Modernizing the U.S. Federal Data System: Insights from Nancy Potok, Chief Statistician, Office of Management and Budget

By Michael J. Keegan

An effective and efficient U.S. federal government requires evidence about where needs are greatest, what works and what does not work, where and how programs could be improved, and how programs of yesterday may no longer be suited for today. Having access to timely, accurate, reliable statistical data enables the federal government to make reasoned and disciplined decisions about where to target resources to get the largest possible return for the American taxpayer. The federal government's statistical agencies and programs play a vital role in generating that data. Timely, accurate, and relevant statistical data are the foundation of evidence-based decision making.

How is the U.S. federal government leveraging data as a strategic asset? How is the federal government building the infrastructure for evidence-based policymaking? And what does the future hold for the federal data and statistical communities? Nancy Potok, chief statistician within the Office of Management and Budget (OMB), joined me on *The Business of Government Hour* to share her insights on these topics and more. The following is an edited excerpt of our discussion, complemented with additional research.

Would you tell us more about the work of the Statistical and Science Policy Office and your duties and responsibilities?

When I tell people that I'm the chief statistician of the United States, there is this pause. People say, that's the coolest title I ever heard in government. There is another pause with an immediate, "What do you do?" I don't produce statistics. It's a policy job. It was established as part of the Paperwork Reduction Act and put into the Office of Information and Regulatory Affairs (OIRA).

My job is threefold. First and foremost, my job is to safeguard the integrity of federal data. I am charged with making sure that federal statistics are objective, unbiased, not politically influenced, accurate, timely, and relevant. All statistical directives that outline standards and rules on handling federal statistical data come from my office. My office puts out methods and standards that federal agencies have to follow if they're going to assert that their statistical data is official U.S. government data.

Given we have a decentralized statistical system in the U.S., I also coordinate all federal statistical agencies. I'll give you a sense of the size and scope: 13 principal federal statistical agencies and three recognized statistical units—agencies whose principal mission is to produce official federal statistics—are joined by over 100 other federal programs in statistical activities, spanning measurement, information collection, statistical products, data management, and dissemination. I head an Interagency Council on Statistical Policy (ICSP), which promotes integration across the Federal Statistical System.

My third role is to represent the U.S. internationally. I lead the U.S. delegation to the UN Statistical Commission. I also represent the U.S. at the Organization for Economic Co-operation and Development (OECD) on statistical matters. These are very important partnerships. We collaborate with the international statistical community and have good working relationship with my international counterparts.

"First and foremost, my job is to safeguard the integrity of federal data. I am charged with making sure that federal statistics are objective, unbiased, not politically influenced, accurate, timely, and relevant."

—Nancy Potok

Given your role and responsibility as chief statistician of the U.S., what would you say our top challenges are and how have you sought to address those challenges?

A top challenge is the pace of change around information and the ways in which information can be made available. Things have changed considerably in the last 10-15 years. My challenge is to make sure that the federal statistical system stays relevant in an ever-changing environment. A common perception of a statistician is someone with a green eyeshade who is calculating variance and standard deviations. This is not an accurate portrayal; statistical activity really defines information that is used to describe groups, even though it comes from individuals. It's business data or social data.

If you want high-quality information, it's very important that you think about the mature system of quality measurement that the statistical community has developed over decades. In some quarters, I think there is a view that if you have enough data, you're going to get to the right answer eventually. But a lot of people use big data sets that aren't complete, that have biases built into them. The statistical system has a very mature framework that is important to use. How do we take these traditional statistical methods that rely primarily on surveys and modernize them for using other types of data? It's a big challenge.

As the world produces better, faster, more granular statistical products, what do you worry about?

Re-identification. Protecting confidentiality is a big, big challenge these days because technology, computing power, and the availability of open data really create a different environment than we had 30 years ago. The intake side is a big challenge in terms of new data sources and the rapidity at which you can create products to meet increasing demands for timely and granular data.

What are your strategic priorities? How have external trends informed and shaped your strategic direction?

One of my key priorities is to modernize the data collection methods in order to be able to get data out faster. Surveys take a long time to process and they're expensive. Also, people increasingly don't like to answer surveys. It's an intrusion. It's hard to collect information that way. We also have a proliferation of data accessible in less traditional ways that can be used for statistical purposes. For example, if you are releasing a monthly retail sales economic indicator and you want to put it out faster than, say, six weeks after you complete each monthly survey asking businesses about their retail sales, you can start to look at data from companies that aggregate credit card records, because more and more purchases are on credit cards. The Census Bureau and the Bureau of Economic Analysis have done research in using aggregated credit card records to calculated retail sales. The individual purchases are de-identified because they're aggregated, but you can see what was purchased using credit cards in Chicago or in New York the day after the purchases took place. That's how fast the data are aggregated. You no longer have to go to the businesses to ask about sales, because you can see the sales from the purchase end. But you need to be careful that you are not missing sales that are paid for by means other than credit cards in the released indicator.

