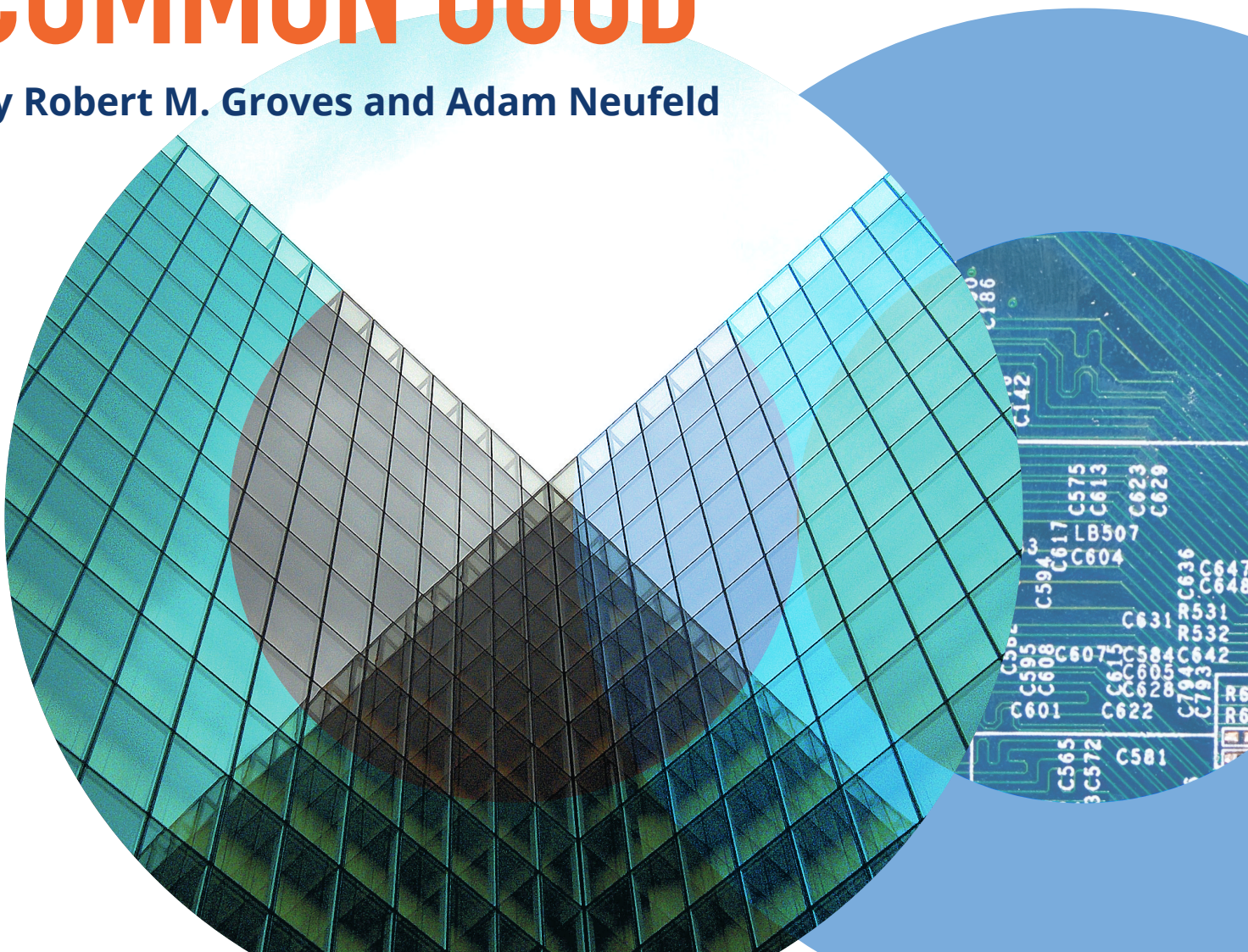


# ACCELERATING THE SHARING OF DATA ACROSS SECTORS TO ADVANCE THE COMMON GOOD

By Robert M. Groves and Adam Neufeld



## Authors

**Robert M. Groves** is the Provost of Georgetown University and Gerard Campbell SJ Professor in the department of Mathematics and Statistics and the department of Sociology. He was Director of the Census Bureau from 2009-2012.

**Adam Neufeld** is a Senior Fellow at Georgetown University's Beeck Center for Social Impact + Innovation and Georgetown Law's Institute for Technology Policy and Law. He was Deputy Administrator of the General Services Administration from 2015-2017.

## Acknowledgements

While this paper reflects only the views of its authors and not the institutions with which they are affiliated, they thank the Alfred P. Sloan Foundation for its financial support as part of its program for the development of an Administrative Data Research Network. We also express appreciation to Alberto and Olga Maria Beeck, whose vision and support have built the Beeck Center for Social Impact + Innovation at Georgetown University. Additional thanks to Catherine Lyons for her assistance as well as Michael Delli Carpini, Amy O'Hara, Marc Rotenberg, John Thompson, and Stefaan Verhulst for feedback on earlier drafts.

## About the Beeck Center for Social Impact + Innovation

The Beeck Center for Social Impact + Innovation at Georgetown University was launched in 2014 through the generosity of Alberto and Olga Maria Beeck. Part lab, part think tank, and part classroom, the Beeck Center works with students and practitioners to innovate new solutions to systemic social problems.

The Beeck Center engages global leaders to drive social change at scale. We conduct academic research, hold convenings and provide trainings, classes, and experiential labs for current and emerging leaders. Building on Georgetown University's core values and global reach, we drive creative thinking around issues at the intersection of data, technology, capital, and policy innovation. We invite a diverse community of thinkers to join us in looking beyond the obvious, to ask the unasked questions, and find adaptable solutions in today's changing world.

# TABLE OF CONTENTS

**Executive Summary ..... 2**

**Overview..... 3**

**The Data Opportunity ..... 5**

**The Transformative Power of Sharing Statistical Data Across Sectors.....8**

**Obstacles to Sharing Statistical Data ..... 11**

**Current State of Public-Private Data Sharing and Linking..... 14**

**The Need For a New Intermediary Focused on Sharing Between Sectors ..... 17**

**Conclusion .....23**

# EXECUTIVE SUMMARY

**The public pays for and provides an incredible amount of data to governments and companies. Yet much of the value of this data is being wasted, remaining in silos rather than being shared to enhance the common good—whether it’s helping governments to stop opioid addiction or helping companies predict and meet the demand for electric or autonomous vehicles.**

- Many companies and governments are interested in sharing more of their data with each other; however, right now the process of sharing is very time consuming and can pose great risks since it often involves sharing full data sets with another entity.
- We need intermediaries to design safe environments to facilitate data sharing in the low-trust and politically sensitive context of companies and governments. These safe environments would exist outside the government, be transparent to the public, and use modern technologies and techniques to allow only statistical uses of data through temporary linkages in order to minimize the risk to individuals’ privacy.
- Governments must lead the way in sharing more data by re-evaluating laws that limit sharing of data, and must embrace new technologies that could allow the private sector to receive at least some value from many sensitive data sets.

By decreasing the cost and risks of sharing data, more data will be freed from their silos, and we will move closer to what we deserve—that our data are used for the greatest societal benefit.

# OVERVIEW

**Every day, we provide data about ourselves to companies and governments—the digital trail of our financial transactions documents our purchases and our financial health. Our phone calls show our location and our communication patterns. Our search terms and social media postings reveal what we are interested in and whom we associate with.**

Similarly, our responses to government surveys provide personal details, such as our family structure and whether we have been a victim of a crime. Our answers to mandatory government forms indicate our earnings and company structures. Our use of government benefit programs shows our medical transactions and unemployment periods. In other cases, we just pay for the data to be collected—whether it's NASA satellites funded by taxes or a company's R&D project paid for by customer profits.

The recipients of our data use the information for narrow purposes that we at least technically agree to, either through terms of service or through policies built by elected representatives. The data are used in important ways, but generally limited to serving the needs and interests of the data recipient in administering programs (whether Medicare or a credit card program), making decisions (whether where to locate an ATM or an additional Veterans Affairs hospital), or understanding consumer behavior (whether targeted advertising or educational outreach). By such a siloed approach to data provided by the public, multiple agencies and companies waste individuals' time by repeatedly asking the same questions.

This narrow use misses much of the value of the data we have provided or paid for. Data about individuals can be used for two distinct purposes: a) to take specific actions on individuals based on their data, or b) to construct statistical information describing large groups in the population. The first use often raises legitimate concerns about the privacy of individuals. The second use is inherently uninterested in individual data, and raises fewer privacy concerns; instead it is the source of



aggregate descriptions of large groups of individuals.<sup>1</sup> In this discussion we are solely focused on statistical uses of data. We argue that the data we have provided could be used much more effectively to provide better planning and policymaking for our own benefit.

Sharing<sup>2</sup> data with other users or combining data sets can yield statistical information that sheds light on key societal challenges or opportunities. This is especially true for inter-sector data sharing, because the two sources of data differ significantly and are complementary, with private sector data more likely to be spatially granular, real-time, and growing, but having fewer measured attributes and changing more over time than data from the public sector. Only by combining data across sectors can we get the full value of our data. For example, understanding and addressing the opioid crisis requires connecting private sector data (such as employer policies, use of addiction centers, and private insurance

*Sharing data with other users or combining data sets can yield statistical information that sheds light on key societal challenges or opportunities.*

---

transactions) with government data (death records, workers' compensation filings, Medicaid and Medicare transactions). Helping the un- or under-employed also requires connecting data outside the government on student achievement, courses, and credentials with data inside the government on earnings and employment. Similarly, designing effective ride-sharing, autonomous vehicles and other private sector transportation options requires understanding where, when, and how people get around, which can be found in data that are collected at various levels of government through public transportation systems, traffic sensors, toll records, as well as employment and home address data.

Society need not miss out on this broader value of our data. The current narrow use of our data is a response


to constraints that are not inherently necessary. As a result, we propose creating a safe environment for sharing data that removes or at least mitigates many of those constraints. These safe environments would exist outside the government, be transparent to the public, and use modern technologies and data techniques to allow only statistical uses of data through temporary linkages in order to minimize the risk of intruding on individuals' privacy. For example, an intermediary could help companies and governments link data to address the pressing question of what employers can do to help their workers avoid or recover from opioid dependency and overdose. An intermediary could link employers' data on relevant benefits offered and used by employees; states' data on which employees have filed for workers' compensation (and thus might have an injury that could make them at risk of taking opioids) and Medicaid and Medicare transactions identifying which publicly insured employees received opioid prescriptions; private insurers' data on privately insured employees who received opioid prescriptions; and the U.S. Centers for Disease Control and Prevention's data tracking deaths. Rather than these owners transferring their data to one another, the intermediary would provide a temporary data linkage among them for immediate analysis of whether certain employee benefits decreased the chance that injured employees became dependent, overdosed, or died from opioids.

The government must lead and show itself willing to share far more data with others, enticing them to share their own data in return. Even when government does share microdata, it often only shares with academics and not the rest of the private sector. Governments must re-evaluate laws that allow no or excessively narrow sharing of their data, and must embrace new technologies that could allow the private sector to get at least some value from many sensitive data sets.

Only by creating safe environments, with the government stepping up first, will we ever increase the amount of data shared between sectors. This will require hard work to build trust, clarify laws, and overcome the culture of undersharing. However, our data are too valuable—and the societal issues and opportunities too great—not to try.

# THE DATA OPPORTUNITY

**The public has given or paid for the government to have an incredible amount of data. For example, the 13 principal federal statistical agencies spend over \$6 billion annually to conduct surveys and collect data on topics ranging from population to gross domestic product to crime victimization. Individuals and companies spend about 10 billion hours annually providing information just to the federal government.<sup>3</sup>**



In addition, state and local governments produce various statistics, such as infectious disease incidence, police reports, and student performance. In addition to statistical data, governments collect “administrative” data to operate programs, such as the transactions of 50 million Medicare beneficiaries or the nearly 2 million people who received unemployment insurance benefits at a given moment. Finally, governments also collect data on physically observable phenomena. For example, NASA’s Earth Observing System Data and Information System collects 6.4 terabytes of images of Earth each day.<sup>4</sup> State and local governments also collect sensor data to better understand and mitigate operational inefficiencies. For example, Los Angeles has 145,000 smart LED streetlights alone.<sup>5 6</sup>

While the public sector continues to have a significant role in collecting data, private sector data have become increasingly important as a result of digitization of processes and the advent of the “internet of things.” Google has insight into people’s interests and concerns through the 2 trillion searches it processes a year.<sup>7</sup> Foursquare knows the daily whereabouts and favorite places of its 10 million users, who “check in” more than 8 million times a day.<sup>8</sup> Visa knows where and on what millions of Americans spend their money through the 150 million transactions recorded a day—up to 24,000 per second.<sup>9</sup> On one day in 2016, Amazon processed more than 600 transactions per second.<sup>10</sup> By 2040, 80 percent of traffic signals in the U.S. are predicted to have sensors that can constantly collect information on vehicle speed, traffic volume, weather conditions, and more.<sup>11</sup>

Neither public nor private sector data are “better.” Rather, they are different—and it is this difference that makes combining them so exciting. Particularly notable differences include:

### DIFFERENCES BETWEEN PUBLIC AND PRIVATE DATA

**Volume:** Depending on the context, private and public sector data volume can differ significantly. When it comes to data on human behavior (besides beneficiaries’ use of government programs), the government usually lags the private sector. For example, the government asks 60,000 households about their job-hunting and employment each month to calculate the unemployment rate. In contrast, 120 million jobs are viewed monthly on Monster.com. Similarly, MIT’s Billion Prices Project “scrapes” millions of online items a day to compute a price index, compared with the 80,000 items reviewed monthly by the government to calculate the official Consumer Price Index (CPI).<sup>12</sup> When it comes to data on physically observable phenomena, the government often has significantly more volume.

**Granularity:** With the exception of the decennial census and data required to administer programs, the government must make cost-benefit analyses on how much data to collect. Sampling allows for good overall estimates, but those estimates lose statistical precision in smaller geographical areas. For example, the federal government publishes a consumer price index for only

27 local areas, and warns users to exercise caution in relying on these indices given the smaller sample size.<sup>13</sup> In contrast, with the increasing national penetration of smartphones and technologically mature companies, private sector data exist on virtually everyone.

**Timeliness:** Most public sector statistical information is available only with some time lag. For example, the Bureau of Labor Statistics releases a monthly unemployment rate for each state<sup>14</sup> at least halfway through the following month and for metropolitan areas<sup>15</sup> a week or so later. Similarly, data from the American Community Survey (ACS) is released starting the September after the year of data collection.<sup>16</sup> In contrast, much private sector data are updated in real time based on transactions or movements of users. New data are collected every time your smartphone moves to a new location and every time you withdraw money from an ATM, make a phone call, or make an insurance claim.

**Coverage:** Most public sector data collection is designed to provide a comprehensive understanding of an issue in order to



accurately assess the situation or to inform policy. This means that the government goes to great effort to reduce survey nonresponses and ensure a statistically representative sample. For example, the Census Bureau follows up over the phone and in person with people who do not respond to the ACS, and even has the ability (which it has never used) to prosecute or fine nonrespondents. Additionally, the government collects data from people who have little connection to the formal economic marketplace. In contrast, the private sector's data collection is understandably constrained by its profit motive, meaning that it may collect less data on issues, people, and behavior further removed from existing and potential economic transactions. As a result, significant caveats exist in treating the private sector's data as representative of a population for governmental purposes.

**Data richness:** To inform policy, governments often ask for rich demographic or other data for both administrative and statistical purposes. For example, in calculating the unemployment rate, the federal government does not just inquire whether a respondent has a job, but also

collects data on personal characteristics such as age, sex, race, marital status, educational level, and veteran status. In contrast, a private company will collect only those data fields necessary for its administrative processes.

**Stability:** Understandably, most private sector companies collect data to serve their own administrative processes, not for external analysts. As a result, data fields can change as business operations or strategy changes. In contrast, the government changes data collection or methodology very rarely. For example, the government changes survey methodology or administrative data fields infrequently, and usually only after advance notice and a long public comment period.

