Modernizing U.S. Data Infrastructure: Design Considerations for Implementing a National Secure Data Service to Improve Statistics and Evidence Building

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.

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.

Contents

Data Foundation (DF)

Description:

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.

Stakeholder(s):

Nick Hart, Ph.D. :
Co-Author — President, Data Foundation

Nancy Potok, Ph.D. :
Co-Author — Former Chief Statistician of the United States, White House Office of Management and Budget; Board Member, Data Foundation

Alfred P. Sloan Foundation :
The Data Foundation and authors thanks the Alfred P. Sloan Foundation for its generous support for this project.

Policymakers :
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 :
NSF can administratively implement aspects of this option today, assuming adequate resources are appropriated by Congress or identified within existing authorized levels.

Congress :
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.

OMB :
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 interagency 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.

Office of Information and Regulatory Affairs :
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.

American People :
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.

Vision

Statistics and evidence building are improved

Mission

To establish a national secure data service

Values

Transparency: TRANSPARENCY AND PUBLIC TRUST -- 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.

Trust

Authority: LEGAL AUTHORITY FOR PRIVACY PROTECTIONS -- CIPSEA-eligible Privacy Act Ad hoc use limitations

Privacy

Independence: INDEPENDENCE -- Quasi-independent, subject to some NSF Director discretion and independent oversight body. Less likelihood of being prioritized for NSF activities.

Accessibility: ABILITY TO ACCESS AND ACQUIRE DATA -- Yes, within CIPSEA framework

Scalability: SCALABLE AND TIMELY -- Yes, working with flexible NSF business processes. Allows for rapid development, particularly for acquiring capacity and skilled workforce.

Timeliness

Sustainability: SUSTAINABILITY -- 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.

Oversight: OVERSIGHT AND ACCOUNTABILITY -- Subject to GAO, IG, and congressional committee oversight. Additional interactions from governing board and an advisory committee.

Accountability

Cooperation: INTERGOVERNMENTAL COOPERATION -- Flexible collaboration model in an FFRDC; would need to establish new relationships, building on NSF partnerships.

Data Service

Establish a public-private partnership through a Federally-Funded Research and Development Center.

Stakeholder(s)

National Science Foundation :
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. 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.

Department of Commerce :
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.

General Services Administration

FFRDCs :
FFRDCs are widely used by the federal government, including 12 federal agencies that currently sponsor more than 40 FFRDCS. 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. FFRDCs include research and development laboratory capacity, analytical operations, and some system engineering capabilities.

OMB :
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." 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. ..

Federal Agencies :
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.

State Agencies :
An FFRDC could also be expanded to enable inter- and intra-governmental collaboration, as needed...

Local Agencies

Oak Ridge National Lab :
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.

FFRDC Administrators :
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.

Evidence Commission :
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

Stakeholders (continued)

infrastructure capabilities. 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.

CNSTAT Panel :
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.

Oversight Agencies :
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.

Inspectors General

GAO

Congressional Authorizing Committees

Sponsoring Agencies :
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.

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 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.

1. Responsibility & Organization

Assign responsibility and determine the organization of the FFRDC.

Responsibility and Organization within NSF.

Stakeholder(s):

NSF SBE :
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 CISE :
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).

National Center for Science and Engineering Statistics (NCSES)

1.1. Safeguards

Provide suitable safeguards for the data service.

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.

Stakeholder(s):

Congress :
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.

Evidence Commission :
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.

NSF :
As a contractor to NSF, 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.

OMB

1.2. CIPSEA Authorization

Obtain designation to operate as a statistical agency under CIPSEA authority.

2. Transparency

Make transparency paramount in all aspects of the data service's design and operations.

Transparency Mechanisms for the FFRDC. Transparency must be considered paramount in all aspects of the data service's design and operations.

2.1. Reports

Provide periodic reports about projects and activities.

Stakeholder(s):
Congress :
Congress should receive periodic reports about projects and activities that may be relevant for future decision-making.

2.2. Projects & Decisions

Summarize projects and decisions about data access and use.

Stakeholder(s):
OMB :
OMB officials should similarly have access to information that summarizes ongoing projects and decisions about data access and use.

2.3. Communication

Communicate with the American public.

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.

Stakeholder(s):
American Public

2.4. Directory

Publish a searchable directory of ongoing projects.

Stakeholder(s):
Evidence Commission :
One aspect of transparency, recommended by the Evidence Commission, includes that the FFRDC publish a searchable directory of ongoing projects.

3. Accountability

Monitor and review compliance with stated practices, processes, and goals.

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.

Stakeholder(s):

GAO :
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.

NSF Inspector General

Congressional Staff :
Congressional staff should provide insights through regular authorizing committee dialogues about the priorities and needs for the data service, including an annual oversight hearing.

3.1. Feedback

Enable feedback from users.

Stakeholder(s):

Data Service Users :
A mechanism for regular feedback from the users is also needed.

4. Cooperation

Gain cooperation of agencies across the federal government to provide data needed for research.

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.

4.1. Direction

Issue an executive order or policy memorandum articulating support and direction.

Stakeholder(s):

Executive Office of the President :
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.

OMB Director :
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.

4.2. Support

Issue a statement to federal agencies encouraging data sharing, improved access, and support for the FFRDC approach.

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.

Stakeholder(s):

Interagency Council on Statistical Policy

Interagency Council on Evaluation Policy

Chief Data Officers Council

4.3. Projects

Establish a process for project approvals

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.

Stakeholder(s):

NSF :
In addition, NSF will need to establish a viable, streamlined business process for project approvals, particularly when projects require data from multiple agencies.

5. Contract

Specify the criteria for awarding a contract for a data service FFRDC.

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.

6. Governance

Establish a governance board to provide general guidance on policies and practices implemented by the FFRDC.

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.

Stakeholder(s):

Federal Statistical Community :
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.

Data Service Governance Board :
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.

Data Service Governance Board Members :
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.

6.1. Reports

Produce publicly reports.

6.2. Strategic Planning

Participate in strategic planning.

6.3. Audits

Review periodic audits.

6.4. Resources

Consult with NSF and OMB about resource allocations.

Stakeholder(s):

NSF

OMB

7. CIPSEA

Designate the FFRDC as a CIPSEA agent.

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.

Stakeholder(s):

Congress

NCSES Director :
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.

Data Service Executive Director :
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.

8. Executive Director

Ensure the Executive Director of the data service is experienced in government data activities and deploying privacy protections.

Data Service Leadership. The Executive Director of the data service should have experience in government data activities and deploying privacy protections.

8.1. Authority

Authorize the Executive Director to carry out activities.

The Executive Director would need authority to carry out activities such as the research project approval 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.

9. Training

Support training and education for potential users of the data service.

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.

10. Establishment

Establish the data service with ongoing input from the advisory committee established in the Evidence Act.

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.

10.1. Performance Assessment

Assess the performance of the FFRDC at the 5-year mark.

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.

11. Learning Agenda

Publish a government-wide learning agenda for researchers.

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.

Stakeholder(s):

Researchers

Administrative Information
Start Date:
End Date:

Publication Date: 2020-07-30
Source: https://www.datafoundation.org/modernizing-us-data-infrastructure-2020

Submitter:
Given Name: Owen
Surname: Ambur
Email: Owen.Ambur@verizon.net
Phone: