Harvard Data Science Review • Issue 1.2, Fall 2019

Deep Policy Learning: Opportunities and Challenges From the Evidence Act

Nancy Potok¹

¹United States Office of Management and Budget, Executive Office of the President of the United States, Washington DC, United States of America

Published on: Nov 08, 2019

DOI: <https://doi.org/10.1162/99608f92.77e63f8f>

License: [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

1. U.S. Federal Data Modernization Opportunities

If you are reading this, you are most likely already a fan of accessible, high-quality data. Whether you are a producer or a user or both, new opportunities are opening up to access and work with U.S. federal data, thanks to the recently enacted *Foundations of Evidence-Based Policymaking Act of 2018*¹ and the inaugural [U.S. Federal Data Strategy](#). The Evidence Act and the Data Strategy are a powerful match. Their collective vision: to create partnerships between U.S. federal agencies, state and local governments, academia, and industry to realize effectively the value of shared federal data—accomplished by putting ‘open’ nonsensitive data in the hands of the public and using secure technology to increase legitimate researcher access to more restricted, sensitive data while still protecting privacy and confidentiality of those data.

U.S. federal data modernization efforts have opened a vast array of opportunities. These opportunities are available whether you are in government, academia, or the private sector. As agencies continue to make progress on the modernization agenda, you might consider becoming involved in making a meaningful contribution to data science in the public policy and management arenas.

This article explains some key activities underway to improve the quality of and access to federal data. The purpose of these activities is to advance evidence-based decision making in government, use data to improve government operations, and put more open data into the private sector and in the hands of the public, so that these data can be commercialized and used to help the public become more informed. The U.S. government collects vast amounts of data, from weather satellites to the population census to crop estimates to Medicare and Medicaid data, to name a few. Open data available to the public can be found through [the government's data portal](#).

2. U.S. Federal Data Strategy Takes a Holistic View

To help agencies leverage their data as a strategic asset, the Federal Data Strategy includes four components. These components are the building blocks and guides for federal agency actions over the next several years. The first component, *Enterprise Data Governance*, includes—but is so much more than data management as we often think of it—standardizing metadata, creating inventories, safeguarding confidentiality and privacy, and so on. The more expansive governance vision includes collaboration across agencies and agency program silos in order to bring multidisciplinary expertise together to formulate and address the ‘big questions’ that have been so difficult for agencies to tackle. Agencies know what questions are important, but often lack access to the data they need to answer those questions, because the data are held by multiple agencies or are hard to find and use. To be successful means changing federal agency cultures not only to ask priority questions that are meaningful and specific to the agency, including operational and mission-strategic questions, but also to share data across silos within and across agencies. The change for many agencies will be that the

priority questions to be answered must drive the research methods, rather than methods being determined by what data have been readily available in the past.

The second component of the data strategy focuses on *Access, Use, and Augmentation* of data. It calls on agencies to make data available to the public more quickly and in more useful formats. In addition, agencies should be using the best available technologies to increase access to sensitive, protected data while protecting privacy, confidentiality, and security, including the interests of the data providers. The Evidence Commission envisioned a National Secure Data Service that would be a center of excellence, designed to keep up with emerging technologies in areas such as secure access to data and privacy protection. The strategy's action plan calls for the creation of toolkits and methodologies to help agencies build their own competencies as well. Agencies would also be expected to seek out new sources for building data sets, which could include commercially available data and data from state and local governments.

Decision Making and Accountability addresses the need for policy and decision makers to increase their use of high-quality data and analyses to inform evidence-based decision making and improved operations. In addition, increased government accountability and transparency should be achieved by providing accurate and timely spending information, performance metrics, and other administrative data. Agencies are expected to use the most rigorous methods possible that align and are appropriate to answering the identified 'big' questions. Agencies may answer questions using existing evidence, including literature reviews, meta-analyses, and research clearinghouses, but are encouraged to also explore opportunities for acquiring new evidence, including utilizing outside expertise.

Finally, the federal agencies need to facilitate the use of government data assets by external parties, such as academic researchers, businesses, and community groups. To accomplish this through *Commercialization, Innovation, and Public Use* will require agencies to reach out to partners outside of government to assess which data are most valuable and should be prioritized for making available. There are many examples of entrepreneurial companies that have taken public data to create new apps that benefit the public as well as founding new economic engines, such as weather and geographic mapping companies. This part of the strategy seeks to accelerate that long-standing practice by releasing more data to the public.