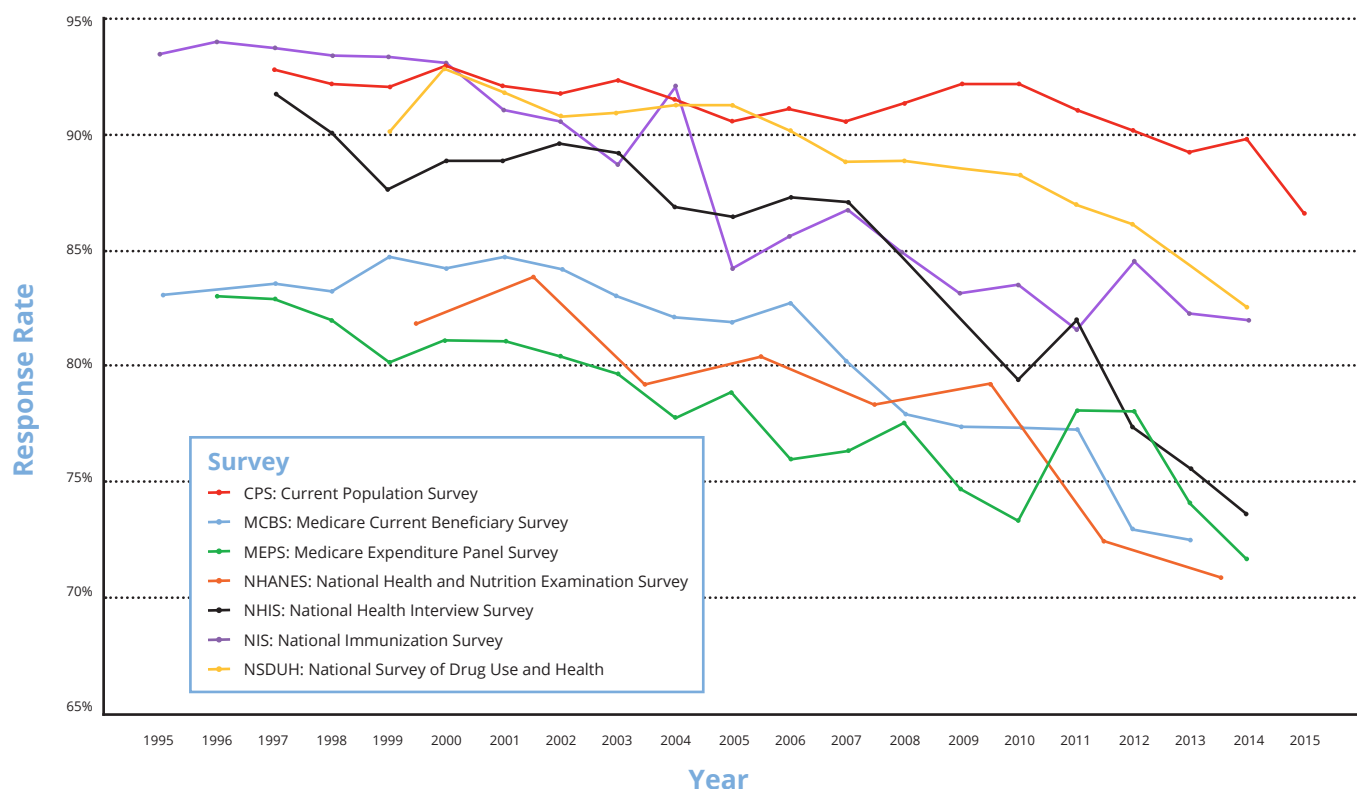
# THE TRANSFORMATIVE POWER OF SHARING STATISTICAL DATA ACROSS SECTORS

**Sharing of data across sectors and levels of government can help society address urgent problems or seize exciting opportunities.**



The federal Commission on Evidence-based Policymaking discussed the importance of improving sharing within and among different levels of government, and we wholeheartedly agree. Here, we complement their work by focusing on the potential of sharing data between government and the private sector. Thoughtful sharing of data between sectors is critical for advancing the common good so that all sectors make economic decisions and develop policies, programs, and offerings based on the best data. First, by getting access to private sector data, the government can improve the accuracy and utility of statistics.<sup>17</sup> Federal, state, and local governments use the data to evaluate past efforts (for instance, has poverty declined since a program began?) and to make policy decisions (would new programs better serve our aging population?). Private companies also use statistics from government data, such as retail firms that use census estimates to target specific demographic groups with new stores. Finally, quantitatively focused academics rely on the federal government's large and long-standing data to study topics ranging from drivers of income mobility to climate trends. However, the state of government statistics is weakening. As seen in the chart on response rates over time for several federal statistical surveys, fewer people are responding to government surveys,<sup>18</sup> which increases survey costs and raises concerns about accuracy:

## THE TRANSFORMATIVE POWER OF SHARING STATISTICAL DATA ACROSS SECTORS



Source: Mathematica Policy Research<sup>19</sup>

More importantly, many statistics no longer meet the needs of public and private users, because the low frequency and lack of granularity are ill matched to a fast-changing world with significant variations by region and subgroup. In some cases, delayed statistical indicators can mislead policymakers and the public. For example, during the Great Recession, the government's indices on prices were very slow to reflect changes in the economy. However, private sector data from credit card transactions, point-of-sales systems, and other sources could have offered more timely and granular estimates. For example, the Billion Prices Project scrapes online prices for 15 million items every day, providing a daily price index for at least a large portion of the economy.<sup>20</sup> Similarly, federal housing surveys do not capture America's housing market in a sufficiently granular or timely way, as they provide just annual estimates on only 25 metropolitan areas. Blending these data with commercial loan processing and other sources could offer greater insight.

In addition to improving government statistics, limited access to private sector data could allow the government to make more informed policy and operational decisions. A diverse group of policymakers and stakeholders, from House Speaker Paul Ryan to the Obama administration and from Bloomberg Philanthropies to the American Enterprise Institute, have encouraged greater use of data in government decision-making. Most have focused on the government's use of its own data. As useful as that is, combining public and private sector data will make the government even more effective. For

*Limited access to private sector data could allow the government to make more informed policy and operational decisions.*

example, the telecommunications company Telefonica helped Mexico study its efforts to prevent the spread of infectious diseases by analyzing phone records to see how movement of people changed in response to different government alerts.<sup>21</sup>

*The private sector has shown that consumers are willing to pay for products and services where companies have made government data more accessible or useful.*

---

Second, government data can also provide additional information to the private sector in making business decisions, potentially increasing the chance of success (and thus growing the economy). Businesses already use official government statistics. For example, Target considers the annual ACS data on household composition in deciding what to offer in each store. Similarly, more than 20 companies accessed Medicare and Medicaid microdata in an effort to identify new business.

Simultaneously, the private sector can benefit immensely from government data, as seen in the nascent Open Government effort. First, the private sector has shown that consumers are willing to pay for products and services where companies have made government data more accessible or useful. For example, the availability of open GPS data in the 1980s created an entire subsection of businesses that contributed an estimated \$90 billion to the U.S. economy. Additionally, the National Weather Service's continuous and open data were fundamental in the creation and growth of the weather industry. Over several decades, the public has gone from relying on the government entirely for its weather data to 95 percent of the public using private companies for weather updates. AccuWeather alone is estimated to be on 1.5 billion devices worldwide.<sup>22</sup> However, we are only at the tip of the iceberg. For example, the McKinsey Global Institute estimates that more than \$3 trillion in economic potential could be unlocked annually worldwide if data became more open as a result of increased transparency and subsequent new businesses.<sup>23</sup>

# OBSTACLES TO SHARING STATISTICAL DATA

**Despite the transformative potential of private-public data sharing, several obstacles stand in the way, even when there is mutual interest.<sup>24</sup> These obstacles are typically complex, requiring significant attention from multiple stakeholders. We have not built the necessary policy infrastructure, standards or norms for sharing data across sectors—leading to our society receiving far less benefit than it could from the data that it paid for or provided.**

## Motivation

Neither governments nor companies generally have an incentive to prioritize the hard work of sharing granular data with other sectors. The benefits of providing data to other sectors typically do not fall to the data owner, at least in the short term, making it hard to justify taking energy and resources away from achieving the core societal mission of the government agency and the profit-making mission of companies. Moreover, stakeholders, whether Congress and the public or the board of directors and customers, do not generally focus on data sharing, but rather on performance, customer satisfaction, and profits.

However, in some cases, the motivations of governments and companies may be changing. An increasing number of companies are sharing, analyzing, or encouraging research on their data to serve the public good in various forms.<sup>25</sup> For example, Uber provides certain de-identified data from 2 million Uber trips through Movement;<sup>26</sup> JPMorgan Chase created a think tank that analyzes the company's extensive data to publish reports on matters of social and policy importance;<sup>27</sup> and Twitter provides grants for research using its data.<sup>28</sup> Some suggest that we may be witnessing the beginning of corporate "data philanthropy," analogous to corporate responsibility programs or pro bono work.<sup>29</sup> Additionally, some companies share their data as part of their business model, either by earning revenue from selling their data or providing their data freely to increase related



revenue opportunities (for example, Google shares its maps' Application Programming Interface (API)). In terms of government sharing its data with the private sector, the last decade has seen the rise of "open data" in government, with the public and stakeholders beginning to expect and reward the public release of data. However, these data will last only so long as the data provider stays in business and wishes to share the data, risks that users such as governments must consider.

### Legal and policy

Even if the parties are motivated, they still must address a raft of complex legal and policy issues primarily focused on privacy. These issues are particularly prominent in data that have information that could identify an individual. Company sharing of data must address and answer legal questions such as liability in case the data are illegitimately accessed, defining the scope of access, compliance with privacy and other laws, impact of freedom of information laws, and whether sharing may violate terms of service. On the other side, law or policy may prohibit or limit agencies from sharing certain information in order to protect privacy, as the Obama administration and the Commission on Evidence-Based Policymaking documented.<sup>30</sup> For example, Congress allows analysis of federal tax data only for tax administration purposes, as opposed to the many other potential uses. Similarly, laws sometimes limit how the government can release or share data. For example, the U.S. Census Bureau allows access to microdata only for academics who have become sworn employees of

*Privacy risks must be weighed when sharing or linking data, and will sometimes be a significant obstacle to sharing—especially when there is no trusting, transparent relationship between the user and the provider of the data.*

the Census. Additionally, even when no law prohibits or limits data sharing, the legal uncertainty on privacy and other laws can require significant legal and policy discussion. In some cases, agencies may be risk-averse and decide not to share in the absence of explicit legal permission.<sup>31</sup> Privacy regulations and federal oversight of information requests of the public place further restrictions on the use of respondent-provided personal data. Generally, such regulations prohibit uses outside those specified in the original request. The Privacy Act does have an exception that allows some additional use of federally collected data for statistical purposes.<sup>32</sup>

### Privacy

There are two types of privacy risks in sharing or linking data, both of which can pose obstacles. First, the more individuals or systems that have access to data, the more possible ways the data can be breached and unauthorized actors obtain identifiable information (e.g., names and social security numbers). These data could then be used in ways (whether legal or illegal) that hurt those individuals. Indeed, Americans give low marks when it comes to organizations protecting their data. For example, in one survey, 28 percent of respondents said they were not at all confident that the government has the ability to protect their personal information.<sup>33</sup> Second, in addition to breaches of identifiable information, sharing of data, even statistical data, can increase the chances that individuals can be re-identified by providing more information on the relevant populations.<sup>34</sup> As a result, there can be a privacy risk when datasets are linked and identifiable information is stripped. In many cases, the data that would be most useful for advancing the common good also raise the most significant concerns if individuals can be re-identified. In contrast to their concern about breaches, the public does not appear as concerned about this type of privacy risk when done for public policy purposes. For example, in a Census survey, less than 20 percent of respondents opposed combining data by the government in order to improve government accountability or efficacy.<sup>35</sup> Recent technological developments have allowed us to better measure and minimize privacy risk as well as to extract greater statistical benefit with the same amount of privacy risk.

Use of these approaches, such as query tools, synthetic data, and multiparty shared computing, create largely untapped opportunities to mitigate privacy risk. These privacy risks must be weighed when sharing or linking data, and will sometimes be a significant obstacle to sharing - especially when there is no trusting, transparent relationship between the user and the provider of the data. We are fortunate that recent developments in computer science and statistics provide tools to address these privacy threats in more effective ways than were possible in the past.

## Trust

Governments and companies are sometimes in adversarial relationships, whether in inspections, regulations, or lawsuits. As a result, data owners may fear that users will not respect the agreed-upon scope or the spirit of the data sharing, resulting in legal problems or government use of the data in a way that hinders a business opportunity. For example, a company may fear that employee salary data shared with the Census Bureau for statistical purposes may lead to litigation if the data suggest discrimination, even if the law or the data-sharing agreement clearly prohibits such a use. Additionally, a company may believe that the government is incapable of securing its network or limiting access, and thus fear accidental public release of sensitive data that help the company's competitors. That is, despite trust in accepting the government's intentions, they may fear that it does not have the capability to fulfill its commitments to data security. On the other side, the government may fear that data it receives from the private sector may be low quality or even manipulated for ulterior purposes. Similarly, the government may fear that granular data shared will be used in disfavored ways, beyond the scope of the agreement, or to publicly embarrass the government about data quality or poor program performance.

## Technical

The mechanics of sharing data also present obstacles, particularly for organizations that only rarely do so. Cleaning data can be both complicated and expensive. Questions include how to provide access to data, whether to build in privacy protections such as synthetic data,<sup>36</sup> how to link and harmonize different data sets, and whether both parties have sufficient skills and capacity. Solutions to the technical barriers have come in various forms, each with different advantages and disadvantages. For example, the Census Bureau provides confidential microdata access only to approved researchers in secure facilities, while the Centers for Medicare and Medicaid Services (CMS) allows virtual access for researchers.<sup>37</sup> Wharton Research Data Services stores data on-premises, while the University of Chicago's Center for Data Intensive Science stores data in the cloud.

This is a formidable list of barriers that generally requires signoff from individuals in an organization including operational leadership, chief information officer, general counsel, and government affairs. But they are not insurmountable.

# CURRENT STATE OF PUBLIC-PRIVATE DATA SHARING AND LINKING

**Unsurprisingly, these barriers have stopped much public-private data sharing, keeping us far from realizing the potential of much of the data we have provided or paid for.**

The government uses private sector data in four main ways, each with significant limitations.

- **Publicly available data:** Because of business interests, technological limitations, or altruism, many companies have data that are publicly available as a data set either directly (such as through API or publication) or indirectly (such as through web scraping). For example, Uber recently released certain de-identified data on its rides, although the data are a small fraction of what would be useful for governments. Zillow releases a significant portion of its data to approved researchers, which could include the government. Additionally, the Department of Justice recently redesigned its method of counting arrest-related deaths by relying on open information sources that scrape online news articles for keywords.<sup>38</sup> However, overall, only a small portion of valuable granular administrative data is publicly available.
- **Commercially available data:** The government frequently purchases commercially available data. For example, the U.S. Department of Agriculture buys point-of-sale data from Nielsen, and the Federal Reserve Board purchases payment processing data from Firstnet to understand domestic spending.<sup>39</sup> Similarly, a leading international cellular provider developed tools to understand effectiveness of public health alerts that it now offers governments.<sup>40</sup> Most of these arrangements provided aggregated data to a government agency, not person-level data. Purchasing data from companies or third-party vendors that have a business model can be easy for government since there is likely to be greater customer service, data quality, standardized

process, and stability in the data. However, much valuable data is not commercially available at all or at the level that allows for desired data linking. Additionally, the government may lack the resources to pay for certain valuable data.

- **Mandate:** Governments sometimes mandate that people or companies provide certain data (such as the decennial census, addresses, and other personal data for driver's licenses). Unfortunately, the information is often provided in a manner that is not integrated with other systems the public uses, creating a potentially duplicative cost.
- **Ad hoc agreements:** In rare cases, governments and companies have overcome the barriers listed

above to provide the government access to data that are not otherwise available. In most cases, data are not shared with the government per se, but the company runs a desired analysis and the output is shared. In some cases, the sharing goes further. As described in the callout box, the Census Bureau, the National Association of Manufacturers, and the National Student Clearinghouse are working to combine data sources to assess whether available manufacturing credentials are making workers more productive. These ad hoc agreements can take many forms and usually involve a long and difficult process with many players to produce individualized agreements.