What big initiatives are you working on?

Data sharing between agencies, and safeguarding that data, is one. There are large amounts of information that the government has already collected on people. It resides not just in Social Security records, but also Medicare and Medicaid, veterans' records, housing records. Why would you spend all this money to go out in a survey and recollect the information if the information that you wanted has already been collected? Doing more data sharing between agencies is a big focus and a key requirement of the Foundations for Evidence-Based Policymaking Act (Evidence Act). Another priority is to safeguard it at the other end and make sure that you're protecting confidentiality and privacy. That also includes assuring that these sensitive data are only used for statistical purposes and not to identify or take action against any individuals. These are two big strategic initiatives that we're working on cross-agency, along with the federal data strategy.

Would you tell us more about the federal data strategy?

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 is enterprise data governance. It includes 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. 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 requirements of the Evidence Act are geared towards a fundamental change in the way agencies think about what they're doing. The law enacts 11 of the 27 recommendations of the Commission on Evidence-Based Policymaking. I was one of the Commissioners. I know these recommendations. I'm very excited about this law."

— *Nancy Potok*

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 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 for statistical activities that support evidence building. 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 statistical data sets, which could include commercially available data and data from state and local governments.

The third component—decision making and accountability. It 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. 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 they are also encouraged to 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.

Would tell us more about the implementations of the Foundations for Evidence-Based Policymaking Act (the Evidence Act)?

The federal data strategy and the Evidence Act are a powerful match. Their collective vision: to create partnerships between U.S. federal agencies, State, tribal, and local governments, academia, and industry to realize effectively the value of shared federal data—accomplished by putting 'open' non-sensitive 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.

The requirements of the Evidence Act are geared toward a fundamental change in the way agencies think about what they're doing. The law enacts 11 of the 27 recommendations of the Commission on Evidence-Based Policymaking. I was one of the commissioners. I know these recommendations. I'm very excited about this law. The purpose is to address the fractured federal statistical landscape. It illustrates a shift in thinking. This is not about starting with the metrics and then creating your metrics based on data you have. It is about focusing on what you want to know such as: are your programs changing anything in society as they were intended to do? What about operational efficiency?

Can you elaborate on this fundamental shift in thinking?

Sure. 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.

The Act also requires 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.

How are these learning agencies created and carried out?

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. In September, OMB held an orientation for people in these new roles to help them understand their responsibilities. We found the orientations to be very interesting because of the wide range of backgrounds, skills, and agency mission objectives across government, which highlighted the challenges we are facing. The good news is that although this is a heavy lift for many agencies, and a major culture change, there is a lot of enthusiasm, commitment, and pockets of deep knowledge that people are eager to share.

We have also established an OMB Data Policy Council to further this effort. We have been working towards meeting some statutory deadlines we have in the Evidence Act for releasing Notices of Proposed Rule Making (NPRM) for public comment. The first one we hope to release defines the roles and responsibilities of statistical agencies, emphasizing their need to be able to produce objective, unbiased, relevant, timely high-quality data without political interference. The second sets out criteria on how units within agencies, such as evaluation offices, can qualify to get the official designation of a statistical unit. This designation comes with many advantages, such as access to confidential data for statistical purposes.

These changes are quite significant, and the Act envisions that they will occur with some standardization across the federal government. We have a first-year implementation plan, which should be out by the end of this year, and has very specific things that we want agencies to do.

What does the future hold in the federal data and statistical communities?

There are tremendous opportunities for agencies to take advantage of now, with an eye on the future. There's the rise

of artificial intelligence and machine learning. Agencies are drowning in old unstructured data some of it handwritten in PDF form. Technology like machine learning can help. You can do things like text analysis with machine learning. Advances in technology takes us away from more manual processes to more automated processes. It seems to me a natural next step for the statistical and data communities is to start relying on such technologies as artificial intelligence and the machine learning. We just need to ensure we address issues of transparency and bias in order to assure that we continue to provide high quality data that can be relied upon to guide important decisions.

To learn more about the Office of Management and Budget, go to omb.gov.

To hear the interviews from *The Business of Government Hour*, go to businessofgovernment.org/interviews.

To download the show as a podcast, go to PodcastOne or iTunes and search for *The Business of Government Hour.*

To view excerpts of the show, go to youtube.com/businessofgovernment.