MODERNIZING U.S. DATA INFRASTRUCTURE:
Design Considerations for Implementing a National Secure Data Service to Improve Statistics and Evidence Building
Acknowledgments

The Data Foundation and authors thanks the Alfred P. Sloan Foundation for its generous support for this project. The Data Foundation would also like to thank the dozens of experts who participated in discussions and roundtables to inform the options and recommendations presented in this report. The Data Foundation especially thanks 12 reviewers who provided valuable feedback and advice on earlier drafts.

Disclaimer

This paper is a product of the Data Foundation, with funding support provided by the Alfred P. Sloan Foundation. The findings and conclusions expressed by the authors do not necessarily reflect the views or opinions of the Data Foundation, its funders and sponsors, or its board of directors.

Disclosure

Both authors of this paper were affiliated with the U.S. Commission on Evidence-Based Policymaking. Dr. Potok was an appointed member of the Evidence Commission and Dr. Hart served as the policy and research director.
The need for using data to generate insights that can help improve American society is vast and urgent. Increasingly, researchers need capabilities to link together data collected through formal surveys, federal program administration, and non-governmental data sources. However, the lack of coordination throughout the federal government’s decentralized data infrastructure and statistical system limits the ability to generate the relevant, timely information demanded by policymakers.

Building on the initial recommendations from consensus panels of experts in recent years, we propose a strategy for developing a National Secure Data Service that would revolutionize the federal government’s data analysis capabilities, while promoting and even expanding privacy protections available today. The data service would modernize the country’s antiquated, inefficient, and often ineffective data infrastructure for research to develop a modern, cutting-edge system that would substantially advance evidence-based policymaking capabilities in the United States.

This paper explores the nuances of multiple approaches for implementing a data service, based on core criteria for developing a successful infrastructure. The criteria include consideration of attributes for transparency and public trust, legal authority for privacy protections, independence, ability to access data, scalability, sustainability, accountability, and intergovernmental cooperation. Four different approaches are weighed:

1. Establishing a New Agency at the Commerce Department.
2. Re-Tasking an Existing Agency at the Commerce Department
3. Creating a New Federally-Funded Research and Development Center (FFRDC) at the National Science Foundation (NSF)
4. Launching a Public-Private Partnership in a University Consortium

In reviewing the capabilities for each option, each of which would be improvements from the status quo, we recommend launching a new center in NSF. The FFRDC approach has many advantages over other options. It would afford considerable oversight, transparency, agility, and accountability for the envisioned activities paired with technical skills, credibility, and infrastructure offered by NSF, one of the country’s leading institutional sponsors of research. NSF operates within existing legal frameworks for protecting privacy, yet still advances strategies for enabling researchers and data users to conduct evidence-building activities. In addition, NSF’s deep ties to researchers in the natural and social sciences, as well as its support of advances in computer and data science and security, lead us to conclude that NSF would most successfully launch and sustain a researcher-centric data service. A final section of the paper outlines a roadmap for implementing such an approach and poses questions for consideration by the Federal Advisory Committee on Data for Evidence Building.

While aspects of the proposed National Secure Data Service could be implemented administratively by NSF, we call on Congress to authorize the project and specify oversight, transparency, and accountability preferences. The burden for using statistical and other protected data is high, even while the benefits are tremendous. If evidence-based policymaking is to succeed long-term in the United States, the American people must retain public trust in a system that serves their interests, protects their information, and advances policies that improve their quality of life, our economy, and society. Rapid action on a data service does just that by modernizing our country’s data infrastructure for the public good.
Introduction

The challenges facing decision-makers in modern society are complex, multi-faceted, and constantly evolving. In the public sector, finding solutions to today’s problems requires collaboration and partnership between industry, the research community, nonprofits, and government. The required coordination can be excessively challenging. Any single solution may involve multiple government agencies, different levels and branches of government, and other non-profit and private-sector entities. The organization and number of actors can complicate the measurement of how well proposed solutions address and solve defined problems once implemented.

This complexity fuels the ongoing need and growing demand for high-quality evidence that informs decision-making and public policy. Federal, state, and local implementing officials need to know how best to direct their constrained resources to achieve expected outcomes when addressing their greatest problems. Additionally, in order to hold government accountable, members of the public must be armed with sufficient knowledge to ensure that programs and policies are effective, agencies are operating efficiently, and regulations are designed with relevant, reliable information. Legislators need to understand which programs are achieving their goals and how to allocate limited funding.

While formidable, the goal of aligning the data needs of potential users and decision-makers with timely, appropriate capacity and resources is attainable. Over the past decade, researchers and government staff have successfully aligned capabilities to produce meaningful evidence that informed major policy actions in virtually every policy domain. The result has been measurable improvements in economic mobility, public health, food safety, environmental quality, child welfare protections, and homelessness policy.1 Unfortunately, the studies and research for these projects were often conducted at considerable expense, with delays caused by navigating data sharing and use agreements, and over long time periods measured in years or, in some cases, decades. Few were built on designs incorporating rapid succession, iteration, or sustainability, nor did they ensure a data infrastructure was in place to provide recurring insights and knowledge for policymakers.2 These types of lengthy, expensive studies are sometimes possible if and when all the right factors align perfectly. However, the factors rarely align in the most relevant time period for evidence to be useful to decision-makers—this is a fundamental problem.

The challenge for government officials and researchers is to jointly address how to meet the demand from decision-makers for useful, high-quality evidence with insights that are timely, relevant, reliable, and valid and that are detailed enough to inform policy while still protecting the privacy of individuals and businesses. Despite expertise and dedication from a highly-skilled data workforce inside and aligned with government, today our government is ill-equipped for meeting the information needs of modern society.

The historic approach to addressing government’s data needs is antiquated, inefficient, and too often, ineffective. Decision-makers at all levels of government are ill-served by a decentralized, under-coordinated, or fragmented data infrastructure. While decentralization admittedly offers some benefits for aligning data uses across agencies with topical needs and interests expressed by decision-makers, the approach has limits for data sharing across agency silos.3 Many of our existing systems and processes were established during the last century to collect and report data or statistics on a quarterly, annual, or biennial basis. Today’s information demands are different. Decision-makers are often confronted with problems that need insights within days, or even hours, and there is a growing expectation that insights from high quality data combined from multiple sources should be available much more quickly.

Data collected by one government agency or one level of government are often relevant for another when analyzing and understanding policies that span across traditional organizational structures and hierarchies. However, the infrastructure for
state and federal agencies to share and link sensitive data for social, behavioral, economic, science, and health research is understandably incomplete and fragmented. Many parts of the infrastructure are outdated and unable to keep up with demands to protect privacy, while allowing additional access, increased transparency, and reproducibility of results. The current situation is not one that was intentional or planned; it evolved over decades due to accretion that produced a patchwork of legal authorities for federal data access, use, and protection. The current situation embodies weak incentives for high-quality data reporting, particularly local and state data. The current situation is also no one’s fault in particular, because it involves the entirety of government - virtually every federal agency, multiple congressional committees, complex technical issues, political sensitivities, and countless stakeholders.

Even though the challenges are vast, taking a first step by creating a capability to more easily share data and provide privacy-protected, secure research access is desirable, possible, and within realistic reach. Solutions for securely connecting data are available. We provide a brief overview of recent proposals to devise improved capabilities for data sharing and use, largely focused on the federal government but with implications for state and local governments, researchers, and industry. Based on the recent proposals, we explore options for detailed implementation strategies that can be considered by policymakers in enabling data sharing and use for the public good through a National Secure Data Service. We conclude with a recommendation about a technically sound and politically feasible approach for rapidly implementing such a data service in the federal government.

**Principles for an Improved Data Infrastructure in the U.S.**

The vision of a modernized data access and linkage infrastructure in the United States for researchers inside and outside of government is within reach, aligning with consensus principles that inform the attributes described in this paper. Echoing the U.S. Commission on Evidence-Based Policymaking’s principles for data use and practice, any design of such an infrastructure must protect privacy and prevent reidentification of confidential data published in findings, be transparent regarding which data are being used and for what purposes, and enable the capacity to deploy cutting edge approaches and techniques for research and analysis. A National Academy of Science’s panel also echoed certain principles for the federal statistical system that are relevant, including relevance of the work for policy issues, credibility for data users, trust among information providers, and independence from external influences. Selected approaches should improve evidence-building timeliness and cost-effectiveness by streamlining access to sensitive or restricted data. Capabilities are also needed to deploy emerging analytical methods for machine learning, search functions, rich text analysis, metadata publication and curation, assessments of data quality, and or techniques.
Proposals for New Government Data Infrastructure Capabilities

The challenges and barriers to expanding the federal government’s capabilities for data access and sharing are not new. Over the past 25 years, various attempts to better integrate and prioritize government’s data capabilities were addressed periodically by congressional committees in authorizing legislation, including some government-wide reforms in the Paperwork Reduction Act of 1995, the E-Government Act of 2002, and the Confidential Information and Statistical Efficiency Act of 2002. OMB also issued memoranda to agency heads to promote data sharing including guidance for improving how agencies leverage existing data to facilitate agencies’ programmatic work and better serve the American public and strongly encouraging federal statistical agencies to build interagency collaborations around sharing data.

Evidence Commission Recommendations

More recently, in 2016, with leadership from Senator Patty Murray and then-House Speaker Paul Ryan, Congress passed the Evidence-Based Policymaking Commission Act, which specifically established a panel of experts charged with developing a strategy for more effectively using the data already collected by government.

With the launch of the U.S. Commission on Evidence-Based Policymaking in 2016, the federal government established a formal advisory body suited to conduct a comprehensive study for how to improve evidence building related to federal policymaking. The Evidence Commission’s 15 politically-appointed members were specifically required by law to consider whether a “clearinghouse” for program and survey data should be established and, if so, how to create such a clearinghouse. But the duties of the commission and the charge defined by Congress were expansive in law, beyond exploring the clearinghouse question. The Evidence Commission also weighed strategies for building program evaluation capacity in agencies, what data should be prioritized for supporting policymaking, what infrastructure should be used, approaches for improving access for research, and strategies for enhancing privacy protections, as well as identifying barriers and incentives for interagency data sharing.

Over the course of 18 months, the Evidence Commission studied the issues within its charge, gathering feedback from the public, researchers, and federal agencies. In late 2017, the Evidence Commission provided its final report to Congress and the President with 22 unanimous recommendations and a comprehensive strategy for expanding access to government data, enhancing privacy protections, and building capacity for evidence building in government.

“Without evidence, the federal government is an ineffective fiduciary on behalf of the taxpayer. Unfortunately, in many instances, federal decision-makers do not have access to the data necessary to best inform decisions. In such instances, agencies are unable to show the benefits or impacts of the programs they administer and cannot determine what, if any, unintended consequences are created by programs, or whether programs can be improved.”

House Committee on Oversight and Government Reform (2015)
In its findings, the Evidence Commission recognized that it is difficult and time consuming to collect, combine, and understand key information that can shed light on the effectiveness of government programs and spending. The tangled relationships between federal agencies, states, and local governments create many barriers when it comes to collecting and sharing data on individuals and businesses. Just as important, sharing data must be done in a way that is secure, maintains privacy, protects the confidentiality of data, and assures that information is used appropriately.

The linchpin for the Evidence Commission’s proposed strategy, around which the remaining recommendations and suggested action items were based, was in response to Congress’ request that the experts consider a federal data clearinghouse. The “clearinghouse” or warehouse Congress asked the commission to explore could be imagined as a single repository for government-collected data assets. The Evidence Commission rejected a large-scale data warehouse model due to its untenable privacy risks and practical limitations for implementation. Instead, the experts encouraged the establishment of a National Secure Data Service as a shared service for conducting temporary data linkages for exclusively statistical purposes. The data service envisioned by the commission members was a new entity in the federal government designed to facilitate access to data for qualified researchers and approved purposes, while also ensuring privacy and transparency for the data service’s activities. The commission’s proposed data service was also envisioned to support and enhance the existing infrastructure with new capabilities, rather than duplicating existing efforts.

Multiple recommendations from the Evidence Commission address different elements of the design for the data service, while leaving other details for future discussion and study. In total, 10 of the commission’s 22 recommendations explicitly referenced the data service in the recommendation or explanatory text, though the Evidence Commission

---

**National Secure Data Service Recommendations from the Evidence Commission**

The Evidence Commission provided 22 recommendations, 10 of which referenced the National Secure Data Service, including:

**Rec. 2-1:** Goal to facilitate data access (and linkage) for evidence building, or “statistical activities,” while also encouraging applications of privacy-preserving technologies.

**Rec. 2-2:** Envisioned as a shared service not a data warehouse, for temporary data linkage.

**Rec. 2-3:** Enable the Data Service to operate within the authorities of the Confidential Information Protection and Statistical Efficiency Act (CIPSEA), with modifications to the Privacy Act.

**Rec. 3-2:** Coordinate with federal departments on applications of privacy-preserving technologies.

**Rec. 4-1:** Established in the Department of Commerce as an entity unique from the Census Bureau.

**Rec. 4-2:** Operate with support of a steering committee.

**Rec. 4-3:** Operate with “unparalleled” transparency, including to disclose projects and audits of established rules and procedures.

**Rec. 4-4:** Operate with administrative flexibility, including for public-private partnerships and supported by user fees.

**Rec. 4-5:** With temporary linkage and a minimal core of data, rely on agency data inventories to identify relevant data assets including for data sensitivity.

**Rec. 5-5:** Authorized with sufficient resources to support, beyond user fees exclusively.
envisioned all of the recommendations working in tandem to support either the data service itself, or the infrastructure required to enable evidence-based policymaking to succeed in the federal government.

National Academy CNSTAT Recommendations

As the Evidence Commission was conducting its study, the National Academies of Science, Engineering, and Medicine convened a panel of experts under the auspices of the Committee on National Statistics (CNSTAT) to examine approaches for increasing data sharing among the federal statistical agencies. The focus was intently on improving the production of federal statistics, a purview encapsulated by the Evidence Commission in its broader scope, but which generated specific suggestions to enhance federal statistics. An initial consensus report produced by CNSTAT was available to the Evidence Commission in January 2017 as a resource in developing its recommendations.

Like the Evidence Commission, the CNSTAT panel's consensus report recognized the challenges created by a decentralized federal statistical system. It also recognized both the need and capability to simultaneously improve data access while strengthening privacy protections. Among the six recommendations offered by CNSTAT in the panel's initial report, it recommended that a new or existing entity should be designated or created to facilitate access to data. The recommendation acknowledged that while individual agencies could engage in data sharing and linkage activities, the panel perceived the best strategy would be for a shared resource across agencies. The recommendations also focused on the role of private-sector data sources to support production of official federal statistics and the role for emerging privacy-preserving technologies.

The CNSTAT panel published a second report that weighed different approaches and models for a data service-like entity, including additional recommendations for suggested attributes. Publication of this report coincided in late-2017 with the Evidence Commission’s final report.

Overlap Between the Evidence Commission and CNSTAT Recommendations

While neither the Evidence Commission nor CNSTAT provided analysis of how the recommendations related or overlapped, both groups of experts converged on common themes and approaches:

1. the establishment of a new entity in the federal government to support data linkage and combination activities,
2. mechanisms for improved data access, and
3. enhancements for privacy using new and emerging technologies.

The consensus represents a compelling convergence of approaches to consider in improving government data infrastructure.

Data Linkage, not Data Warehouse

Both panels of experts produced common, related findings and recommendations that qualified researchers, inside and outside government, should have more efficient mechanisms available to link existing, protected data. Further, the federal government could and should significantly improve how it provides secure access to those data for exclusively statistical purposes in connection with approved projects. The Evidence Commission was clear that it envisioned the entity as a service, not a data warehouse or clearinghouse; the CNSTAT panel reached the same conclusion. Instead, the service would link data for specific studies, requiring stringent privacy qualifications to ensure that data continued to be effectively protected while improving government officials’ ability to understand the impacts of programs on a wider range of outcomes. Agencies would continue to collect and store their own data separately and only provide the specific data needed for approved research. The Evidence Commission emphasized that the new entity should enhance efforts of the federal statistical agencies and assist in reducing the costs and increase the value of national statistics by integrating data from multiple data
sources including surveys; federal, state, and local administrative data; and private-sector data.

This recommendation was highly consistent with the final CNSTAT panel recommendation, though the Evidence Commission envisioned even wider access to researchers for evidence-building activities beyond just the production of official federal statistics. The CNSTAT panel examined several options and emphasized that any new entity should reinforce and enhance efforts of the federal statistical agencies to produce and improve statistical data. Therefore, the CNSTAT proposal meant that a new entity would need the same legal protections as a federal statistical agency, and data accessed through the entity would be used only for statistical purposes.

Enhanced Privacy Protection

Both the Evidence Commission and the CNSTAT panel recognized that protecting privacy needed to be a fundamental mission of a data service entity. Both envisioned that the entity would use a modern database, cryptography, privacy-preserving, and privacy-enhancing technologies, and be transparent with the public. Both emphasized that data accessed through the entity would need to be used only for statistical purposes and should not be used for any administrative, enforcement, or regulatory purpose that would affect the rights, privileges, or benefits of any individual person, business, or organization.14

The consistency in design of these varied groups of experts reflects the emerging recognition that gaps can and should be addressed. Reasonable and responsible solutions exist for enhancing data infrastructure that can achieve both privacy and accessibility expectations. Subsequent to the release of these reports, additional experts have recognized and echoed similar challenges.15

Evidence Act Enactment and Implementation

Following the publication of both the Evidence Commission and CNSTAT reports, a major legislative proposal called the Foundations for Evidence-Based Policymaking Act of 2018 (Evidence Act) was filed in Congress. The Evidence Act, which became law in 2019, was championed by then-Speaker Ryan and Sen. Murray as a starting point for improving federal data infrastructure; it was not intended as a final statement on what was needed but rather a strategic effort to rapidly begin making progress on the Evidence Commission’s vision with politically-viable legislation.17

In sum, Congress incorporated half of the commission’s recommendations in the Evidence Act. Among the provisions in the Evidence Act are the commission’s core capacity recommendations, such as new leadership roles like chief data officers, evaluation officers, and statistical officials; expectations for promoting data accessibility with a single application portal for researchers to access data, requirements for agencies to inventory and disclose existing data, and outlining certain expectations for openness in government data; and establishing new processes to connect supply and demand for data and evidence in the federal government. The Evidence Act also reauthorized and strengthened the Confidential Information Protection and Statistical Efficiency Act (CIPSEA), a privacy framework for use by statistical agencies and units in government that enables unique data collection and analysis capabilities in conjunction with strong privacy and confidentiality protections, enforceable with civil and criminal penalties. CIPSEA also positioned the federal statistical agencies as trusted agents for federal government data sharing efforts by mandating:

1. sharing of federal agency program data with statistical agencies for statistical purposes unless specifically prohibited by statute;

2. creation of a single application for researchers to request permission to access protected sensitive data;

3. creation of agency data inventories accessible by the public; and

4. classification by agencies of the sensitivity of their data and providing access appropriate to the level of sensitivity.
The Evidence Act did not establish the Evidence Commission-recommended National Secure Data Service. However, Congress did clearly express its interest in further exploring and advancing the concept once the foundational groundwork in the Evidence Act was implemented by Executive Branch agencies. The law included the establishment of an Advisory Committee on Data for Evidence Building (ACDEB), associated with one of the commission’s recommendations for establishing an advisory and oversight body for the data service.\textsuperscript{18} The ACDEB is specifically charged under the law with providing advice to the White House Office of Management and Budget (OMB) on implementation of new CIPSEA authorities, expected to be implemented for a data service to become operational.\textsuperscript{19} The ACDEB is also charged with providing recommendations about data linkage, data sharing, interagency coordination, and applications of privacy techniques in government within two years. The committee, chaired by the Chief Statistician of the United States, includes representatives from federal statistical agencies, chief data officers, evaluation officers, privacy officers, performance officials, and state and local government, as well as non-governmental, data and privacy experts.

The intent of Congress in including the ACDEB was explicitly to encourage further progress on implementation and exploration of a data service. The ACDEB charge in the Evidence Act aligns with the need to establish a clear vision, jointly acceptable to Congress and the President as well as viable for agencies to implement. An initial report from the ACDEB is expected in mid-2021 and a final report in 2022.