### SHARING DATA TO EXPLORE THE IMPACT OF EDUCATIONAL CREDENTIALS

The National Association of Manufacturers (NAM), a trade association representing U.S. manufacturing interests, wanted to know whether certain certificates and community college programs in skills such as welding and machining result in higher wages and better career opportunities.<sup>41</sup> This data would help prospective employees choose the most effective credentials, which translates into a better and more productive workforce. Answering this question required both private and public sector data. Through an innovative national data-sharing partnership with the National Student Clearinghouse, third-party credentialing companies, and the U.S. Census Bureau, NAM is working to link data on credentials and education with employment and wage history. NAM persuaded leading credentialing companies as well as some states to share data with the Student Clearinghouse. The Student Clearinghouse then transferred that data to the Census Bureau's secure Federal Statistical Research Data Centers

(FSRDCs), where it will be linked to IRS wage data and Census business data. The process of getting to this partnership involved significant effort. Data from the third-party credential programs had to be reformatted for compatibility, and missing data from higher education institutions had to be accounted for. The IRS had to be brought on board to provide critical wage data. The data had to be securely transferred among three different entities, each with its own technical capacities. The association worked to standardize data-sharing agreements and legal documents to cement the data-focused partnership among these public and private institutions. NAM led this important work not only to provide insight into the relationship between credentialing and employment, but also because of the other potential applications to address key manufacturing challenges. As the NAM lead said, "We built this because we envision data sharing as a service to the industry on a host of important issues."

In terms of the private sector having access to the government, the main mechanisms are listed below. As with government access to private sector data, the mechanisms each have significant limitations.

- **Open data:** Governments make a significant amount of data available to the public. These are not personally identified data. As technology has developed, press releases with summary statistics have evolved into spreadsheets with more granular data and then, in some cases, into APIs. For example, the federal government's data.gov has information ranging from income tax returns by ZIP code to complaints received about financial products. Additionally, certain data on physically observable phenomena is available at a very granular level in real time, such as weather information. The public has also scraped certain government websites for social or business purposes. For example, many companies take government information that is public but not part of the government's open data effort (such as procurement solicitations or campaign finance information) and make it more accessible and add services. However, most open data are not at the individual level because of privacy concerns, making linking difficult and preventing the release of most information on many government programs.
- **Freedom of information laws:** Governments also make certain data available in response to requests under their respective freedom of information laws. The federal government and all states have some form of freedom of information law, all of which provide some degree of protection for personal privacy and sensitive business information. Courts have generally interpreted the federal privacy exemptions broadly to cover disclosure of information that could reveal privacy through subsequent data linkage.<sup>42</sup> As a result, although freedom of information requests are common,<sup>43</sup> they generally will not yield much additional granular data. Additionally, the release of data often takes a long time, may come in less accessible formats, and may need to be repeated in order to make longitudinal comparisons.
- **Research:** All statistical agencies and some other agencies provide some restricted access to granular, sensitive information for research under limited circumstances on an agency-by-agency basis.<sup>44</sup> The most common example is the Federal Statistical Research Data Centers (FSRDCs), which are secure offices where approved researchers who pass a background check can access certain otherwise restricted microdata for on-premises research, and then export only statistics that cannot be used to re-identify anyone. In addition, agencies will occasionally share their government data with researchers in an ad hoc manner that must address many of the same obstacles as when government tries to get data from the private sector.<sup>45</sup> However, in many cases, data can be made available only for research that advances the agency's mission (rather than fulfill another societal or business purpose). For example, federal tax data can be shared only for research that can improve tax administration, and research on Census Bureau data must benefit Census Bureau operations. However, some agencies have provided information for research aiming to provide broader societal value, and have even allowed companies to conduct the research. Despite these meaningful efforts, the government generally suffers from weak incentives, insufficient resources, or varied technical capacity to undergo the hard work of preparing the data, reviewing applications, and working with researchers.

In total, neither the government nor the private sector has access to anything close to the types or volume of data from the other that would be useful—a lost opportunity to capitalize on the financial and privacy costs incurred in collecting the data.



# THE NEED FOR A NEW INTERMEDIARY FOCUSED ON SHARING BETWEEN SECTORS

**While the question of *whether* data should be shared can be complex, we need to evolve *how* we share data in those cases where parties want to share. There is no default path or standard options for sharing microdata for statistical use, leaving parties with additional hurdles to negotiate, such as lack of trust, privacy concerns, absence of legal standards, and technological capacity. Unsurprisingly, much potential data sharing fails to overcome these obstacles.**

Only by adding a new means of sharing data that addresses the identified obstacles will we ever see a significant improvement. We do not expect the creation of new intermediaries to result in every company or government allowing linking of their data. Rather, we hope to shift the data-sharing equilibrium point towards more accessibility by decreasing the impediment of sharing data for statistical purposes.

Facing a similar problem with low sharing of data *within* the government and with academics, the Commission on Evidence-Based Policymaking proposed creating a federal statistical agency that would help facilitate temporary linkages among government data sets in a privacy-preserving manner.<sup>46</sup> Many of the arguments for creating such an intermediary—called the National Secure Data Service (NSDS)—apply even more to the context of sharing across sectors. The creation of entities outside the government not only would help promote more and better inter-sector data sharing, but also could make the NSDS even more effective.

Most companies and government agencies do not engage in inter-sector data sharing enough to develop competence and confidence with the array of complex legal, policy, and technical issues. These organizations might have a few one-off agreements or solutions, but, even if effective, these were not necessarily designed to scale. The arrangements usually involve complicated nondisclosure agreements and involvement of R&D staff, corporation attorneys, and others. This can result in subpar sharing practices that may carry a greater risk to privacy or provide less utility than necessary. Additionally, the lack of competence and confidence can result in risk aversion toward approving data sharing because, generally, the costs and risks of

*By establishing themselves as experts and a default option for data sharing, intermediaries could help create a standardized path that would make data owners more comfortable that sharing is being done in an appropriate, privacy-preserving, and secure manner.*

data sharing fall mostly on the data owner while the benefits fall mostly to society at large.<sup>47</sup> Intermediaries can address both issues.

Because the intermediaries' focus would be data sharing, they could invest both money and human capital in identifying and maintaining best practices. By establishing themselves as experts and a default option for data sharing, intermediaries could help create a standardized path that would make data owners more comfortable that sharing is being done in an appropriate, privacy-preserving, and secure manner. Just as the Census Bureau's FSRDCs helped establish norms, rules, and security for statistical agencies to share their data, nonprofit intermediaries would do the same for inter-sector data sharing. Additionally, intermediaries would address the risk aversion of most organizations toward data sharing by providing a political (and potentially legal) safe harbor. If something went wrong, the intermediary would shoulder much of the blame. Overall, by lowering the transaction costs for data sharing, intermediaries would help increase the total amount of responsible sharing for statistical purposes.

Several excellent nongovernmental entities exist to facilitate data sharing to advance the social good, from intermediaries to collaboratives to prizes. Some are focused on a particular industry (for example, the National Student Clearinghouse focuses on higher education) or issue (the University of Michigan's IRIS focuses on research on innovation and science), while

others are more general (NORC at the University of Chicago). Most entities focus on sharing data for academic purposes, and transfer data or provide general access to data.

