUTHOR

Gillian Diebold was a policy analyst at the Center for Data Innovation. She holds a B.A. from the University of Pennsylvania, where she studied Communication and Political Science.

ABOUT THE CENTER FOR DATA INNOVATION

The Center for Data Innovation studies the intersection of data, technology, and public policy. With staff in Washington, London, and Brussels, the Center formulates and promotes pragmatic public policies designed to maximize the benefits of data-driven innovation in the public and private sectors. It educates policymakers and the public about the opportunities and challenges associated with data, as well as technology trends such as open data, artificial intelligence, and the Internet of Things. The Center is part of the Information Technology and Innovation Foundation (ITIF), a nonprofit, nonpartisan think tank.

Contact: info@datainnovation.org

datainnovation.org