In the years since the Evidence Commission and CNSTAT reports were finalized, progress in improving data infrastructure across the federal government continued. The number of calls for more rapidly enhancing the infrastructure has also increased.\textsuperscript{20} During 2019 and 2020, OMB issued three memoranda to agency heads, laying out how to implement changes from the Evidence Act:

- **M-19-15** updated the implementation guidance for the Information Quality Act of 2000, to incorporate assuring the high quality of government data when using linked data from multiple sources.\textsuperscript{21}

- **M-19-23** gave instructions to agencies on Evidence Act implementation, including development of learning agendas to be informed by data, establishment of three new agency positions (Statistical Official, Evaluation Officer, and Chief Data Officer), data access for statistical purposes, and program evaluation.\textsuperscript{22}

- **M-20-12** gave further direction to agencies on best practices and principles for program evaluation.\textsuperscript{23}

Also during 2019, OMB published the Federal Data Strategy, a 10-year effort for federal agencies to better leverage data for improving their mission fulfillment and service delivery.\textsuperscript{24} The strategy contains 10 principles and 40 best practices agencies are expected to make progress on in coming years. Among the 10 practices outlined in the strategy, agencies must identify data needs to answer key agency questions, use data to guide decision-making, and convey insights from data. In addition, each year OMB intends to publish a new action plan for agencies with discrete steps, many of which directly align with expectations and requirements in the Evidence Act and prior OMB guidance. For example, in 2020, agencies are expected to begin to create annual learning agendas to articulate key questions relevant to policymakers and to devise the capabilities to answer these questions, calling upon coordinated federal data assets and shared data between state, local, and tribal governments and federal agencies to evaluate programs and policies.

We expand here on the past work from the Evidence Commission and CNSTAT, applying insights from implementation of the Evidence Act, OMB memoranda, and the Federal Data Strategy along with other contemporary issues facing federal agencies. We build on the idea of creating a shared service and infrastructure capable of providing modern access to sensitive data for statistical research purposes. Then we examine the necessary functions and structure of such an entity, the options for its organization and placement, and recommend an approach to making this much-needed capability a reality.
Improved data sharing and linkage of federal, state, and local government data, as well as academic and private sector data collections, requires an organized, cohesive effort to enable efficient and cost-effective use of relevant data. Both the Evidence Commission and the CNSTAT panel recognized the critical importance of establishing a new infrastructure quickly, but with appropriate feedback and careful design. Today, the ability to combine data within a secure environment that protects privacy is a vital component of the evidence-building community’s capacity to meet demand from policymakers for actionable insights. In the absence of a comprehensive service filling this need, gaps will certainly persist, affecting the production of reliable, timely information. This, in turn, limits the country’s ability to formulate effective strategies to respond to emerging crises and issues and improve government program outcomes for the American people.

DATA CHALLENGES FOR COVID-19 PANDEMIC RESPONSE

The COVID-19 pandemic demonstrates the critical need for improved data sharing and linkage. The absence of a data service contributed to gaps in timely COVID-19 information, which affected the country’s ability to apply insights for formulating effective strategies for virus response and economic recovery. In addition to medical and epidemiological information about the disease and its spread, information is needed to assess how businesses and individuals are faring, how safety net programs are working, and what is the right path to health and economic recovery. Policymakers at all levels of government need information quickly from multiple sources.

Had a National Secure Data Service existed prior to the onset of the COVID-19 pandemic, gathering and disseminating critical information quickly from multiple agencies and linking it together would have been more efficient and resulted in more timely information for decision-makers. The linked data would have provided multidimensional information that policymakers could have used to assess and project the economic and social effects of the pandemic on the nation. Important sources of data in this instance include the Census Bureau, the Bureau of Labor Statistics, the Centers for Disease Control (CDC) and its National Center for Health Statistics, the National Institutes of Health, the Centers for Medicare and Medicaid Services, the Department of Agriculture, as well as state and local health, labor, and education agencies.

If a data service had been available at the beginning of 2020, researchers inside and outside of government would have been able to more readily and rapidly conduct research using government administrative data, and if available, bring non-governmental datasets to the service for linkages for approved projects. Even with the pandemic underway, a data service could substantially improve the country’s ability to harness data responsive to the pandemic and future crises.
GAPS IN KNOWLEDGE FOR POLICY RESPONSES TO ADDRESS ECONOMIC MOBILITY AND OPPORTUNITY

While economic mobility and opportunity are persistent challenges for many Americans, the occurrence of a global recession and economic shock in 2020 highlighted gaps in data collection and management that limit decision-makers’ capabilities to determine appropriate policy responses. Data collected by different levels of government and private sector firms can produce insights for different demographics, though this information may be siloed, subject to use restrictions, or otherwise unavailable for analysis.

During the first half of 2020, 16 percent of American households reported they could not cover an unexpected expense of $400. Of those households, 61 percent experienced food insecurity and 31 percent were unemployed. There are also major disparities by race and ethnicity for behavioral responses and impacts in the mental health, food insecurity, and economic well-being domains. The economic shocks from the COVID-19 pandemic did not cause many of the disparities that are now present, but the pandemic has brought to light and exacerbated structural challenges in our society that limit economic mobility and opportunity for too many Americans.26

Data collected by many federal, state, and local agencies are directly relevant to understanding these issues. Income and earnings information is compiled through an array of different data systems, but often reflects a key outcome of interest. Administrative records for anti-poverty programs, including many operated by states on behalf of the federal government, offer insights into the government policies intended to improve mobility. Longitudinal data systems in education provide valuable information about workforce outcomes that may relate to economic conditions. A National Secure Data Service with the capability to link data across these various domains and topics would provide a more complete picture of how programs interact as well as how blended and intersecting policies affect overall mobility and security. Policymakers need to understand the complex dynamics that affect low income and underserved communities in order to tackle the barriers to upward mobility affecting those communities.
A National Secure Data Service would be able to support evidence building with programmatic data sets such as the Supplemental Nutrition Assistance Program (SNAP), the Special Supplemental Nutrition Program for Women, Infants, and Children (WIC), Temporary Assistance to Needy Families (TANF), Medicaid, Economic Development Administration programs like the Public Works and Economic Adjustment Assistance (EAA) and Planning and Local Technical Assistance, federal student loans, and many others. These programmatic data could be combined with valuable statistical data from agencies such as the Census Bureau, Bureau of Labor Statistics, National Center for Health Statistics, National Center for Education Statistics, Statistics of Income Division at the Internal Revenue Service (IRS), and others. The ability to look at cross-functional, cross-agency datasets, and potentially combine these with state-collected data, would be invaluable in understanding the effectiveness of these programs. It would also support new knowledge and insights about how well these policies and programs improve outcomes for the American people and businesses, as well as generate relevant insights to generally understand the population and economic conditions.

The data service’s purpose is not just about conducting entirely new or innovative projects. Once established, a fully operational secure data service could enable more efficient replication or extended research findings from evaluations conducted over the last several decades, including for major demonstration projects where the long-term outcomes initially may not have been analyzed. Data from these studies could be more readily linked with data from federal statistical agencies and other administrative datasets for assessing longitudinal outcomes. This was the precise strategy deployed by researchers who re-analyzed the Moving to Opportunity Demonstration Project revealing that initial evaluation findings missed major insights that could only be incorporated with longitudinal data analytics. In that case, while the researchers were able to successfully navigate the maze of bureaucratic processes to conduct the project, it was an effort that took years just to determine how to access and use the relevant data.

Other examples abound. For example, the Family Options Study completed by the Department of Housing and Urban Development to better understand long-term strategies for effectively reducing homelessness could be extended. The Social Security Administration could better understand policies that encourage return-to-work for beneficiaries of the $140 billion Social Security Disability Insurance program. Workforce training programs could be studied for long-term or multi-generational effects. In the 21st century, there is no reason why government should not be able to support the production of rapid and routine insights that inform how resources are spent addressing society’s greatest challenges.
EXAMPLES OF ANALYSIS FROM LINKED DATA RESULTING IN MEANINGFUL POLICY ACTION

Numerous examples of where combined data sources supported insightful and meaningful evidence to drive policy changes have occurred. However, many of these studies may take years or even decades to complete due to current obstacles and barriers to sharing data.

**FAMILY OPTIONS STUDY.** The federal government allocates billions each year to reduce nationwide homelessness levels. In 2017, more than 150,000 families with children accessed homelessness assistance programs in the U.S., demonstrating the need and scope of the problem in the country. Over the course of a decade, the Department of Housing and Urban Development (HUD) conducted a study in 12 communities to receive different types of services, including vouchers, rapid re-housing, transitional housing, and typical services or supports. To conduct the evaluation, HUD needed access to income and earnings data from the National Directory of New Hires, HUD administrative records, child welfare data provided by states to the Department of Health and Human Services, and new primary data collection conducted by evaluators. Results from the project provided compelling evidence that housing subsidies had significant and cost-effective impacts, relative to other interventions. This led the Obama Administration to propose an additional $11 billion for housing vouchers in the 2017 budget proposal.28

**TRAINING AND EMPLOYMENT POLICY.** When it comes to economic conditions for workers, there is widespread agreement that individuals in the workforce need appropriate skills to satisfy modern conditions and expectations. In 2005, the Labor Department initiated an evaluation project for interventions aimed at modernizing federal job training supports. The study was complicated in its timeliness due to delays from states in supplying requested data, and in which only one-quarter of states ultimately participated. After 11 years, researchers published results from the required study, while interim results were available sooner and used by the Obama Administration in describing the efficacy of existing job training efforts and in subsequent congressional inquiries.29

**REFORMS FOR DISABILITY POLICY.** In 2015 Congress passed the first reforms to the Social Security Disability Insurance Program in a generation. The reforms were designed around insights from a series of evaluations and demonstration projects conducted in the prior decade by SSA and government contractors. In addition, congressional staff were able to submit queries to SSA for near real-time analysis of administrative records available from the agency on payment accuracy, participant demographics, and other summary information relevant to decisions. The analysis of administrative data was foundational in supporting policymakers’ informational needs on potential reforms, and only possible because of efforts to enable data-quality and data linkage activities within SSA’s substantial data infrastructure. The reforms and political agreements reached on the policy averted program participants from experiencing a 20 percent reduction in benefits and ensured the solvency of the program for many additional years.30
The data service concept has also been offered as a mechanism and approach for promoting improved government capabilities for reproducing, replicating, and validating existing and ongoing research used in agency regulations. For example, it would become easier to access data that could answer questions about the economic or health effects of agency regulations for conducting regulatory analysis if a functional data service were in place. Many federal regulations are based on scientific data that are highly confidential and sensitive, thus difficult to access, making study findings difficult to replicate and greatly reducing transparency. The scientific community has called for more transparency and reproducibility to help improve the robustness and quality of scientific studies. A data service operating as a secure data enclave, would enable researchers evaluating federal programs, regulations, and scientific work to have the ability to bring together data from multiple federal agencies and other sources in a secure environment.