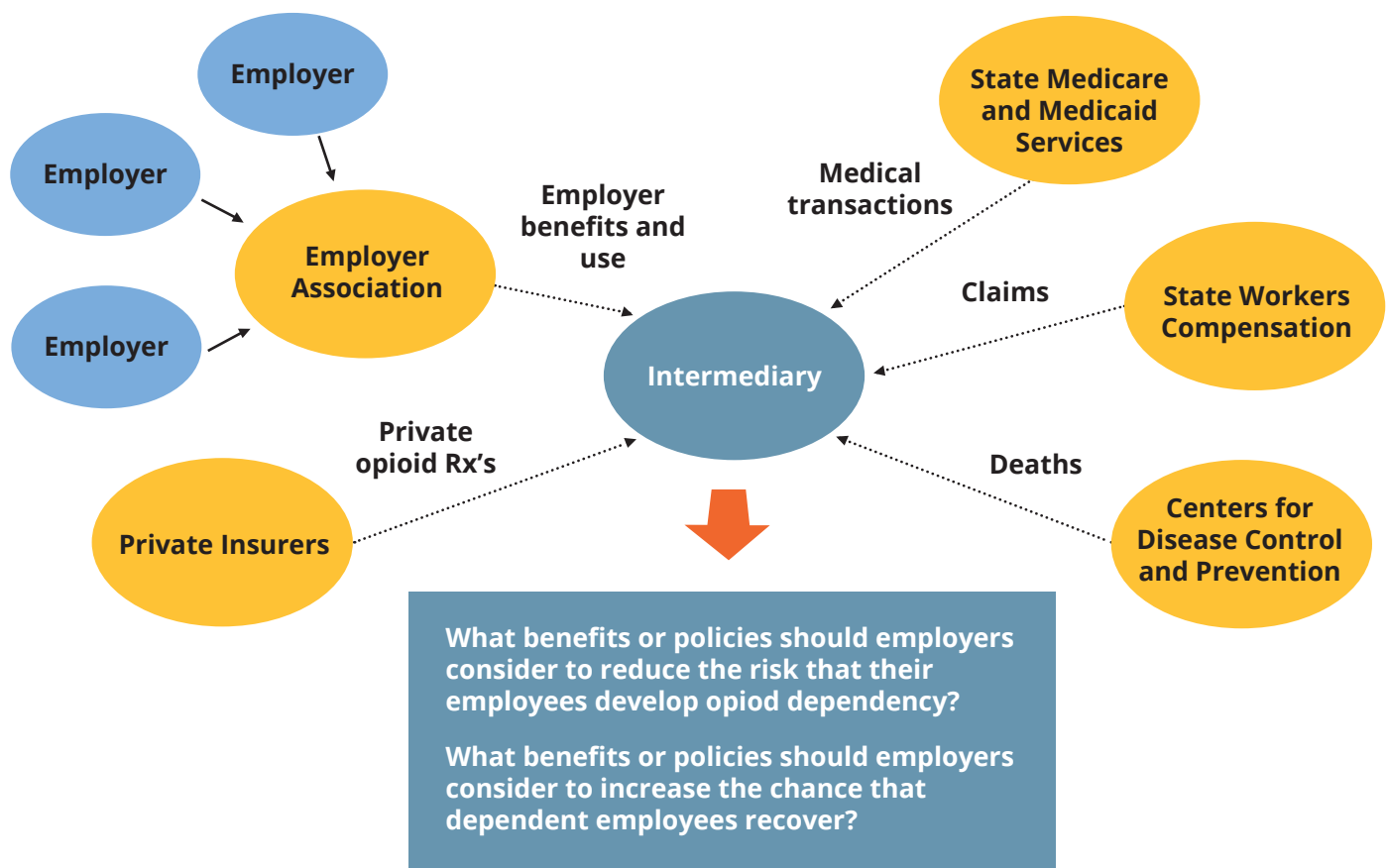
We propose a new type of intermediary designed to facilitate particularly difficult cases of data sharing—situations in which trust is especially low or the privacy concerns are particularly substantial, such as between the private and public sectors.

This focus on facilitating “harder” cases of data sharing drives several differences from most organizations. First, the intermediary provides only temporary linkages among data sources to facilitate immediate statistical analysis, rather than requiring transferring of data or providing broader access, as most existing organizations do. Second, the use must be only for statistical rather than administrative purposes, meaning, for example, the data combining can help identify patterns in who violates laws or may be interested in certain products but cannot be used to identify specific individuals to target. Third, the intermediary itself takes significant independent responsibility for assessing and mitigating privacy risk and for ensuring transparency, since the data sharers may lack significant experience together or shared expectations or norms.<sup>48</sup>

The graphic shows a hypothetical example of how this could work to answer the important question of how employers can help their workers avoid or recover from opioid dependency and overdose. This is a pressing question, since an estimated 75 percent of people struggling with opioid dependency are in the workforce, and employers are struggling with the workplace and productivity consequences.<sup>49</sup>

Answering this question requires many data sets from across sectors. Employers have data on relevant benefits offered and used by employees, such as coverage for alternative pain treatments, sick leave, counseling, and working from home. In this hypothetical, employers have decided to transfer data on a regular basis to an association's data warehouse, so the association rather than the individual employers is the relevant data sharer for the intermediary. States have data on what employees have filed for workers'

## THE NEED FOR A NEW INTERMEDIARY FOCUSED ON SHARING BETWEEN SECTORS



compensation (and thus might have an injury that could make them at risk of taking opioids) and on Medicaid and Medicare beneficiaries transactions identifying which publicly insured employees received opioid prescriptions. In most states, data from insurance companies would be needed to supplement this information by identifying privately insured employees who received opioid prescriptions. Finally, the federal Centers for Disease Control and Prevention has data that track deaths.<sup>50</sup> Rather than these owners transferring their data to one another, the intermediary provides a temporary data linkage among them for immediate analysis, with a clear record of what data were shared and for what purposes.

We do not believe there is one right answer for these intermediaries, and instead encourage competition as data user and provider needs evolve. This competition will be particularly important as we learn what services entities are willing to pay intermediaries for. However, we do believe that intermediaries should share the following core features that increase trust:

**Independence:** As discussed above, a key barrier to sharing data across sectors is trust. Companies want to make sure the government does not use the data in ways that hurt their business interests, and the government taking control of private sector data imposes security and legal risks. The public's privacy concerns also seem to diminish when the government is not the one combining data. The Census found that 62 percent of survey respondents were more likely to support combining data if a third party (rather than the government) was overseeing confidentiality and protecting privacy.<sup>51</sup> Indeed, in many cases, state and local governments are concerned about sharing data with the federal government. As a result, an independent entity is key for improving inter-sector data sharing. An organization overseen by data owners, data users, and privacy experts could provide greater confidence to all parties that data sharing will happen according to best practices and established agreements. As discussed in the callout box, Estonia took this approach in facilitating data sharing between mobile phone operators and the government to estimate tourism statistics. Additionally, companies

already use independent intermediaries occasionally to facilitate data linkages among themselves. Funding for the intermediary could come in various forms, from data users paying to industries paying through associations.

### DATA INTERMEDIARY IN ESTONIA

A third-party intermediary in Estonia facilitates the sharing of data from mobile phone operators with the government to develop tourism statistics and studies. Several academics, urban planners, and architects formed the intermediary, called Positium, to explore uses of mobile positioning data. Positium created software that resides within the phone company's system and samples, cleans, analyzes, and condenses data in a way that protects privacy and sensitive business information. Positium then analyzes or links the data in the desired ways to produce outputs for the government. The statistical output is then shared as often as weekly with the government for use in statistics, program evaluation, and policy decision-making. For example, the government uses the software to make tourism estimates by analyzing calls from within Estonia on phones with SIM cards registered in other countries. Additionally, the government used the mobile positioning data to identify people who might need help during a snowstorm.<sup>52</sup>

**Focus on privacy:** The technology for protecting privacy has evolved substantially since organizations simply deleted names, addresses, and Social Security numbers from spreadsheets. Technical approaches (such as query tools, synthetic data, and multiparty shared computing) and mathematical methods (such as differential privacy) now allow for far more sophisticated ways to reduce the risk of re-identifying

people while enhancing research utility. Many government entities and companies do not face enough data-sharing issues to develop expertise on privacy protection. In contrast, the intermediaries would have the mission, incentive, and resources to stay current on privacy approaches and how best to balance accuracy with privacy. An intermediary is more likely to follow the best current practices, including relying on temporary data linkages rather than permanently housing restricted data as well as allowing the export of statistics only at a sufficiently aggregated level. Additionally, organizations have matured in decision-making on privacy. Leading organizations have disclosure review boards that weigh the costs and benefits of proposed data linkages or research efforts, with many organizations having Internal Review Boards. Some organizations also perform periodic privacy audits to ensure that they are following their internal policies and processes. This evolution of privacy-preserving technical and managerial tools is likely to continue.

**Statistical purposes:** Linking microdata across sectors has many potential purposes, and the intermediary must allow use for only certain objectives. Specifically, the intermediary should allow analysis only for statistical purposes—that is, for understanding the behavior or characteristics of certain groups. In contrast, the intermediary should not allow analysis for administrative uses, such as to compile a list of individuals to target with a company's mailing or to enforce the law against. This has multiple benefits, including that the intermediary's output need only be statistical insights or summaries rather than personally identifiable information.

**Transparency:** People are understandably interested in—and in some cases anxious or fearful about—how their data are used, particularly when the government is involved. As a result, transparency is of utmost importance at all stages of data sharing. The intermediaries must disclose what data can be accessed, what linkages have been performed, how the data have been used, and what redaction methods were employed. A start for transparency is Canada's statistical agency, which publishes an inventory and summaries of data linkages on its website. The public should be able to easily know what data are being

analyzed, for what purpose, and with what outcome. A web portal with advanced search procedures would provide such visibility.

