CENTER FOR DATA INNOVATION

Exploring Data-Sharing Models to Maximize Benefits From Data

By Gillian Diebold | October 16, 2023

Data-driven innovation has the potential to be a massive force for progress. Data sharing enables organizations to increase the utility and value of the data they control and gain access to additional data controlled by others. This report evaluates the advantages and disadvantages of six common data-sharing models and offers recommendations for policymakers to promote greater uptake of these data-sharing models to maximize the economic and social benefits of data in the United States.

Individuals and organizations use data to make better decisions and obtain better insights, leading to benefits in a broad range of areas.[1] But to use data to its fullest potential, individuals and organizations need to be able to combine, augment, and analyze information from different sources. In the private sector, data sharing enables businesses to innovate with partners, such as tackling common challenges and providing better experiences to consumers. In the public sector, it enables government agencies to build on information collected by other organizations to make better decisions, offer personalized services, engage in evidence-based policymaking, and glean new insights. And among academics and nonprofit organizations, data sharing advances scientific breakthroughs and enables data for social good.

But attaining these benefits requires enabling data sharing to its fullest potential so that those who can use data productively have access to it. Unlike most resources, such as land or oil, data is nonrivalrous, meaning the supply of data is not reduced when others use it. Data can be used multiple times and in multiple ways by various entities without being depleted.[2] While many organizations in the United States share data in certain instances, many of these initiatives are ad hoc, and there are few best practices for sharing data. If policymakers want the United States to have a robust AI and data-driven society, they need to take steps to boost data sharing.

This report offers a step in that direction by evaluating the benefits and drawbacks of six different data-sharing models and offers recommendations on how policymakers can implement or expand the use of certain models. Given that different models serve different needs, policymakers do not need to pick a one-size-fits-all solution, but rather should facilitate the adoption of multiple data-sharing mechanisms in their pursuit of a data-driven society.

DATA-SHARING MODELS

Data sharing is the process of making data accessible to others, whether it be between or within organizations or between individuals and organizations. Approaches to data sharing can vary widely and can involve various types of actors and have differing goals. For example, two businesses may use contractual agreements to share data to facilitate collaboration on a large-scale project. Or multiple individuals may share data with through an independent organization for financial gain.

Data-sharing models largely vary based on who contributes the data, who has access to the data, who stores and manages the data, and who benefits from the data sharing. Data contributors can be any actor that owns or creates data, including individuals, private companies, government agencies, nonprofits, and research institutions. Likewise, those same actors can also be the ones receiving data in a data-sharing arrangement. For example, a government agency might share health data with pharmaceutical companies investigating new drugs, or a pharmaceutical company might share its data on vaccine distribution with the government or public health researchers. These agreements can be one-way, where one actor shares data with another party in order to receive specific insights, or reciprocal, where each actor receives data. Lastly, data-sharing models differ depending on who receives and stores the data, such as the data owner or an intermediary institution. These factors are the core differences among data-sharing models.

The following section explores and evaluates six different models, illustrating their respective strengths and weaknesses and offering recommendations for U.S. policymakers on how to best implement and increase data sharing across the nation.

Data-Sharing Partnerships

Data-sharing partnerships involve collaborative efforts between different entities, such as academic institutions, research organizations, industry partners, individual consumers, and government agencies, to share and exchange data for the purpose of conducting research, collaborating on new products, and enhancing evidence-based decision-making. These partnerships aim to leverage the collective expertise, resources, and data holdings of multiple parties to address complex questions and generate valuable insights. This type of data-sharing arrangement usually requires clear agreements to define data access and usage rights and the ownership of intellectual property (IP), but specific characteristics may vary depending on the type of data involved and the nature of the collaboration.

For example, health care is one field in which partnerships between organizations such as hospitals, research institutions, and medical providers can help leverage data analytics and artificial intelligence (AI) in health care research, ultimately improving patient outcomes and optimizing service delivery.[3] The 23andMe Patient-Centric Research Portal allows customers to voluntarily contribute their genetic and self-reported health information to research studies.[4] This partnership between 23andMe, a genomics and biotechnology company, and its customers enables the advancement of scientific understanding of various diseases and traits. Researchers link genetic data in order to study topics such as ancestry, traits, and even rare diseases.[5]

Data-sharing partnerships have a number of benefits for all partners. For researchers, such partnerships provide access to greater data for analysis than they would have on their own, allowing for greater insights. These partnerships also help overcome the limitations of a single dataset that may be too small for certain types of statistical analysis or be missing relevant information needed for investigation. In fields where data is often highly sensitive, such as health care, research partnerships protect the sensitive nature of patient information while allowing institutions to collaborate and aggregate insights. Such partnerships also mean less duplication of data, saving researchers time and money.

At the same time, this data-sharing model has some constraints. For example, when data-sharing partnerships are between two competing institutions, there are often IP and competition concerns. Likewise, such collaborations may involve datasets of varying quality and standards. These issues must be resolved before any sharing can occur.

Recommendation: Facilitate data-sharing partnerships with model contracts.

Partnerships between two entities are the most basic model of data sharing and should be supported by policymakers. When it comes to data-sharing partnerships, organizations are often forced to reinvent the wheel and go through a new contract and negotiations process each time a new opportunity for collaboration comes up. Policymakers in federal agencies should alleviate this roadblock and facilitate more data-sharing partnerships by developing a contract template that organizations can adopt and customize (e.g., type of data, retention terms, IP rights, etc.). Some countries such as Singapore already provide this type of guidance for data-sharing partnerships and have accelerated research and innovation as a result.[6] Moreover, the Federal Trade Commission and Department of

Justice should provide guidance on complying with antitrust rules on collusion when using these model partnership agreements.

Data Consortia

Data consortia allow organizations to pool their data for the benefit of the group.[7] Whereas data-sharing partnerships involve bilateral agreements, data consortia constitute a series of reciprocal sharing agreements. These consortia can exist to address a specific issue or for the general and ongoing exchange of information. For example, a group of towns along a river might form a data consortium to share data about bacteria in the water, or a group of hospitals might form a data consortium to share data about a specific rare disease. Likewise, online marketplaces might form a data consortium to exchange data about third-party sellers that are selling counterfeits.[8]

Data consortia have long played a role in filling data gaps. For example, the Clinical Research Data Sharing Alliance exists to accelerate drug discovery by sharing data collected throughout the clinical development process.[9] Members of the consortium include biopharmaceutical companies, academic institutions, nonprofits, and patient advocacy groups, which come together around the globe to provide collective access to clinical data and help diversify study populations. Another example of the utility of data consortia is the Linguistic Data Consortium (LDC) at the University of Pennsylvania.[10] This group of universities, libraries, corporations, and government labs was founded in 1992 "to address the critical data shortage facing language technology research and development." Members of the LDC share language resources, such as speech and text databases, lexicons, and other resources, that play a big role in training large language models.

The primary benefit of data consortia is that it promotes more data sharing and aggregation. Only members of a given consortium can access the data, and consortium members typically must contribute to the group. Eventually, a data consortium will create a tipping-point effect in which it is more beneficial to be in the collective than out. Once a tipping point is reached, a consortium ensures ongoing data sharing and generally promotes a pro-data-sharing world.

Data consortia do have some drawbacks. Before a critical mass is reached and a tipping point effect occurs, some organizations might be better off hoarding their data for their exclusive use. This means consortia need to consider joining incentives in the early days of an effort.

Recommendation: Survey and identify opportunities for cross-sector data consortia.

Federal agencies should catalog data consortia that exist within specific sectors and facilitate the creation of new cross-sector consortia for critical areas. Data consortia can provide policymakers with access to a broader and more diverse range of data sources, including from other agencies and the private sector. For example, policymakers in interdisciplinary agencies such as the Federal Emergency Management Agency should create consortia that bring together relevant stakeholders from agencies such as the Environmental Protection Agency, the Department of Agriculture, and the Department of Housing and Urban Development as well as private sector organizations to coordinate ongoing data sharing and ensure more coordinated and effective disaster response.

Data Trusts

Data trusts are a type of data governance framework that manage, protect, and share data for an agreed purpose on behalf of individuals and organizations.[11] Although there can be conflicting definitions of data trust, the characteristics of this type of data-sharing mechanism remain the same. At the core of a data trust is the delegation of data rights to an independent intermediary, known as a trustee, who makes data-sharing decisions with researchers, private companies, and public-sector bodies that benefit the data subjects.[12] Data trusts give structure and rules for managing and using aggregated data and help unlock its value for the public interest.

Data trusts are an emerging model with a number of variations being piloted around the world. The United Kingdom has taken a particular interest in the data trust model for health care applications. For example, the UK Biobank manages the genomic data of more than 500,000 individuals who have donated their data for use in research.[13] The data is anonymized and made available to researchers around the world to accelerate scientific discovery and improve public health. The Biobank acts as a trustee for this data—in other words, it has a fiduciary responsibility to hold and share the data for the benefits of the UK public. Additionally, the UK's National Health Service (NHS) is developing an NHS Federated Data Platform to aggregate all health data, including personal health records, clinical data, and public data, in one centralized platform individuals and the private sector alike can access.[14]

There are a number of benefits to data trusts, including increasing societal benefits from data, streamlining processes, and unlocking more value from data by enabling secondary use. Overall, data trusts are an institution that multiple entities can contribute to and access, thereby facilitating ongoing transparency and accountability and consistent rules for the reuse of data. Governments can therefore access private sector data in key areas under a clear set of agreements, and vice versa. In the context of AI, they can facilitate access to diverse and high-quality datasets, enabling AI developers to train and validate models on more comprehensive and representative data. Overall, data trusts provide a trusted framework for managing data responsibly.

At the same time, data trusts do come with certain challenges. Given the often-sensitive nature of the data held by a trust, data trusts can be difficult to implement and sometimes are met with resistance. A lack of social trust in data sharing can lead to projects being held up or even canceled, such as in the case of InBloom, a proposed data trust for education that was met with so much stakeholder resistance that it failed to launch.[15] Moreover, data trusts can be resource-heavy, requiring a lot of financial, technical, and human resources. Lastly, data trusts can be at odds with data protection laws focused on safeguarding individual rights—which are common in many Western countries—because they focus on collective empowerment and benefit. This collective model can be difficult to implement in the context of stringent data privacy laws.

Recommendation: Implement sector-specific data trusts.

There are specific domains in the United States, such as health care, transportation, education, and environmental research, in which the establishment of sector-specific data trusts could provide significant benefits to society. These trusts could bring together stakeholders from relevant sectors to pool and govern data, enabling research, improving service delivery, and driving societal outcomes in specific areas. Consolidating data assets within a specific sector would enable a comprehensive understanding of sector-specific challenges, trends, and opportunities. Federal agencies, such as the Environmental Protection Agency, the Department of Health and Human Services, and the Department of Education, should establish programs to create and operate data trusts in their respective domains. By providing guidance and capacity-building support, federal agencies could help the data trusts navigate the legal and regulatory frameworks specific to each industry.

Data Cooperatives

Data cooperatives are a form of bottom-up data governance in which individuals voluntarily pool their data to negotiate collectively with private companies and other entities. Members of a data cooperative establish rules on data sharing designed to benefit those in the group. These cooperatives often aim to monetize members' collective data and are funded by the revenue generated from data-sharing agreements. Data cooperatives are similar to agricultural, housing, and consumer credit cooperatives in which the organization is owned and jointly managed by its members, who share the benefits.