Numerous capabilities currently exist that can support or extend conceptualization of a National Secure Data Service, including Federal Statistical Research Data Centers, academic centers, and non-profit data enclaves. While these partnerships are important and useful for researchers and eventually policymakers, their capabilities for sharing and linking government data are constrained and limited. A data service that capitalizes on the existing capabilities while enabling new efficiencies and innovative approaches could rapidly enhance the entire evidence ecosystem.

The Evidence Act and the reauthorization of CIPSEA with expanded data sharing capabilities within the privacy framework set the stage for further reforms. As agencies are implementing the Evidence Act and bringing the foundational elements to daily practice through the Federal Data Strategy, policymakers must advance the next phase of this work. That phase is to create the bridge across the government’s decentralized data capabilities with a new entity that jointly maximizes data access responsibilities with confidentiality protections—a National Secure Data Service.
Resources and Existing Capabilities for Data Sharing to Support a Data Service

The National Secure Data Service would be building on the existing capabilities of the Federal Statistical Research Data Centers (FSRDCs), academic centers such as the Institute for Research on Innovation and Science (IRIS), Inter-university Consortium for Political and Social Research (ICPSR), and the Administrative Data Research Facility (ADRF), and non-profit efforts such as the NORC data enclave. There are currently 29 FSRDCs managed by the U.S. Census Bureau under the auspices of the Interagency Council on Statistical Policy. Established in the 1990s, they are secure brick and mortar facilities located at universities and federal agencies that support qualified researchers working with restricted-access microdata while protecting data confidentiality. Project work runs on servers maintained by the Census Bureau at its computer center. Work conducted at an FSRDC must be research for a statistical purpose. Seven agencies, mainly statistical agencies, allow researcher access to their restricted data through the FSRDC network.

IRIS is located at the University of Michigan. Through an agreement with the Census Bureau, IRIS links university administrative data with the demographic and business data from the Census Bureau, and these linked data can be accessed at the FSRDCs. ICPSR is also located at the University of Michigan within the Institute of Social Research and is an international consortium of more than 750 academic institutions and research organizations. It stores, curates, and provides access to scientific data so others can reuse the data and validate research findings. Data at ICPSR are enhanced with meaningful information to make it complete, self-explanatory, and usable for future researchers. ICPSR hosts data in a repository with powerful search capabilities. Indexed by all the major search engines, ICPSR data are easily discoverable and widely accessible to the public. Data at ICPSR are screened for confidentiality and privacy concerns. Stringent protections are in place for securing and distributing sensitive data. ICPSR hosted the OMB pilot project to provide one common application to researchers requesting sensitive data from statistical agencies, which is required by the Evidence Act.

The ADRF was initially created with Census Bureau funding but currently is operated independently by the Coleridge Initiative, a collaboration between researchers at New York University, University of Maryland, and the University of Chicago. It provides a secure platform on the Amazon cloud to host and provide remote access to confidential microdata and is certified to host sensitive federal data. The ADRF works with both state and federal agencies to link data and build evidence and also includes extensive, hands-on training of client agency personnel on data linkage and data science applications as part of its many services.

NORC at the University of Chicago is a non-profit social science research organization that offers a high-performance computational, storage, and analytical environment that a small number of federal agencies use to provide secure remote access to their confidential microdata for researchers.

Other academic institutions have collaborated with state and federal agencies to link data around specific research projects. There are a number of universities that have partnered with states to work with state administrative data to gain insights that better inform state planning.
Necessary Attributes of a Data Service

While neither the Evidence Commission nor the CNSAT panel developed detailed proposals with every feature for what a new entity or data service would look like, both offered suggestions for key characteristics, functions, and attributes. The Evidence Commission generally recommended the creation of a new legal entity within government to manage secure data linkages in a transparent way that protects privacy, while improving knowledge about the insights gained and exploring privacy-preserving technologies. As envisioned by the commission, the data service entity would have an exclusively statistical mission and a focus on data access and protecting privacy and confidentiality for building evidence across many topical areas. It would operate as an effective and efficient service that could be held accountable by policymakers and the American public.

The CNSTAT panel was primarily focused on sharing administrative data collected by agencies with the federal government’s statistical agencies, as well as sharing data among statistical agencies to improve federal statistics. The panel’s final report offered some specific perspectives and options for establishing such an entity, and its findings were consistent with the Evidence Commission’s broader view of the need for enhanced researcher access.

Accomplishing the broad range of expectations outlined by the commission and the CNSTAT panel suggests a series of at least eight distinct, interrelated attributes and functions necessary for a data service in practice. The identified attributes and functions include:

1. transparency and trust,
2. legal authority to protect privacy and confidentiality,
3. independence,
4. legal authority to collect data from agencies,
5. scalable functionality,
6. sustainability,
7. oversight and accountability,
8. and intergovernmental support.

Each is discussed further below.

Transparency and Trust

In order to maintain the public’s trust, the entity needs the capability to explain clearly what activities it is undertaking, the benefits of those activities, be guided by a set of standards for maintaining public trust (e.g., Statistical Policy Directive No. 1), and include a representative oversight infrastructure. This likely includes a steering committee or advisory board with representation from the public and other stakeholders. Such a board would provide guidance and oversight on the entity’s policies and procedures. To enable transparency, the entity would support a portal for sharing information about ongoing and completed evidence-building activities. The portal would have information about how confidential data are being used, create documentation for audits, monitor and report on compliance with policies and practices, and support information about the government’s data inventories. Researchers would be able to apply for access to data and obtain data documentation to help determine which data were needed to conduct studies. This would include consistent metadata, commentary on the quality of data and fitness for various uses, and information on other similar research projects.

Legal Authority to Protect Privacy and Confidentiality

Strong authority would be required to protect the privacy of data that are accessed and prevent misuse. At a minimum, that authority would need to be commensurate with CIPSEA and the Privacy Act, but
it may also require new legislation with additional safeguards. Staff of the entity would have legal authority to access key data sources, technical expertise to clean, curate, and link data, and would also provide technical assistance to federal statistical and program agencies, state and local program agencies, and external researchers. Staff would need to keep up with new data security practices to ensure that the information technology used to link and analyze data, updated as technology and methods advance. This includes the development, promotion, and application of new privacy-preserving technologies, and coordination with other federal agencies to develop and deploy the approaches which may require explicit demonstration authority and resources. The staff would need to innovate on ways to prevent inappropriate disclosure or reidentification of data once research results are released, in collaboration with other federal agencies.

**Independence**

Much like Inspectors General are afforded considerable independence within the Executive Branch’s operations, the entity should be empowered to set strategic priorities and to operate apart from policy and related offices. The entity would need to be able to support objective analyses and protect privacy and confidentiality at the same level as a principal statistical agency. Independence is also an attribute that relates to maintaining public trust, including ensuring that data are used only for approved statistical purposes and not used to take enforcement or administrative action against any individual or organization. The entity also would need the ability to prioritize support for evidence building across multiple agencies, rather than any one department.

**Legal Authority to Acquire Data from Agencies**

Legal authority to acquire and use data would need to be minimally consistent with capabilities provided to statistical agencies in the Evidence Act under CIPSEA’s Part D. Although the Evidence Act did not alter agencies’ existing statutory prohibitions to sharing data or program information, it provided statistical agencies a presumption that federal data should be made available to statistical agencies for evidence-building purposes, unless expressly prohibited by statute. The entity would need to rely on this authority and may well need stronger presumptions to be successful. Such authority could also include expanded capabilities for some existing data restrictions in Title 13, Title 26, and Title 42 (e.g., National Directory of New Hires), among other federal laws that currently include use restrictions.

**Scalable Functionality**

Scalability has several dimensions and is necessary in order to accommodate demand for high-quality evidence. The IT architecture should be such that cost-effective expansion is enabled without substantial capital investment as demand grows. The data service entity must be able to recruit, hire, and retain skilled personnel who understand data curation, record linkage, machine learning, statistical computing, data transmission, data encryption, the legal and regulatory framework around various data sets, identity disclosure avoidance, as well as IT architecture and cybersecurity. It also needs business and administrative practices that are secure, but avoid creating bottlenecks for the entity staff in conducting normal operations. The entity needs to also ensure efficiency for researchers in the project approval and start-up process.

**Sustainability**

A secure source of funding is necessary to assure continuity, oversight, and the ability to meet future demands. Researchers and potential system users will need some level of certainty, including among federal agencies, that if they are to invest time and resources to support developing the infrastructure that it will be sustained long-term. Long-term and stable funding will likely be best accomplished with partial or full direct federal appropriations. Direct authority to collect some fees and reimbursements is appropriate to support operations in conjunction with the Economy Act, however, as the service is established caution should be applied in how collections are assumed to support system development in advance of establishing full operational capacity. In addition, fees should not be prohibitive for
Collectively these eight attributes represent core capabilities and responsibilities envisioned for a data service, outlining the role and function it would play along with initial expectations for legal authorizations and policies that may be needed for the entity’s success.

19

researchers who may not have large funding grants, particularly at the start of their careers.

Institutional placement is another important element of sustainability. Whether the entity is placed within an existing federal department or operates as an independent agency could have a significant influence on the priority placed on its activities by the executive and legislative branches of government. Whichever strategy is used for the design, the data service will need a clearly-articulated and established business model that recognizes the substantial up-front investment needed to develop an entity but also the ongoing operational costs.33 Sustainability also requires sufficient staffing infrastructure with relevant and appropriate expertise to support data systems management, statistical methods, innovative approaches and applications, strict disclosure avoidance, and sound administrative management.

**Oversight and Accountability**

The capabilities of the data service should be coupled with strong oversight and accountability mechanisms. Congress, the Executive Branch, and other stakeholders need mechanisms by which they can be assured that the entity is responsible, ethically, and legally complying with its charge and mission. Capabilities for oversight, in addition to the governing body, could include mandatory external data audits, Government Accountability Office compliance audits, Inspector General performance reviews, periodic reporting to congressional authorizing committees, and OMB coordination. The Advisory Committee on Data for Evidence Building could also support initial oversight and accountability for the entity.

**Intergovernmental Support**

Recognizing the important role of state and local data collection and management relevant for federal policies and programs, the data service will need capabilities to coordinate and collaborate across governmental entities beyond the federal government. Such efforts should include formal mechanism for involving state chief data officers, workforce and employment agencies, and other key data providers and partners.
Organizational Options for a National Secure Data Service

While structural considerations about where a National Secure Data Service entity would organizationally be situated are not necessarily first-order relative to the questions about function and purpose, an organizational structure must be decided prior to developing authorizing legislation or policies to move a successful proposal forward for policymakers to consider. Under the options for different structures there are clear trade-offs, but also recognizable benefits. There are also common elements that can exist in any of the options, if appropriately specified in any new legal authority. Understanding and weighing the trade-offs is a core factor for determining which option is preferable in the near term.