**Legal clarity:** Executives and general counsels are justifiably concerned about the many legal obstacles and legal risks associated with data sharing. The lack of experience in data sharing by most organizations, combined with the lack of established norms on how to allocate legal rights, results in high transaction costs for agreements. Through templates or their terms of service, the intermediaries would establish default legal terms (or a limited number of options) that could simplify the range of legal issues that the organizations face. Similarly, the intermediaries would develop expertise in relevant federal and state laws, such as the Privacy Act, Family Educational Rights and Privacy Act (FERPA), and Health Insurance Portability and Accountability Act (HIPAA).

**Security:** Decisions on technical infrastructure are intertwined with privacy, cybersecurity, and confidentiality concerns. While sometimes there is a clear correct answer, in most cases, organizations have different values that could result in different trade-offs. For example, the Census Bureau decided to share its data only with researchers working inside a small number of secure offices, while CMS allows for virtual access to data. Intermediaries must strive to meet the desired research goal with minimum security and privacy risks by focusing on data linkage rather than data warehousing. Similarly, the federal government has different standards for hosting different types of data. The intermediaries must ensure that their technology is sufficient for government to provide access for statistical purposes. Additionally, the intermediaries must have strong auditing and management controls to detect misuse of data. For example, the business-focused Wharton Research Data Services platform has strong processes to prevent data misuse, such as keystroke logs and standard automated compliance checks.

**Nonprofit:** The goal of inter-sector data sharing here is to serve the public good, whether in the form of better-designed policies or new companies that help grow the economy. Sharing data among companies for the private good is important, and some

companies already offer that service. A nonprofit is best positioned to advance inter-sector sharing for the public good because of financial incentives and greater ability for transparency. A nonprofit would also be more likely to stay focused on inter-sector data sharing, rather than private-to-private data sharing, which could be a bigger business. The nonprofit could be a stand-alone organization or be housed at a university, such as Brown University's Rhode Island Innovative Policy Lab, which serves as an intermediary for public and private sector data. At least some data also suggests the public has higher levels of trust in nonprofits. In the 2017 Edelman Trust Barometer, 58 percent of those surveyed in the United States said they trust nongovernmental organizations, while only 47 percent reported trust in the government.<sup>53</sup>

*The technology for protecting privacy has evolved substantially since organizations simply deleted names, addresses, and Social Security numbers from spreadsheets.*

**Flexibility:** The scope and methods of inter-sector data sharing will change in response to new technologies, opportunities, needs, and concerns. As a result, intermediaries should be designed with flexibility top of mind. The dynamic needs of data owners and users also argue for a dynamic marketplace where intermediaries are competing in terms of offerings, security, and customer service, but against a backdrop of interoperability. For example, intermediaries may differentiate themselves based on offering data hosting in addition to data linking; the level of sensitivity of data they work on; subject matter focus; domestic or international scope; or technical assistance in crafting analyses or accessing government data.

While governments might need to change their own laws to allow for full use of intermediaries, governments can encourage the creation of useful intermediaries even without any new legislation or



increased budget by issuing guidance clarifying the conditions they would need to provide access to data through an intermediary. Given the complexity of such requirements, a more technically competent agency such as the Census Bureau should consider drafting requirements first. The government also could consider creating a safe harbor that would protect companies from certain legal risks if they link to government via an intermediary.

As the legal landscape has evolved over the decades, a complicated set of federal laws and regulations, all seeking to protect the privacy of individual data, has been constructed. In most all of those the federal regulatory framework permits only one-way sharing of data. That is, it is possible for state or local governments, as well as private sector firms, to pass their data to a federal agency for blending with federal data, but the blended data set is not permitted to exit federal control. The federal purposes and control usually trump those of any other partner. In that sense, we have constructed a protective framework for individual data that prevents the federal agencies from being true partners. Such asymmetry is a chilling influence on partnerships that could serve the American public more fully.

As a result, governments share only a small portion of their data with the private sector. Given the high cost people pay in both time and money for data collection, governments must ensure that society gets the most out of this data while preserving individual privacy and organizational competitiveness by adopting new technologies and changing laws.

New privacy-preserving technologies and approaches can allow government to make much otherwise-sensitive data publicly available. Differential privacy techniques are emerging as a useful framework. The best example of this is the Census Bureau's use of synthetic data to create OntheMap, a web interface that allows the public to conduct detailed analyses based on data on where people live and work. To prevent the risk that people could be identified from the data, OntheMap does not use the actual underlying Census data. Rather, it creates a synthetic version that introduces random noise to the data in such a way that analyses of the data are still statistically valid.<sup>54</sup>

This or similar approaches could allow governments to provide valuable—even if not perfect—open versions of additional statistical or administrative data, such as Veterans Health Administration transactions or unemployment insurance records.

In addition, governments should re-evaluate and change laws to broaden what types of limited research can be done on microdata. Right now, many agencies can allow only research that benefits their particular mission—so broader societal benefits are irrelevant. That should be changed in most, if not all, cases. In addition, the law and policy should clarify that companies, and not just academics, can use the data for research. The innovative researcher program at CMS is the best example of this broader approach. CMS offers virtual access to microdata for research intended to develop innovative products or tools that could improve health care operations. Researchers must receive approval from CMS, they can download only aggregate statistics and summaries, and derivative products and tools are reviewed to ensure that they will not exploit beneficiaries or enable fraud.<sup>55</sup>

Finally, governments should ensure that agencies are sufficiently incentivized and resourced to provide data to the private sector, whether in an open or a more limited manner. The costs of providing data access can be significant, and often funds or management attention must be diverted from other priorities.<sup>56</sup> As a result, government should consider expanding reimbursable authorities so that agencies can charge for preparing, sharing, and validating data for the private sector, which will allow data provision to scale to meet market demand.

# CONCLUSION

**Building a better world through data is crippled by our using only a small fraction of the existing data. We deserve more from the data we provide and pay for. The sequestering of private and public data hurts society by making both private enterprise and government far less effective than they could be. Sharing data across sectors can help us better tackle societal problems and grow the economy.**



Greater inter-sector data sharing is possible only by creating a safe environment for data to be shared; otherwise, good intentions will fizzle out when confronted with the complex issues involved in sharing. Nonprofit intermediaries can help decrease the transaction costs of inter-sector data sharing by providing legally, technically, and politically acceptable options for data owners and users. Not only will these intermediaries increase sharing, but they will also ensure that sharing uses the latest privacy-preserving technologies and policies.

Fostering intermediaries can happen in part even without major changes, such as creation of the National Secure Data Service or streamlined researcher approval process. Rather, nonprofit intermediaries are complementary, and would allow the government and the private sector to start improving data sharing even while exploring other improvements. The government must step up first and show that it is willing to partner with the private sector by embracing new technologies and changing laws, enticing companies to share their statistical data in turn.

Without these changes, governments and companies will continue operating in the dark when it comes to critical policy and business decisions—something that society cannot afford. Only with safe inter-sector data sharing can we achieve the full societal or economic benefits for the country.