For example, Driver's Seat Cooperative pools gig economy workers' smartphone and mobility data, allowing them to optimize their incomes.[16] The cooperative functions through an app that links multiple sources of an individual driver's data and analytics, then aggregates that data for all members of the collective. Driver's Seat also sells these group insights to local governments looking for data to help transportation planning decisions and splits dividends among members. This type of data-sharing arrangement is primarily designed to empower workers to use their data as a collective bargaining mechanism. Data cooperatives also exist in the agricultural sector as a means of empowering farmers with shared knowledge. Cooperatives such as the National Agricultural Producers Data Cooperative and the Grower's Information Service Cooperative operate at the national level and pool data from producers, small businesses, public universities, and nonprofits in order to provide farmers and growers with agricultural data and help enhance the sustainability of their operations.[17]

One challenge is that the economics of data cooperatives do not always work.[18] The value of each individual data contributor might be relatively small, but without widespread participation from many data holders, the cooperative will fail. Data cooperatives therefore have to carefully choose how to compensate members—too little, and not enough contributors will join; too much, and it is not sustainable.

Data cooperatives are a relative novelty and may have limited applications. However, just as labor unions provide an important mechanism for collective bargaining for workers, data cooperatives can allow individuals to collectively negotiate benefits for their data.

Recommendation: Explore data cooperatives in areas where nondata cooperatives and collective bargaining occurs.

Data cooperatives are useful when individuals may be reluctant to share their data because they fear a third party will use it against their interests, such as small farmers who are concerned that large agricultural companies will use their data, and insights from their data, to profit at their expense.[19] Forming data cooperatives can give these data holders more negotiating power to share data with others on their preferred terms and overcome reluctance to collect and share data. Federal agencies that already provide oversight or support for various types of nondata cooperatives, such as the U.S. Department of Agriculture, the National Credit Union Administration, and the National Labor Relations Board, should convene stakeholders on the potential value of data cooperatives in their respective areas to promote greater data sharing.

Federated Data Analytics

Federated data analytics is a way to allow data analysis to occur even when organizations are unable or unwilling to share their data. Federated data analytics refers to a distributed approach to data processing in which data is analyzed in disparate locations and only the aggregated insights are brought to a centralized location. For example, a company might use federated data analytics to analyze data stored on its customers' devices without moving any of that customer data to the company. Instead, the company would only receive the results of the analytics.[20] Federated data analytics, including methods such as federated learning, allows data insights to be obtained without sharing the data itself. By not sharing the data, this method can assuage fears about organizations accessing or storing sensitive data, such as concerns about privacy for individuals or proprietary company information.[21]

For example, biopharmaceutical company Boehringer Ingelheim and precision medicine software company Lifebit Biotech have partnered to build a scalable federated analytics platform to capture genomics insights from biobank data.[22] Federated analytics in this case preserves the privacy of the highly sensitive biomedical data but still allows researchers to access insights from individual data stores and generate medical innovations.

Federated analytics has a number of benefits, including providing a new way to access large quantities of data. For example, biomedical research and clinical trials require patient data that is typically held by a number of different health care institutions and bound by strict privacy laws.[23] Federated analytics enables privacy-preserving analyses of datasets without revealing any specific patient data; and each data provider retains control. This type of data sharing can enable precision medicine and is critical in situations where one dataset from one provider won't be enough to identify meaningful patterns, such as in the case of rare disease research.[24] Eventually, federated health data networks can facilitate large-scale analysis across institutions, regions, and borders, a possibility being considered by the EU-U.S. Trade and Technology Council.[25]

At the same time, federated analytics has a few drawbacks, including problems of cost, scalability, stakeholder resistance, and lack of technical readiness in institutions. As an emerging data-sharing technique, computing costs for federated data analytics may be high and prohibitive for many applications. There are also limited large-scale examples of federated analytics at work, particularly in fields such as health care, which creates problems of scalability given the lack of a blueprint for some sectors. Moreover, not all institutions are receptive to federated analytics or technically equipped to enable it.[26] Federated analytics requires a scalable platform that can deal with large quantities of data, as well as advanced application programming interfaces (APIs) that allow for coordination between platforms.[27]

Recommendation: Continue funding R&D for federated analytics.