We consider four practical options for organization and structure:

1. establishing a new statistical agency within government, specifically the Commerce Department,
2. re-tasking an existing government operation within the Commerce Department,
3. establishing a new public-private partnership within the National Science Foundation (NSF),
4. and developing a new university-based consortium.

These four options narrow considerably from the range of potential ideas and elaborations that could be suggested in a blue sky exercise or blank slate formulation. However, because we are building on the suggestions from the Evidence Commission and CNSTAT, we use the panels’ frameworks to inform the options with recognition of the realities of current government systems and processes. Each of the four options is weighed against the eight functions and attributes presented above.

Of note, we narrow consideration for which federal Cabinet-level agencies should support the data service in Options 1-3 to exclusively focus on the Commerce Department and NSF. Discussed further below under the respective options, these organizations have a combination of expertise, capacity, and certain existing legal authorities that uniquely position them within the federal government to support variants of a data service model. While the General Services Administration (GSA) provides a number of government-wide supports that conceptually align with the need for interagency data sharing capabilities offered as a service, GSA does not traditionally support the research community or intragovernmental collaboration. GSA also does not have experience with operating statistical agencies and working with data protected by CIPSEA, Title 13 (Census) or Title 26 (IRS).

Similarly, OMB provides interagency support and coordination, and houses key roles for government-wide information policy (e.g., Chief Statistician of the United States, federal Chief Information Officer, Privacy Branch) and some digital consulting (U.S. Digital Service). The U.S. Digital Service, while the names offer some similarities, has focused on valuable customer service enhancements for data systems and websites, rather than as an operator of data management infrastructure. We also strongly discourage the data service from being placed at OMB to ensure satisfaction of the principles outlined by the Evidence Commission and CNSTAT panel, including both real and perceived objectivity for the entity’s operations. OMB does not have an historic or legal framework suitable for supporting ongoing activities akin to the envisioned data service. Thus, we narrow the options to Commerce and NSF.

Some may question why a stand-alone agency newly created was not included in the option set. In short, our view which is consistent with the Evidence Commission, is that a stand-alone data service entity would not be achievable in short-order to meet the rapid need for its capabilities. Such a stand-alone
That said, once established and operational such an approach may be appropriate for further consideration or action.

The section that follows explores the four options for potential placement and structure of a data service that could be developed rapidly.

### DATA SERVICE ORGANIZATIONAL OPTIONS AND SATISFACTION OF KEY ATTRIBUTES

<table>
<thead>
<tr>
<th>OPTIONS → ATTRIBUTES ▸</th>
<th>OPTION 1: NEW AGENCY AT COMMERCE</th>
<th>OPTION 2: RE-TASK AGENCY AT COMMERCE</th>
<th>OPTION 3: FFRDC AT NSF</th>
<th>OPTION 4: PUBLIC-PRIVATE PARTNERSHIP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TRANSPARENCY AND PUBLIC TRUST</strong></td>
<td>Create new mechanisms, co-opt existing approaches from Census and BEA, and be covered by SPD1.</td>
<td>Create new mechanisms, reallocate from existing mechanisms, co-opt existing approaches from Census and BEA, and be covered by SPD1.</td>
<td>Create new mechanisms guided by the National Center for Science and Engineering Statistics (NCSES) and federal statistical principles and practices and be covered by SPD1 through NCSES.</td>
<td>Create new mechanisms, borrow from existing academic experience, and mimic principles in SPD1.</td>
</tr>
<tr>
<td><strong>LEGAL AUTHORITY FOR PRIVACY PROTECTIONS</strong></td>
<td>CIPSEA-eligible Privacy Act Ad hoc use limitations</td>
<td>CIPSEA-eligible Privacy Act Ad hoc use limitations</td>
<td>CIPSEA-eligible Privacy Act Ad hoc use limitations</td>
<td>Ad hoc arrangements in data sharing agreements</td>
</tr>
<tr>
<td><strong>INDEPENDENCE</strong></td>
<td>Subject to some Commerce Secretarial discretion and independent body oversight. May be at risk of political interference and prioritizing Commerce activities instead of interagency priorities.</td>
<td>Subject to some Commerce Secretarial discretion and independent body oversight. May be at risk of political interference and prioritizing Commerce activities instead of interagency priorities.</td>
<td>Quasi-independent, subject to some NSF Director discretion and independent oversight body. Less likelihood of being prioritized for NSF activities.</td>
<td>Quasi-independent with independent or third party oversight body.</td>
</tr>
<tr>
<td><strong>ABILITY TO ACCESS AND ACQUIRE DATA</strong></td>
<td>Yes, within CIPSEA framework</td>
<td>Yes, within CIPSEA framework</td>
<td>Yes, within CIPSEA framework</td>
<td>Limited</td>
</tr>
<tr>
<td><strong>SCALABLE AND TIMELY</strong></td>
<td>Yes, but would navigate existing business processes at Commerce. Likely slow development.</td>
<td>Yes, but would navigate existing business processes at Commerce and retool existing infrastructure. Likely slow development.</td>
<td>Yes, working with flexible NSF business processes. Allows for rapid development, particularly for acquiring capacity and skilled workforce.</td>
<td>Yes, with flexible and rapid development and hiring processes.</td>
</tr>
</tbody>
</table>

Table Continued on next page
<table>
<thead>
<tr>
<th><strong>OPTIONS ➔</strong></th>
<th><strong>OPTION 1: NEW AGENCY AT COMMERCE</strong></th>
<th><strong>OPTION 2: RE-TASK AT COMMERCE</strong></th>
<th><strong>OPTION 3: FFRDC AT NSF</strong></th>
<th><strong>OPTION 4: PUBLIC-PRIVATE PARTNERSHIP</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SUSTAINABILITY</strong></td>
<td>Can be funded by direct appropriations in addition to user fees and transfer authority across Commerce budget accounts. May experience funding competition with other Commerce activities. Existing and relevant staff expertise accessible via Commerce. Hampered by federal hiring barriers, including salary limits. Would rely on Commerce shared services for administrative support.</td>
<td>Can be funded by direct appropriations in addition to user fees and transfer authority across Commerce budget accounts. May experience funding competition with other Commerce activities. Existing and relevant staff expertise accessible via Commerce, but also would require reskilling or reducing current staff of re-tasked entity. Hampered by federal hiring barriers, including salary limits. Would rely on Commerce shared services for administrative support.</td>
<td>Can be funded by direct appropriations in addition to user fees and transfer authority across NSF budget and program accounts. Existing and relevant staff expertise accessible via NSF. High capacity for quickly recruiting and retaining highly skilled workforce with up-to-date skills. FFRDCs are operated on a contract basis, subject to competition; shifting contracts could challenge a sustainable service model.</td>
<td>May be eligible for direct appropriation contingent on structure and authority. Likely majority funded by user contributions and fees.</td>
</tr>
<tr>
<td><strong>OVERSIGHT AND ACCOUNTABILITY</strong></td>
<td>Subject to GAO, IG, and congressional committee oversight. Additional interactions from governing board and an advisory committee.</td>
<td>Subject to GAO, IG, and congressional committee oversight. Additional interactions from governing board and an advisory committee.</td>
<td>Subject to GAO, IG, and congressional committee oversight. Additional interactions from governing board and an advisory committee.</td>
<td>Subject to independent oversight. Additional interactions from governing board and an advisory committee.</td>
</tr>
<tr>
<td><strong>INTERGOVERNMENTAL COOPERATION</strong></td>
<td>Would need to establish new relationships, building on some existing Census partnerships; may jeopardize some existing cooperation among state and local governments.</td>
<td>Would need to establish new relationships, building on some existing Census partnerships; may jeopardize some existing cooperation among state and local governments.</td>
<td>Flexible collaboration model in an FFRDC; would need to establish new relationships, building on NSF partnerships.</td>
<td>Flexible collaboration model in academia; would need to establish new relationships.</td>
</tr>
</tbody>
</table>
Option 1: Establish a New Statistical Agency in the Commerce Department

In its final recommendations, the Evidence Commission suggested that the data service be established within the Department of Commerce. The recommendation aimed to avoid expanding the size of government while also building on existing knowledge and insights on the topic at the Census Bureau. When the commission was developing findings and recommendations, its members recognized that the Census Bureau had developed some infrastructure through the (now-defunct) Center on Administration Records Research and Applications in addition to longstanding efforts on data linkage and privacy protection. The Evidence Commission also recognized that a data service could benefit from efficiencies by leveraging administrative support services provided by the Commerce Department and perhaps take advantage of utilizing existing professional staff in the department’s two existing statistical agencies (the Census Bureau and Bureau of Economic Analysis), which have already established public trust in data activities and operational knowledge relevant for a data service.

Notably, the Evidence Commission preferred this option to both send a signal that the commission did not favor establishing the entity within the Census Bureau (thus the higher order statement about the Commerce Department). The commission also recognized unique capabilities at the Commerce Department and the need for specificity in the recommendations upon which policymakers could take action. In fact, the Evidence Commission did specifically consider whether the entity should be attached to the Census Bureau and determined the approach was not optimal for a number of reasons (discussed further in Option 2).

Under this option, existing statutory authorities that enable statistical data sharing, linkages, and privacy protection would be available for use by a new agency at the Commerce Department, designated as a statistical agency. Traditional oversight mechanisms at the Commerce Department Inspector General, OMB, GAO, and Congress would exist. This option could also allow for the use of extensive expertise in the Census Bureau, Bureau of Economic Analysis, National Institute of Standards and Technology, and the National Oceanic and Atmospheric Administration on core data infrastructure needs and issues. Use of an existing organizational structure bodes well for development and sustainability, even if existing processes limit or constrain the entity’s capabilities to be nimble and efficient to launch a new entity in rapid order.

The designation of the entity as a formal statistical agency ensures the use of the strong CIPSEA privacy framework, while also promoting ease for collaboration and cooperation with other statistical agencies. The new agency would also be able to participate in the Interagency Council on Statistical Policy, the Federal Committee on Statistical Methodology, and other inter-agency statistical activities, which would help establish relationships with existing federal statistical and program agencies to enable and support collaborative data management and sharing. Collaboration through the interagency bodies would also promote the creation of an organizational culture of service to other agencies and the research community.