The Federal Data Strategy was developed by work teams consisting of people from many federal agencies, under the guidance of the U.S. Office of Management and Budget (OMB), Office of Science and Technology Policy (OSTP), Department of Commerce (DOC), and Small Business Administration (SBA). As the strategy was being developed, parts of it went out for public comment. Several roundtables and town halls were held to gather further input from the public, industry, and academia. The resulting products consist of a vision, a mission statement, 10 guiding principles, a set of 40 practices to guide agencies, and a first-year action plan to implement the strategy. The mission statement, principles, and practices were issued by OMB as a memorandum to agencies (M-19-18).²

3. Why the Evidence Act Matters

Turbocharging the effort to leverage the power of federal data are the new statutory provisions in the Evidence Act. Why does the Evidence Act matter so much? First, it creates a new model for agencies to rethink how they build evidence and use data. It focuses the leadership of the agencies on collaborative problem solving, forcing cooperation across organizational data silos. And it builds on long-standing principles underlying federal data infrastructure and policies.

How does the Evidence Act accomplish this? It starts by requiring agencies to develop evaluation plans tied to their strategic goals. Agencies then create learning agendas focused on first asking the big questions, and then getting the information needed to answer those questions. What kinds of questions might agencies have? The act envisions that agencies will begin to understand the longer-term societal outcomes of their programs, be able to visualize the results of multiple federal programs in various geographic areas, improve their operations, and better serve the public. Statistical data agencies, such as the Census Bureau and the Bureau of Labor Statistics, should be able to acquire and combine data more easily from multiple sources to create new data products. To help create and carry out their learning agendas, agencies are required to designate a Chief Evaluation Officer, a Senior Statistical Official, and a Chief Data Officer. The qualifications, data governance requirements, and general expectations around these activities are laid out in the OMB guidance memorandum to agencies, [M-19-23](#).

The Evidence Act was formulated around 11 of the recommendations made by the bipartisan Commission on Evidence-Based Policymaking, which submitted [its report to Congress](#) in September 2017. The commission was very concerned with increasing access to data for statistical purposes, defined as the description, estimation, or analysis of the characteristics of *groups*, without identifying the individuals or organizations that comprise such groups, for activities such as producing statistics, program evaluation, and policy analysis.

The Evidence Act requires that agencies become more transparent with their data and create a comprehensive data inventory and data catalogue accessible to the public that can be accessed through a single site for the federal government. In addition, each agency must create an Open Data Plan. To help facilitate easier access to protected statistical data, the act mandates that a single application be developed and put in place for researchers to request access to statistical agency data. Currently, each agency has its own application, making the process cumbersome for researchers. In addition, agencies must categorize data into tiers according to their sensitivity and allow access based on that sensitivity. For example, the most protected data, such as microdata collected under a pledge of confidentiality in a census or survey, would only be accessed through [Federal Statistical Research Data Centers](#) (FSRDCs) or other secure data enclaves or facilities. Open data or public use files could be put out on the Web. New for most of the statistical agencies: other agencies must share data with a statistical agency upon request, unless a statute expressly prohibits sharing the data. The statute is not detailed on how this occurs, rather it charges OMB with developing regulations on how the sharing will take place.

These changes are quite significant, and the act envisions that they will occur with some standardization across the federal government. How will that take place? As mentioned, the act mandates that OMB issue guidelines and regulations on how agencies should provide data to statistical agencies, as well as how agencies should categorize their data into tiers or ‘buckets’ according to sensitivity. In addition, as mentioned above, OMB is to develop a single application for researchers to use when requesting data from a statistical agency for their research projects. The process for approving these research projects will follow existing procedures that are in place at each agency but are expected to evolve over time as experience is gained. Finally, the act requires OMB to establish an Advisory Committee to review, analyze, and make recommendations on how to promote the use of federal data for evidence building. The Chief Statistician of the United States serves as the chair of the Advisory Committee. The Advisory Committee will evaluate and provide recommendations to the OMB director on how to facilitate data sharing, enable data linkage, and develop privacy-enhancing techniques, and review the coordination of data sharing or availability for evidence building across all agencies. One concept for how this could be done is exemplified by the commission recommendation for establishment of a National Secure Data Service. The Advisory Committee has two years to develop recommendations.