The White House Office of Science and Technology Policy recently released a "National Strategy to Advance Privacy-Preserving Data Sharing and Analytics," which outlines the importance of federated analytics and other privacy-enhancing technologies to improve the value of data for the public benefit while protecting individual privacy.[28] While such a report is important to organizing policy response to federated data-sharing models, federated analytics should not be put on a pedestal or considered a silver bullet to data-sharing dilemmas. In some ways, federated analytics exists as a technical solution to a social problem of distrust in the collection and use of data or the legal barriers to data sharing between different countries' data protection laws. However, federated analytics does make sense as an option when data sharing is otherwise not feasible, so policymakers should fund research and development (R&D) efforts in this field to develop this capability and determine which sectors stand to benefit the most from federated analytics.

Cooperative Research and Development Agreements

Cooperative Research and Development Agreements (CRADAs) are partnerships between the government and private sector in which government research institutions share data with private sector partners to facilitate the commercialization of R&D projects.[29] Under these agreements, the government provides personnel, services, facilities, IP, equipment, data, and more to their collaborators—but no funding. The partners, which can be private corporations, nonprofits, universities, or even state governments, provide the same services but can also provide funding for the R&D efforts.[30] CRADAs allow private sector partners to file for patents and retain patent rights while ensuring the government gets a license for any commercialized product.[31]

For example, the National Ocean and Atmospheric Administration (NOAA) uses CRADAs to leverage the value of its datasets and better ensure public access to data. The NOAA Big Data Initiative began in 2015 to enlist private companies such as IBM, Microsoft, and AWS to develop solutions to increase utilization and access to its data.[32] Due to budgetary and security constraints, NOAA could not keep up with public demand for access to critical datasets of things such as weather radar data, satellite imagery, historical climate data, information on fisheries, etc.[33] The agency wanted to promote the use of its data and democratize access, while collaborators had the infrastructure expertise. This partnership reduced loads on NOAA systems and budgets and created new business opportunities for the partner companies.

CRADAs have a number of benefits. First, they have no impact on government budgets, as they primarily utilize existing infrastructure and simply make government and private sector collaboration more efficient and effective. CRADAs also protect IP and allow partners to monetize solutions, meaning they have built-in incentives. At a high level, these agreements help expand and enhance the existing expertise of government, industry, and academia and contribute to overall national competitiveness and spur more innovation.

CRADAs are highly focused and contractual in nature, meaning they need authorization and careful negotiation of terms related to liability and IP.[34] For example, federal research must be made available to the public, so CRADAs must work out ways to withhold research results for a certain period of time in order to allow the private partner to patent any inventions for commercial use. While federal research institutions are all entitled to undertake this type of research agreement, not all may have opportunities for such collaboration, or even want it. Given that CRADAs occur on a project-by-project basis, they do not necessarily ensure ongoing public-private partnerships.

Recommendation: Evaluate areas where R&D can benefit from CRADA arrangements.

Many government agencies already use CRADAs, particularly in the physical and medical sciences, and federal agencies should continue to promote and authorize them. Given the many benefits of this type of data-sharing agreement, Congress should ask the Government Accountability Office to identify what types of R&D currently benefit the most from this type of collaboration and which agencies should make better use of CRADAs.

CONCLUSION

As evidenced by these examples, data sharing can create social and economic value for businesses, governments, and individuals. Enabling data sharing means more people and organizations can access vast amounts of data and put it to productive use, leading to innovations and discoveries, an empowered citizenry, and better data-driven decision-making by policymakers.

There are a number of barriers to data sharing, including privacy and transparency concerns, technical challenges, and economic constraints. However, various data-sharing models can overcome these barriers and increase the value of data to individuals, organizations, and the government alike. Some of these models are more geared toward protecting sensitive information by default than are others, but each of them can use technical and legal measures to protect the confidentiality of shared data. Policymakers should do more to support the development and adoption of different data-sharing models in order to increase the overall amount of data sharing occurring in the United States, including:

- Providing model contracts for data-sharing partnerships
- Surveying and identifying opportunities for cross-sector data consortia
- Implementing sector-specific data trusts
- Exploring data cooperatives in areas where nondata cooperatives and collective bargaining occurs
- Continuing to fund R&D for federated data analytics
- Evaluating areas where CRADAs can benefit R&D activities

The United States needs to get serious about enabling and expanding data sharing across all parts of society if it hopes to lead in the age of AI. Data sharing can help close critical gaps.[35]

REFERENCES

1. Daniel Castro and Travis Korte, "Data Innovation 101," *Center for Data Innovation*, November 3, 2013, https://datainnovation.org/2013/11/data-innovation-101/.

2. Hodan Omaar, "No, The Data Economy Is Not A Barter Economy," *Center for Data Innovation*, September 13, 2021, https://datainnovation.org/2021/09/no-the-data-economy-is-not-a-barter-economy/.

3. Shannon L. Sibbald, Hosung Kang, and Ian D. Graham, "Collaborative health research partnerships: a survey of researcher and knowledge-user attitudes and perceptions," *Health Research Policy and Systems*, (2019), https://health-policy-systems.biomedcentral.com/articles/10.1186/s12961-019-0485-3.

4. "23andMe Research," 23andMe, accessed June 2023, https://www.23andme.com/research/.

5. "23andMe Patient-Centric Research Portal," DataCollaboratives, accessed June 2023, https://datacollaboratives.org/cases/23andme-patient-centric-research-portal.html.

6. "Trusted Data Sharing Framework," Infocomm Media Development Authority, last modified August 3, 2023, https://www.imda.gov.sg/how-we-can-help/datainnovation/trusted-data-sharing-framework.

7. "What are data consortiums?" Experian, accessed July 2023, https://www.experian.co.uk/privacy/consumer-information-portal/using-data/data-consortium/.

8. Sujai Shivakumar, "How Data-Sharing Partnerships Can Thwart Counterfeits on Online Marketplaces" (ITIF: March 2021), https://www2.datainnovation.org/2021-data-sharing-counterfeits.pdf.

9. "About," Clinical Research Data Sharing Alliance, accessed September 2023, https://crdsalliance.org/about/.

10. "About LDC," Linguistic Data Consortium, accessed July 2023, https://www.ldc.upenn.edu/.

11. Jack Hardinges et al., "Data Trusts: Lessons from Three Pilots" (Open Data Institute: April 2019), https://theodi.org/article/odi-data-trusts-report/.

12. Ada Lovelace Institute and AI Council, "Exploring Legal Mechanisms for Data Stewardship" (Ada Lovelace Institute: March 2021), https://www.adalovelaceinstitute.org/wp-content/uploads/2021/03/Legal-mechanisms-for-data-stewardship_report_Ada_AI-Council-2.pdf.

13. "UK Biobank," accessed June 2023, https://www.ukbiobank.ac.uk/.

14. Bernard Marr, "Building the World's Largest Healthcare Data Platform," *Forbes*, March 23, 2023, https://www.forbes.com/sites/bernardmarr/2023/03/23/building-the-worlds-largest-healthcare-data-platform/?sh=7942d0477692 and "NHS Trusts and Sites," NHS England Data Catalogue, accessed July 2023, https://data.england.nhs.uk/dataset/ods-nhs-trusts-and-sites.

15. Monica Bulger, Patrick McCormick, Mikaela Pitcan, "The Legacy of InBloom" (working paper, Data&Society: February 2, 2017), https://datasociety.net/pubs/ecl/InBloom_feb_2017.pdf.