The Evidence Commission specified that the selection of Commerce should be paired with sufficient independence from other Commerce Department operations and policy activities. Since 2017, the Commerce Department has demonstrated that such an approach is likely unrealistic as envisioned by the Evidence Commission, which could jeopardize the prioritization, implementation, and apolitical activities of such an entity. The Evidence Commission’s suggestion was optimistic that the Secretary of Commerce would politically and practically lend support to the evolution and development of the model, enabling the entity to set strategic priorities distinct from the Commerce Department without unwarranted influence from policy offices. Factors in the contemporaneous environment suggest such optimism may no longer be simply assumed.

Locating the data service in the Commerce Department would concentrate considerable influence over
cross-agency and intergovernmental statistical data and research studies under the sole control of the Secretary of Commerce. A similar consolidation of control was one of the objections raised when the President proposed moving the Bureau of Labor Statistics from the Labor Department to the Commerce Department for sharing of economic statistics. Location in the Department of Commerce could also subject the data service to budget cuts and reprioritization in line with the Secretary of Commerce’s priorities rather than government-wide priorities.

Most countries have a consolidated statistical system, usually a single office of national statistics. The most highly regarded of international statistics programs are given a measure of independence that is not found in the decentralized U.S. system, where statistical agencies are under the jurisdiction of the agency in which they are located. Although OMB is charged with coordinating these disparate statistical activities and setting standards for the quality and integrity of federal statistical data, the Chief Statistician at OMB has little leverage to influence cabinet secretaries on budget and organizational decisions, as well as priorities set for their agencies that affect the statistical activities under their jurisdiction.

Staffing a highly-skilled workforce in a rapidly-evolving data and computer science arena could also prove challenging to a federal agency. Currently, the federal government does not even have a data scientist job classification, although the U.S. Office of Personnel Management has been considering publishing a designation. Federal hiring can be slow, salaries for needed skills are not always competitive, and acquisition and implementation of modern IT is often time-consuming and burdensome. Bureaucratic constraints in current law and practice could impede the effectiveness of the data service mission.

In the context of coordinating with state and local governments, this option has potential to be successful, although it would be dependent on building out capabilities and establishing a new culture beyond the approaches at the Commerce Department today. While possible under this model, the lack of ownership or funding support for state-collected datasets relevant to the service within the Commerce Department’s domain of control may be limiting for ensuring long-term accessibility is sustained.

Above we rejected the possibility of situating a data service at GSA, which operates many government shared services. Despite GSA’s recognized role in coordinating management functions across agencies and even in supporting some interagency data activities (e.g., acquisitions, running the IT Modernization Fund), implementation of the envisioned data service extends far beyond simple coordination of officials or awarding grants to agencies. The culture, expertise, and strictness needed to operate a research-oriented statistical agency within the stringent limitations of CIPSEA is not well-suited for GSA’s operational model.

Option 2: Re-task an Existing Governmental Unit within the Commerce Department

Re-tasking an existing governmental entity within the Department of Commerce to assume the functions of a National Statistical Data Service could involve transforming an existing bureau to support implementation of the new entity. This would have many of the same benefits and challenges as Option 1, while also some unique considerations.

A re-tasked entity at the Commerce Department could realize capabilities for transparency, privacy protections, and data access under current law, including the ability to be designated as a statistical agency. Oversight and accountability mechanisms would also likely be substantially similar to Option 1. But any re-tasking or reallocation of responsibilities will inevitably face some resistance for the unit’s existing supports and program champions. However, using existing infrastructure and capabilities has the practical benefits of some existing operational processes, staff, and resources, as well as potentially some related or relevant legal authorities. Choosing among existing entities poses challenges, particularly given that existing entities carry existing cultural contexts and perceptions of missions that can be difficult to rapidly change.
Public trust could be difficult to build, or re-build, depending on the existing unit’s focus and degree of trust among relevant stakeholders in the American public, as well as congressional officials and inter-agency partners. Perhaps the greatest administrative obstacle faced by a re-tasking is recognizing that no existing unit in the Commerce Department is perfectly suited to fulfill the responsibilities and objectives for a data service. A re-tasking would likely result in leadership for any entity dedicating considerable time and energy to reskilling or replacing existing professional staff, which could distract from enthusiasm and motivation for focusing on the core mission of the data service.

One approach to implementing this option could be to re-task the National Technical Information Service (NTIS), as has been suggested by at least one non-profit advocacy organization with a focus on data policy. NTIS has the mission to help agencies make better decisions with data through supplying a structure to help agencies store, analyze, sort, and aggregate data. Originally created to be a repository for scientific research, NTIS is now an agency that primarily uses its authority to enter into joint ventures with the private sector to help agencies increase their capabilities. NTIS continues to receive critiques that the agency’s own existing infrastructure and approaches are obsolete. At least one bipartisan congressional proposal in recent years proposed shuttering the agency.

For addressing the key data service attributes, the NTIS model may be appealing because the existing statutory authority allows considerable flexibility to promote collaborative endeavors, and the agency has an existing mission relevant to the data service’s intended purpose. However, transforming NTIS into a statistical agency may not be any faster or simpler than creating a new agency in the Department of Commerce under Option 1. In fact, for NTIS this could be difficult and untimely because of the need to retrain or replace a substantial portion of the existing staff to meet the data service’s needs.

Alternatively, designating an existing statistical agency such as the Census Bureau as the data service entity would overcome some of the issues with relying on NTIS, but present still yet other issues. The Census Bureau’s lack of ability to share data under the restrictions within its authorizing legislation (Title 13 of the US Code) is limiting for operating a government-wide infrastructure; Title 13 would need to be amended, which is fraught with political and policy challenges. In addition, the Census Bureau would need to significantly change its focus away from prioritizing its own extensive portfolio of activities in order to give high priority to cross-governmental priorities that may have little to do with the Census Bureau’s core mission. Understandably, the Census Bureau currently prioritizes its own mission over those of the research community and other agencies; this prioritization is particularly acute during the years immediately prior to and during a decennial census. Researchers complain about the difficulties of using the FSRDCs managed by the Census Bureau, excessive wait times for project approvals and background checks, the antiquated IT infrastructure available to researchers, and the inflexibility and lack of transparency surrounding many activities and services desired by researchers. Even if these challenges were more perceptual than real, the frustrations experienced by external partners could substantially limit buy-in and willingness to participate in a new infrastructure. To the extent these limits are real, each could independently suggest the lack of capacity and processes to sufficiently support a data service. Further, the elimination of CARRA by the Census Bureau also suggests a lack of interest among leadership in serving in a government-wide service capacity that could be sustained long-term.

**Option 3:** Establish a Federally-Funded Research and Development Center Through the National Science Foundation

The establishment of a public-private partnership through a sponsored Federally-Funded Research and Development Center (FFRDC) offers the appeal of creating a quasi-governmental entity that could be responsive to intergovernmental needs as well as potential academic and industry users. Non-profit organizations that currently have state-of-the-art
Organizational Options

MODERNIZING U.S. DATA INFRASTRUCTURE

capacity for providing the functions of a secure data service, or the capability to rapidly develop them, could operate the FFRDC. There have been some calls from supporters of a data service to specifically use an FFRDC model.43

FFRDCs are widely used by the federal government, including 12 federal agencies that currently sponsor more than 40 FFRDCs.44 Originally launched to support national security purposes, FFRDCs today provide technical staff to support discrete projects, including those that may be long-term in focus or highly complex.45 FFRDCs include research and development laboratory capacity, analytical operations, and some system engineering capabilities.46 Guidance from OMB in 2011 reinforced that FFRDCs can support government operations, including in fulfillment of agency missions to conduct what OMB then termed “inherently governmental functions.”47 OMB’s guidance suggests that even if a data service’s activities were identified as an inherently governmental function, an FFRDC model would be permissible under current law and practice.

The Evidence Commission weighed the creation of an FFRDC during its deliberations, but noted that the classic FFRDC model tends to focus on specific projects for research activities, not operations for long-term or sustained infrastructure capabilities.48 The commission observed that a challenge with the FFRDC model might be that confidential or sensitive data would leave strict governmental control, though the existing OMB guidance on FFRDCs suggests this may be permissible under the FFRDC model.49 The CNSTAT panel similarly identified the FFRDC option’s appeal, in part to address cultural barriers to innovation in federal statistical agencies which tend to focus on data production for many key national indicators and may conflict with the need to innovate in methods, data sources, and techniques. Even prior to both panels’ recommendations, FFRDCs were considered a viable option for improving the federal statistical system.50

An FFRDC offers potential to satisfy key criteria for independence, while being partnered with a federal agency that also lends credibility, existing trust, a transparency infrastructure specified under government contract regulations (including specific FFRDC regulatory provisions), and an ability to access government data through contract vehicles. The model could scale quickly, building on the efficiency and business processes unfettered by onerous agency bureaucracy. FFRDCs are also subject to oversight not just from the sponsor agency, but because of the nexus to a federal agency also OMB, Inspectors General, GAO and congressional authorizing committees.

Whether an FFRDC model for a data service is sustainable depends on the availability of resources from a sponsoring agency. By design the contract for an FFRDC would align needed skillsets, along with the contractors presumably greater ease in hiring and skilling than may be possible within the federal government under current hiring authorities. An FFRDC would have fewer restrictions on compensating and retaining the technical staff needed to operate a data service. The fixed nature of the federal pay scale—which is often critiqued as not competitive with the private sector—and rigid hiring practices would make it difficult for a federal agency to attract an appropriate workforce, such as computer scientists, data scientists, IT specialists, and statisticians. While hiring reform in the federal government could address these limitations, absent such reforms the workforce issue will likely remain a substantial limitation.

FFRDCs tend to focus on specific topics rather than serve a broad cross-agency purpose for multiple agencies and external stakeholders. Breaking with this tradition, an FFRDC could serve as a demonstration project for expanded research capabilities for other regulatory, science, and evaluative research. An FFRDC could also be expanded to enable inter- and intra-governmental collaboration, as needed.

NSF is a logical choice to operate an FFRDC for a data service. NSF already sponsors five separate FFRDCs, giving the agency experience as a sponsor and a record of strong oversight that will assure that high quality, objective, independent data will be produced to provide the high quality evidence.51 NSF also has experience with creating and maintaining networks of researchers connected to various topics of interest.
NSF can encourage reproducibility, collaboration, low cost operations, application of innovative tools to enhance privacy and data integration, and a workforce trained in the integration of data science and domain knowledge.

While other agencies could sponsor a data service FFRDC, such as the Department of Commerce or the General Services Administration, neither agency has experience with FFRDCs to build upon. NSF is one of the only domestic, non-security agencies with an extensive record of supporting FFRDCs and an agency that already has a broad, research-focused mission. NSF’s experience in operating past FFRDCs will support rapid development and launch of a new operation.

Existing FFRDCs could also be re-aligned to support a data service, rather than establishing a completely new entity. In a sense, this implementation approach could expedite implementation timelines by relying on existing management and oversight structures. Unfortunately, none of the existing FFRDCs are a well-matched fit at either NSF or any other domestic agency. NSF’s existing FFRDCs tend to be geared to specific natural sciences rather than data infrastructure, which may not align for the objectives outlined here. Existing FFRDCs operated by other federal agencies are not well-suited for activities that may involve cross-agency, intergovernmental sensitive data collected from the American people under a pledge of confidentiality to be used only for statistical purposes. Instead, a new FFRDC could partner with existing operations to leverage capabilities and rapid adoption of the model, including through reimbursable arrangements. The Oak Ridge National Lab, for example, has been rapidly developing artificial intelligence and multi-party computation capabilities based on its hardware infrastructure to support high-computation loads. The existence of such an infrastructure could offer benefits around which other institutional issues could be designed, or could lead to partnering with the new entity to support the development of privacy-preserving technologies.

While there may be some appeal for using existing FFRDCs to achieve expediency, none of these operations have experience in operating a statistical entity protected by CIPSEA, including creating statistical datasets suitable for social science research and program evaluation; that subject matter knowledge will be critically important. Given the targeted activities for which existing FFRDCs have been developed and the cultures of their operations, changing culture and practices in an existing FFRDC would likely prove daunting.

Perhaps the most important reason an existing FFRDC operator should not be immediately assumed for this option is that the appeal for the FFRDC model is that it relies on innovation and ingenuity demonstrated through federal contracting processes. If an existing FFRDC administrator was identified through competition with peer institutions as the leading candidate for a contract, then that organization should pursue development of the data service in conjunction with NSF. Allowing other organizations and collaborative efforts from the non-profit community to propose strategies for effective, efficient implementation is a well-tested approach for ensuring the goals and objectives outlined here are achieved.

Option 4: Create a University-Based Public-Private Research Center

An alternative to the wholly government or quasi-governmental approach is to establish an entirely independent, non-governmental entity to serve as the National Secure Data Service. This option would be akin to the National Student Record Clearinghouse model, where governmental and non-governmental partners collaborate to conduct data linkage and analytical activities to support common goals and objectives. In its current configuration, the Administrative Data Research Facility run by the Coleridge Initiative, a consortium of three universities, also aligns with this conceptual model.52

The greatest strengths of this model are the capability to adopt the data service with some flexibility, while aligning with university-based researchers to promote a network of potential users. In many ways, this approach could yield the strongest intergovernmental and collaborative capabilities
because of the relationships academic institutions already have with state and local governments. The advantages also include the ability to attract and retain a highly-skilled workforce and also create a pipeline of students who could work in the federal statistical agencies. In addition, such an arrangement could be directly supported by multiple agencies, as opposed to a single contract run by one agency and would have natural links across the academic research community.

But the limitations of the model are considerable, including challenges accessing and linking to federal government data. While a university-based center could potentially be designated a CIPSEA unit in new legislation, under current law OMB could not designate a non-federal organization as a CIPSEA unit administratively. If that designation were not possible, the uniform CIPSEA privacy framework envisioned for a data service would not be attainable under this option. Similarly, existing federal directives and guidelines for maintaining public trust in government-collected data would not necessarily apply, even if the standards were voluntarily adopted by the entity. In sum, it is not clear how a university could operate like a federal statistical agency and have the authorities necessary to require agencies to provide data for research projects, particularly agencies with their own data protection statutes such as the IRS (Title 26) and the Census Bureau (Title 13). The university-based entity would have to rely on a patchwork of authorities and funding streams conferred by multiple agencies through individual contracts, which would not be much of an improvement over the current environment.

The Evidence Commission weighed a private entity as a potential option during its deliberations for the data service, but determined the path did not offer the same benefits or capabilities that the proposed governmental approach could have.\footnote{53} Similarly, the CNSTAT panel explored the model of a public-private research center managed by a university, such as IRIS at the University of Michigan. Other proponents of a national infrastructure have similarly suggested the value of relying on the country’s existing land grant university model to advance capabilities.\footnote{54} In order for this option to succeed, legislation would be needed to provide the entity with necessary legal authority to go beyond universities’ existing projects. In a sense, if no action is taken by the government to set up a national secure data service, we are likely to see a continuation of disjointed efforts at the individual agency-university level as a default option that is not likely to achieve the goal of an integrated, efficient infrastructure.
Each of the four options presented here for establishing a National Secure Data Service has its own merits, and each would be a substantial improvement over the status quo. One option rises above the rest as not only being viable, it achieves favorable assessment on most attributes with the fewest identified limitations – the FFRDC at NSF. Based on our review of the potential approaches for implementing a National Secure Data Service, consistent with the parameters outlined by the Evidence Commission in its unanimous recommendations and the CNSTAT consensus panel, we strongly recommend the federal government advance the establishment of the data service as an FFRDC at NSF. This section outlines specific action items and next steps necessary to implement this approach.

In general, authorizing legislation from Congress articulating the specific parameters of a National Secure Data Service as an FFRDC is desirable to reinforce key attributes and strengthen certain transparency and accountability mechanisms. Explicit congressional authorization will allay concerns from the public and bolster the case for the organizations’ demand, utility, and support in working with other federal agencies. In the absence of legislation or congressional action, NSF is well-suited to begin to advance a data service under existing legal and regulatory authorities. A variety of policy frameworks could advance the concept, including presidential executive orders or policy memoranda issued by OMB in order to reinforce certain attributes. NSF has the authority to begin to establish an FFRDC consistent with the agency’s mission, and existing law and regulation.

The Federal Acquisition Regulations (FAR) specify the conditions under which an FFRDC may be contracted by a federal agency. In general, such a contract requires the approval of an agency head, a competitive contracting process, articulation of an oversight and review process, funding and resources to support a contract, satisfaction of a policy statement to meet “some special long-term research or development need,” and a written sponsoring agreement. The sponsoring agreement is required to include a clear statement and purpose of an FFRDC, a series of administrative provisions, and the term of the agreement. However, in the context of the proposed National Secure Data Service, these minimum requirements for an FFRDC are insufficient to satisfy the obligations and expectations that such an entity would have, including to support an array of users. Additional expectations and requirements must be included in any contract issued by NSF, and could also be reinforced in clear congressional authorizations or an executive action.

Below are key aspects for implementing our proposal for establishing the FFRDC at NSF:

- **Responsibility and Organization within NSF.**
  The Social, Behavioral, and Economic Sciences (SBE) and Computer and Information Science and Engineering (CISE) research areas at NSF are ideally suited to sponsor a science and social statistical data FFRDC. NSF’s CISE has the expertise in privacy, confidentiality, and data science. NSF’s SBE has extensive experience and knowledge in human subjects research and includes a CIPSEA-covered principal federal statistical agency, the National Center for Science and Engineering Statistics (NCSES). An FFRDC focused on providing the data service would need suitable safeguards as well as designation to operate as a statistical agency under CIPSEA authority. By placing the FFRDC contract under the jurisdiction of NCSES, the FFRDC could be designated a CIPSEA agent. An administrative CIPSEA designation specifically permits the FFRDC to operate within the CIPSEA framework re-authorized by Congress in 2018, including both data sharing and use capabilities. While this attribute differs from the Evidence Commission’s Recommendation 4-1, in part, it allows for the same intent for the capacity to engage in record linkage and access. As a contractor to NSF, the
MODERNIZING U.S. DATA INFRASTRUCTURE

The FFRDC would be subject to federal oversight, and the NSF would be accountable to Congress and OMB. However, the FFRDC would have a focused mission and not be diverted by ongoing statistical data production activities. This focus would allow the FFRDC to build on NSF networks of researchers and focus on a shared service mission. Having the contract located in a statistical agency within NSF would provide the degree of independence and scientific objectivity that might not be found in a cabinet agency not dedicated to scientific research.

- **Transparency Mechanisms for the FFRDC.**
  Transparency must be considered paramount in all aspects of the data service’s design and operations. Congress should receive periodic reports about projects and activities that may be relevant for future decision-making. OMB officials should similarly have access to information that summarizes ongoing projects and decisions about data access and use. The data service FFRDC must also take steps to communicate with the American public about the value of the projects it is undertaking, the privacy protections that are in place, and what benefits are attributed to produced research. One aspect of transparency, recommended by the Evidence Commission, includes that the FFRDC publish a searchable directory of ongoing projects.

- **Accountability Mechanisms for the FFRDC.**
  Accountability is a series of activities through which organizations and officials can be monitored and reviewed for complying with stated practices, processes, and goals. GAO and the NSF Inspector General should periodically review the FFRDC to ensure adequate compliance with stated processes and CIPSEA authorities through performance and compliance audits. Congressional staff should provide insights through regular authorizing committee dialogues about the priorities and needs for the data service, including an annual oversight hearing. A mechanism for regular feedback from the users is also needed.

- **Interagency Cooperation Supports.**
  Of potential concern is the ability of NSF to gain cooperation of agencies across the federal government to provide data needed for research, through an FFRDC or otherwise. However, this could be a challenge under any option. To address this challenge, the Executive Office of the President should coordinate to issue an executive order or policy memorandum articulating support and direction for implementing the approach. The Executive Order could, for example, specifically reference the use of the OMB Director’s authority under the Paperwork Reduction Act to direct agencies to share data with each other and provide data to a statistical agency under the mandates of the Evidence Act. In conjunction with the executive order, the Interagency Council on Statistical Policy, the Interagency Council on Evaluation Policy, and the Chief Data Officers Council should issue a joint statement to federal agencies encouraging data sharing, improved access, and support for the FFRDC approach. In addition, NSF will need to establish a viable, streamlined business process for project approvals, particularly when projects require data from multiple agencies.

- **Contract Award Process.**
  The contract award for the FFRDC should be open and competitive. One possibility would be for an existing FFRDC to partner with existing non-profit entities to establish a new FFRDC that would take advantage of the management infrastructure and experience of the FFRDC combined with the subject matter expertise of the entities experienced in data linkage and related statistical activities. But any existing FFRDC or partnership should have to compete for a contract and demonstrate capabilities against bidders who may be creating entirely new FFRDCs.