4. Ways You Can Contribute Your Expertise

At this point, you may be wondering how you can contribute to the development of these guidelines and regulations, or even participate in the Advisory Committee. The good news is that there are many opportunities available to participate. Following its standard procedures, OMB will issue draft guidelines and regulations on increasing access and categorizing sensitive data in a series of Federal Register Notices soliciting public comment. There will also be a Federal Register Notice soliciting nominations for members of the Advisory Committee. Look for these to be released during the next few months.

In addition, the new Chief Data Officers will have a cross-agency council to plan activities jointly, share information, and create affinity groups around shared data interests. The new Statistical Officials will be joining the current 13 statistical agency heads on the Interagency Council on Statistical Policy, and the Evaluation Officers will form a cross-agency council as well. The challenge for all of these councils will be for their members to stay closely aligned with the business lines of their agencies so that their expertise can be used to create value for those who are collecting and managing the data to run their programs.

If you are a data scientist, computer scientist, statistician, or researcher in academia, there are numerous ways to partner with federal agencies to move forward with learning agendas, data management, data analysis, advanced techniques to safeguard data and prevent re-identification, new linked data sets, and new machine learning and statistical modeling methods to provide powerful insights into better informed policymaking and more efficient operations. You can also come to work for a statistical agency. The work is challenging, the mission is important, and your colleagues will share your desire to make a major contribution to the field. Look for job openings on [USAJOBS](#), through networking, or by taking a temporary assignment through an arrangement with your organization or through programs such as the [Presidential Innovation Fellows](#).

If you are with a state, tribal, or local government, there are multiple ways to improve partnerships with your federal program counterparts or suggest mutually beneficial joint projects using statistical data. Many states need data from beyond their own state to understand fully the dynamics of activities in areas such as workforce, education, business dynamics, and economic growth. Federal partners can help.

5. Priority Research Areas

Federal statistical agencies have long worked with data from multiple sources, such as surveys, administrative program records, and commercial data (e.g., [Longitudinal Employer-Household Data](#); [Gross Domestic Product or GDP](#); [Consumer Price Index or CPI](#)). However, with advances in machine learning, computing power, increased access to data, and new data sources, several strategic, cross-cutting research priorities have emerged. One of the most critical priorities is around data quality. That is, when data are linked from multiple sources that may have varying degrees of quality, describing the quality to users becomes more challenging than when describing data from a single source. Work is underway to develop standards and methods for measuring and describing data quality for users, including for what uses these new combined data sets are best fit. Transparency is a key element, particularly if machine learning is involved. Also under discussion are ways to expand the standards for what may be called official government statistics, while distinguishing these from the data and information products produced by evaluations and other related statistical activities and studies. In addition, modernizing methods to reduce the risk of re-identification is a high priority imperative. As more open data becomes available and computing power increases, improved approaches to disclosure avoidance and privacy protection are needed. While many promising approaches are on the horizon, much more research is needed in this area to continue to safeguard data entrusted to the government.

In addition to the websites for data.gov, the Federal Data Strategy, and the Federal Statistical Research Data Centers, other places you can check for updates as these activities move forward include the [Federal Register](#); Regulations.gov, where public comments are gathered and posted when solicited through a [Federal Register Notice](#); by looking for agency and stakeholder updates on social media and in newsletters from data stakeholder organizations; and, of course, in this space.

Disclosure Statement

Nancy A. Potok is the Chief Statistician of the United States at the U.S. Office of Management and Budget. The views expressed in this article are those of the author herself and do not necessarily represent the views of the United States Office of Management and Budget or the United States government.

©2019 Nancy A. Potok. This article is licensed under a Creative Commons Attribution (CC BY 4.0) [International license](#), except where otherwise indicated with respect to particular material included in the article.

Footnotes

1. Evidence Act; [Pub. L. No. 115–435](#),
<https://www.law.cornell.edu/uscode/text/44/3563> ↵
2. These and additional information on use cases, proof points, and related activities are available at
<https://strategy.data.gov>. ↵