16. “Driver’s Seat,” accessed June 2023, https://driversseat.co/.
17. “Project Summary,” Ag Data Coop, accessed September 2023, https://www.agdatacoop.org/.
18. Sameer Mehta, “The key to designing sustainable data cooperatives,” *Open Data Institute,* March 20, 2023, https://www.theodi.org/article/odi-fellow-research-the-key-to-designing-sustainable-data-cooperatives/.
19. Michael Gilbert, “Farmers Have Trust Issues With Big Data – For the Same Reasons We All Do,” *Entrepreneur,* March 21, 2022, https://www.entrepreneur.com/leadership/farmers-have-trust-issues-with-big-data-for-the-same/421619.
20. Daniel Ramage and Stefano Mazzocchi, “Federated Analytics: Collaborative Data Science without Data Collection,” *Google Research*, May 27, 2020, https://ai.googleblog.com/2020/05/federated-analytics-collaborative-data.html.
21. The MITRE Corporation, “A Step Ahead: Federated Data Stores,” The Data Administration Newsletter, February 1, 2023, https://tdan.com/a-step-ahead-federated-data-stores/30363.
22. Boehringer Ingelheim, “Boehringer Ingelheim and Lifebit announce partnership to capture transformational value of health data,” news release, January 3, 2022, https://www.boehringer-ingelheim.com/science-innovation/human-health-innovation/partnership-boehringer-ingelheim-lifebit-health-data.
23. David Froelicher et al., “Truly privacy-preserving federated analytics for precision medicine with multiparty homomorphic encryption,” *Nature Communications*, (2021), https://www.nature.com/articles/s41467-021-25972-y.
24. Ibid.
25. Harry Hallock, “Federated Networks for Distributed Analysis of Health Data,” *Frontiers in Public Health,* (September 2021), https://www.frontiersin.org/articles/10.3389/fpubh.2021.712569/full and Luca Bertuzzi and Molly Killeen, “Tech Brief: TTC4 agenda, B2G data-sharing,” *Euractiv*, May 5, 2023, https://www.euractiv.com/section/digital/news/tech-brief-ttc4-agenda-b2g-data-sharing/.
26. Hannah Gaimster, “The complete guide to federated data analysis,” *Lifebit*, accessed September 2023, https://www.lifebit.ai/federated-data-analysis/the-complete-guide-to-federated-data-analysis.
27. Ibid.
28. National Science and Technology Council, “National Strategy to Advance Privacy-Preserving Data Sharing and Analytics” (Washington, D.C.: National Science and Technology Council, March 2023), https://www.whitehouse.gov/wp-content/uploads/2023/03/National-Strategy-to-Advance-Privacy-Preserving-Data-Sharing-and-Analytics.pdf.
29. “CRADAs,” National Institutes of Health, accessed July 2023, https://www.techtransfer.nih.gov/policy/cradas.
30. Homeland Security: Science and Technology, “Cooperative Research and Development Agreements” (fact sheet, Department of Homeland Security), accessed July 2023,

https://www.dhs.gov/sites/default/files/publications/crada_factsheet_20201117-508c.pdf.

31. "CRADAs," National Institutes of Health, accessed July 2023, https://www.techtransfer.nih.gov/policy/cradas.

32. Edward J. Kearns, Andy Bailey, and Shane Glass, "NOAA's Big Data Initiative" (presentation at NWS Partners Meeting, January 2017), https://www.weather.gov/media/wrn/calendar/NOAA_BigData_Session_2017.pdf.

33. "Evolution of the NODD Program," National Oceanic and Atmospheric Administration, last modified August 30, 2023, https://www.noaa.gov/nodd/about.

34. Alexander Kostura and Daniel Castro, "Three Types of Public-Private Partnerships That Enable Data Innovation," *Center for Data Innovation*, August 1, 2016, https://datainnovation.org/2016/08/three-types-of-public-private-partnerships-that-enable-data-innovation/.

35. Gillian Diebold and Daniel Castro, "Digital Equity 2.0: How to Close the Data Divide" (Center for Data Innovation, May 2023), https://datainnovation.org/2023/05/digital-equity-2-0-how-to-close-the-data-divide/.

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