- **Criteria for Data Service FFRDC Contract Award.**
  NSF must specify the criteria for awarding a contract for a data service FFRDC, to include that the awardee be best suited for joint (1) operation of core capabilities for data sharing, linkage, and compliance with CIPSEA protections, (2) development and deployment of current and future privacy-protective technologies in coordination with federal agencies, (3) coordination with federal agencies (sponsor and non-sponsors) as well as the research community and other qualified individuals for approved projects,
(4) operation within the guidance of an oversight committee, (5) demonstrated capability to recruit and retain qualified staff with appropriate and relevant expertise; (6) operation of business processes for project approvals, (7) maintenance of accessible project inventories, and (8) an ongoing program of continuous improvement in meeting customer needs. In addition, the FFRDC should be scalable and flexible enough to explore concepts such as a network of FFRDC spinoffs or partnerships that could be associated with topics, regions, or universities.

- **Governance Board.** Consistent with the Evidence Commission’s Recommendation 4-2, NSF should establish a governance board to provide general guidance on policies and practices implemented by the FFRDC. The board should include representatives from the federal statistical community and major program data providers, chief data officers, evaluation officers, privacy experts, researchers, and state or local government stakeholders. The governance board should produce publicly-available reports, participate in strategic planning, review periodic audits, and consult with NSF and OMB about resource allocations. The governance board should generally be independent from NSF operations and political decision-making processes to encourage and promote objectivity. The members of the governance board should be appointed by NSF in consultation with relevant federal agency stakeholders, but the board should be administratively supported by the FFRDC as part of the contract with NSF.

- **CIPSEA Designation.** Either Congress should designate the FFRDC as a CIPSEA agency in law, or the Director of NCSES as part of the FFRDC contract should designate the FFRDC as a CIPSEA agent. By designating the FFRDC as a CIPSEA agent, it would need to operate like a statistical agency. In practice, the Director of NCSES in NSF would be the federally-designated head of the data service, ensuring alignment with existing statistical agencies and CIPSEA organizations as well as inviting participating in the Interagency Council on Statistical Policy. In addition, the data service would need an Executive Director to run the day-to-day operations, and who would be an employee of the FFRDC and accountable to the Governance Board and NSF for compliance with CIPSEA.

- **Data Service Leadership.** The Executive Director of the data service should have experience in government data activities and deploying privacy protections. The Executive Director would need authority to carry out activities such as the research project proposal and access processes, secure microdata linkage and sharing, disclosure-protected dissemination capability, hiring authorities enabling the recruitment of technical staff, and acquiring the infrastructure to support evidence-building activities. Under an FFRDC contract model, all of these authorities are permissible and readily allocated to an executive director. Specifying these authorities in legislation would bolster the authority and clarity of the role.

- **Training Capabilities.** The FFRDC and NSF must be able to conduct a variety of activities that explicitly support training and education for potential users of the data service. Training should include low- or no-cost educational opportunities for internal government and external researchers, industry stakeholders, non-profits, and government agency staff, and should explicitly communicate the limitations and restrictions imposed by the CIPSEA privacy framework. Potential users should understand these important limitations prior to application processes.

- **Duration of an FFRDC.** Under current regulations FFRDC contracts are reviewed on a quinquennial basis. Potentially, the data service initially could be established with ongoing input from the advisory committee established in the Evidence Act. The performance of the FFRDC would be assessed at the 5-year mark, which could be an appropriate time to recompete or amend the contract.

- **Publication of a Government-Wide Learning Agenda for Researchers.** The Evidence Act’s requirement for agencies to produce strategic plans for research and evaluation can be supported by the activities of the FFRDC, which
should provide researchers a resource for understanding which questions identified in learning agendas could be addressed with available microdata within the existing infrastructure.

This resource could be compiled with analysis of individual agency learning agendas, following stakeholder feedback.

Additional Questions for the Advisory Committee on Data for Evidence Building

While our description in this paper offers many details for consideration of how to implement an FFRDC model, we leave several key questions for further discussion among the members of the Advisory Committee on Data for Evidence Building, which could extend the roadmap further:

• What additional oversight mechanisms might be needed beyond those provided by the governance board, GAO and the NSF Inspector General?

• What, if any, fee mechanisms or reimbursable arrangements should a data service deploy?

• What types of activities would be appropriate to cover with fees from the research community to the FFRDC? Should there be different levels for different types of research? Researchers? Or a waiver process?

• How should a data service prioritize initial project approvals?

• How should a data service establish a meaningful appeals process for projects or researchers that are not approved?

• What additional authorities may be needed to ensure the FFRDC has access to both statistical and programmatic microdata?

• How can existing use limitations in relevant data be considered and prioritized for consideration by lawmakers in exploring opportunities to maximize the utility of the data service?

• What additional incentives should be considered to promote development and adoption of privacy-preserving technologies?

• To what extent would partnerships with other FFRDCs or research collaboratives support the data service in fulfilling its mission, including for adopting privacy-preserving technologies and promoting partnerships?

• What analytic services should be offered an prioritized for the data service, beyond core data linkage activities?

• What services should the data service provide to encourage reproducibility?

• What policies and practices should be in place for the archiving and retention of data?

• How can a data service ensure continued collaboration for researchers and agencies?

• What processes should be considered to enable data discovery and data integration?

• How should user feedback be routinely incorporated?

• What are the most efficient processes for improving data quality in government datasets once users identify potential issues?

• How can the data service most effectively and efficiently ensure automated metadata documentation, linkage, entity resolution, stewardship and collaboration tools are available?

• How should NSF evaluate the FFRDC contract after the initial period of performance?

• What should the role of the Federal Statistical Research Data Centers be in implementing a network of data services and access points, leveraging existing infrastructure?
Conclusion

The data service envisioned here, and the entity designed and proposed by the Evidence Commission and CNSTAT panel, all achieve considerable improvements to the U.S. data infrastructure. Researchers would have more efficient and real access to relevant data, the American public could better understand how data are being used, and policymakers would have access to summary insights relevant to improving decision-making.

The technical capabilities exist to solve our country’s core data infrastructure problems; now policymakers must do their part to take action for enabling the solutions to move forward. The Evidence Act and implementation of the Federal Data Strategy take positive steps in addressing long-recognized gaps in capacity and capabilities. But neither goes far enough in solving the biggest problem: capabilities to access securely linked and combined data from multiple agencies. To solve this problem, the United States needs to move forward without delay to implement a National Secure Data Service. Based on this compilation of attributes and review of the options, placing the data service within an FFRDC at NSF will be cost-effective, timely, and sustainable.

NSF can administratively implement aspects of this option today, assuming adequate resources are appropriated by Congress or identified within existing authorized levels. Congress, OMB, and NSF should work to support and rapidly advance implementation of the FFRDC model. We strongly encourage that Congress promote and encourage development of the data service in the interest of meeting the growing demand for relevant evidence in the near-term, specifying oversight, transparency, and accountability preferences along with other policy parameters. In parallel, OMB and NSF can begin to initiate steps administratively that establish the foundation for implementing the data service as an FFRDC, building on the Evidence Act. Guidance through an Executive Order and a joint statement from relevant inter-agency bodies can encourage agencies to cooperate with NSF and the FFRDC to provide a one-stop shop for government and external researchers, while still enabling the existing infrastructure and capabilities to persist unimpeded. OMB’s Resource Management Offices, in conjunction with the Office of Information and Regulatory Affairs, can take a strong stand on agencies’ learning agendas and evidence-building activities by requiring rigorous evaluations of program outcomes and regulations. By advancing evidence building, insights into solving some of our nation’s most vexing issues can be developed, policies using those insights can be implemented, programs can be measured and reviewed for their effectiveness, and government can improve in service to the American people.

As the work moves forward to establish a National Secure Data Service, we appeal to government to prioritize transparency for the American public, and at all costs ensure that privacy protections are prioritized and enhanced over time. The burden for using data is high, even while the benefits are tremendous. If evidence-based policymaking is to truly succeed in the United States, the American people must retain public trust in the system that serves their interests, protects their information, and advances policies that improve their quality of live, economy, and society. Rapid action on a data service does just that by improving our country’s data infrastructure for the public good.
Endnotes


10. CEP, 2017a.

11. CEP, 2017a.


14. This concept is otherwise referred to as “statistical activities for exclusively statistical purposes.”


20. See, for example, Reamer et al., 2018.


34. We also rely on deliberative materials now publicly available from options considered by the Evidence Commission that provide additional context about information that was available for the commission during its deliberations on final recommendations. See CEP. “Discussion Document: Data Facility.” Memo to CEP Commissioners from CEP Staff. Washington, D.C., CEP, 2017c. See also K. Howell. “Follow-up to Recommendation Memo#2: Establishing a Data Facility.” Memo to CEP Commissioners from CEP Staff. Washington, D.C., CEP, 2017. Available at: http://datafoundation.org/s/Compendium-of-CEP-Staff-Decision-Memos-1.pdf.

35. The U.S. Digital Service which was created during the Obama Administration, is housed at OMB. The digital service model is not a data management infrastructure, however, and instead focuses on supplying and coordinating experts to improve website and data system services across the federal government.


Endnotes (cont.)

38. CEP, 2017a, p. 41.


41. For example, see U.S. Congress. “Let me Google That for You Act” S. 2206 in 113th Congress. Available at: https://www.congress.gov/113/bills/s2206/BILLS-113s2206is.pdf.


43. J. Lane. Democratizing our Data: A Manifesto. Cambridge, MA: The MIT Press, 2020. Note draft legislative proposal was reviewed confidentially by authors but is not currently publicly available.


48. CEP, 2017c.

49. CEP, 2017c.


52. Coleridge Initiative. Available at: https://coleridgeinitiative.org/

53. CEP, 2017c.

54. Lane, 2020.

55. 48 CFR Part 35. See also Gallo, 2017.

56. 48 CFR 35.017-1.

57. The FFRDC director and employees potentially could be designated by NSF as CIPSEA “agents” as defined by section 3561 of the act. The term “agent” is defined as an employee of a contractor with whom a contract or other agreement is executed by an executive agency to perform a statistical activity under the control of an officer or employee of that agency; or who is engaged by the agency to design or maintain the systems for handling or storage of data received under this subchapter; and who agrees in writing to comply with all provisions of law that affect information acquired by that agency.

58. See 44 U.S.C. 3510.
The Data Foundation is a non-profit think tank based in Washington, D.C. that seeks to improve government and society by using data to inform public policymaking. Our research and educational activities proactively and rapidly address relevant, emerging data-related needs in the country with the goal of devising realistic solutions, accelerating policy coordination, and advancing innovation. The Data Foundation values diversity and transparency in pursuit of an equitable, data-informed society.