## Endnotes

- 1 Statistical estimates on groups of individuals can be, and often are, used to inform actions on individuals, whether they're targeting outreach for assistance or denying credit. Many of these uses are societally undesirable, and should be addressed through policy. However, we do not believe a compelling case has been made that preventing data linking generally is the best way to address these problems.
- 2 In this document, we use "share" to mean providing some form of access or connection to data. That access can range from providing a raw data file to providing a temporary data linkage through an intermediary.
- 3 [https://obamawhitehouse.archives.gov/sites/default/files/omb/inforeg/icb/icb\\_2016.pdf](https://obamawhitehouse.archives.gov/sites/default/files/omb/inforeg/icb/icb_2016.pdf)
- 4 <https://earthdata.nasa.gov/getting-petabytes-to-people-how-the-eosdis-facilitates-earth-observing-data-discovery-and-use>
- 5 <http://www.govtech.com/network/Practical-Uses-of-the-Internet-of-Things-in-Government-Are-Everywhere.html>
- 6 Although the distinction among types of government data is important in many instances, this report generally does not distinguish between them because the "how" of data sharing across sectors is generally similar. For the same reason, the report does not generally distinguish between whether the government seeks private sector statistical data for purposes of official statistics (for example, to incorporate into the unemployment rate) or policy or program design (for example, to assess whether a program is effective). That distinction matters for whether a company might want to share the data and whether the government finds the data valuable enough to seek, but not for how agreed-upon sharing occurs.
- 7 <http://searchengineland.com/google-now-handles-2-999-trillion-searches-per-year-250247>
- 8 <https://venturebeat.com/2016/03/24/foursquares-swarm-now-lets-you-make-sense-of-where-youve-been/>
- 9 <https://usa.visa.com/run-your-business/small-business-tools/retail.html>
- 10 <https://www.inc.com/tom-popomaronis/amazon-just-eclipsed-records-selling-over-600-items-per-second.html>
- 11 <http://www.bridgemi.com/business-bridge/smart-road-speed-location-sensors-going-michigan-roads>
- 12 <https://www.bls.gov/cpi/questions-and-answers.htm>
- 13 <https://www.bls.gov/cpi/questions-and-answers.htm> ("Each local index has a much smaller sample size than the national or regional indexes and is, therefore, subject to substantially more sampling and other measurement error. As a result, local-area indexes are more volatile than the national or regional indexes, and BLS strongly urges users to consider adopting the national or regional CPIs for use in escalator clauses.")
- 14 [https://www.bls.gov/schedule/news\\_release/laus.htm](https://www.bls.gov/schedule/news_release/laus.htm)
- 15 [https://www.bls.gov/schedule/news\\_release/metro.htm](https://www.bls.gov/schedule/news_release/metro.htm)
- 16 <https://www.census.gov/programs-surveys/acs/news/data-releases/2016/release-schedule.html>
- 17 See, e.g., <https://medium.com/@sverhulst/better-data-for-better-policy-accessing-new-data-sources-for-statistics-through-data-33a66b2a237d>; <https://unstats.un.org/unsd/trade/events/2014/beijing/Steve%20Landefeld%20-%20Uses%20of%20Big%20Data%20for%20official%20statistics.pdf>; <http://www.oecd-ilibrary.org/docserver/download/5jm3nqp1g8wf-en.pdf?expires=1509392900&id=id&accname=guest&checksum=7874CADED323F2C491FA3CE1E8B04640>
- 18 <https://aspe.hhs.gov/system/files/pdf/255531/Decliningresponserates.pdf>
- 19 Beyler, Amy and John L. Czajka. "Background Paper: Declining Response Rates in Federal Surveys: Trends and Implications." Mathematica Policy Research. June 15, 2016. P. 17.
- 20 <http://www.thebillionpricesproject.com>
- 21 [http://enriquefrias-martinez.info/yahoo\\_site\\_admin/assets/docs/purba2012\\_submission\\_1.12294110.pdf](http://enriquefrias-martinez.info/yahoo_site_admin/assets/docs/purba2012_submission_1.12294110.pdf)
- 22 <https://science.house.gov/sites/republicans.science.house.gov/files/documents/HHRG-114-SY18-WState-BMyers-20160608.pdf>
- 23 <http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information>; see also <http://odimpart.org/developingeconomies.html>
- 24 For a comprehensive list of barriers see, e.g., a study on public health data sharing barriers at [www.biomedcentral.com/1471-2458/14/1144](http://www.biomedcentral.com/1471-2458/14/1144).
- 25 See <http://datacollaboratives.org> for a description of various types of cross-sector collaboration.
- 26 <http://movement.uber.com/>
- 27 <https://www.jpmmorganchase.com/corporate/institute/about.htm>
- 28 [https://blog.twitter.com/engineering/en\\_us/a/2014/introducing-twitter-data-grants.html](https://blog.twitter.com/engineering/en_us/a/2014/introducing-twitter-data-grants.html)
- 29 <http://www.unglobalpulse.org/blog/data-philanthropy-public-private-sector-data-sharing-global-resilience>; <https://medium.com/@sverhulst/data-responsibility-a-new-social-good-for-the-information-age-4463cb15798f>
- 30 <https://www.cep.gov/content/dam/cep/report/cep-final-report.pdf>; [https://obamawhitehouse.archives.gov/sites/default/files/omb/mgmt-gpra/barriers\\_to\\_using\\_administrative\\_data\\_for\\_evidence\\_building.pdf](https://obamawhitehouse.archives.gov/sites/default/files/omb/mgmt-gpra/barriers_to_using_administrative_data_for_evidence_building.pdf)

- 31 <http://reports.opendataenterprise.org/2016opendataroundtables.pdf>
- 32 <https://www.nap.edu/read/24652/chapter/7>
- 33 <http://www.pewinternet.org/2017/01/26/americans-and-cybersecurity/>
- 34 See, e.g., <http://science.sciencemag.org/content/347/6221/479.full>
- 35 Correspondence with the Census Bureau
- 36 <https://arxiv.org/pdf/1705.07872.pdf>; <https://www.theatlantic.com/technology/archive/2015/07/fake-data-privacy-census/399974/>
- 37 <https://www.resdac.org/cms-data/request/cms-virtual-research-data-center>
- 38 <https://www.bjs.gov/content/pub/pdf/ardprs1516pf.pdf>
- 39 [https://www2.census.gov/adrm/fesac/2016-12-09/Aladangady-Presentation-\(FESAC-12-9-16\).pdf](https://www2.census.gov/adrm/fesac/2016-12-09/Aladangady-Presentation-(FESAC-12-9-16).pdf)
- 40 For this and other examples, see <https://www.nap.edu/catalog/24652/innovations-in-federal-statistics-combining-data-sources-while-protecting-privacy>.
- 41 <http://connectingcredentials.org/wp-content/uploads/2016/09/Action-Plan.pdf>
- 42 [https://www.justice.gov/sites/default/files/oip/legacy/2014/07/23/exemption6\\_0.pdf](https://www.justice.gov/sites/default/files/oip/legacy/2014/07/23/exemption6_0.pdf)
- 43 For example, the federal government received nearly 800,000 such requests in 2016. <https://www.foia.gov/index.html>.
- 44 <https://www.cep.gov/content/dam/cep/report/cep-final-report.pdf>
- 45 See, e.g., <https://arxiv.org/pdf/1705.07872.pdf> (sharing of granular OPM wage data)
- 46 <https://www.cep.gov/content/dam/cep/report/cep-final-report.pdf>
- 47 See, e.g., [https://obamawhitehouse.archives.gov/sites/default/files/omb/mgmt-gpra/barriers\\_to\\_using\\_administrative\\_data\\_for\\_evidence\\_building.pdf](https://obamawhitehouse.archives.gov/sites/default/files/omb/mgmt-gpra/barriers_to_using_administrative_data_for_evidence_building.pdf)
- 48 That is not to criticize existing organizations that facilitate data sharing. In many cases, our approach may not be necessary, for example when there is great trust among or between data provider(s) and data users, or when the data are not particularly sensitive. Additionally, some of these organizations could become this type of intermediary. Or they could remain in their current state and use the intermediary when they need to share data with another organization. If and when the NSDS is created, that, too, could use the intermediary when sharing data with others.
- 49 See, e.g., [www.nsc.org/Connect/NSCNewsReleases/Lists/Posts/Post.aspx?ID=182](http://www.nsc.org/Connect/NSCNewsReleases/Lists/Posts/Post.aspx?ID=182).
- 50 In the interest of simplicity, we list only some of the relevant data sources. Additional data sources include 911 calls, use of Naloxone (a drug used to treat overdoses), and hospital emergency department visits.
- 51 Correspondence with the Census Bureau
- 52 <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.461.3362&rep=rep1&type=pdf>; [https://ec.europa.eu/eurostat/cros/system/files/S19P4\\_0.pdf](https://ec.europa.eu/eurostat/cros/system/files/S19P4_0.pdf)
- 53 [www.edelman.com/global-results/](http://www.edelman.com/global-results/)
- 54 <https://onthemap.ces.census.gov>; [https://lehd.ces.census.gov/applications/help/onthemap.html#!confidentiality\\_protection](https://lehd.ces.census.gov/applications/help/onthemap.html#!confidentiality_protection)
- 55 <https://www.resdac.org/cms-data/request/innovator-research>
- 56 <https://govex.jhu.edu/open-data-how-much-does-it-cost/>



Beeck Center for Social Impact + Innovation  
Georgetown University  
3700 O Street, NW, ICC-100  
Washington, DC 20057

[beeckcenter.georgetown.edu](http://beeckcenter.georgetown.edu)  
@BeeckCenter