July 27, 2018

Karen Dunn Kelley
Undersecretary for Economic Affairs
U.S. Department of Commerce
1401 Constitution Avenue NW
Washington, DC 20230

Ms. Kelley,

On behalf of the Center for Data Innovation (datainnovation.org), we are pleased to submit comments in response to the administration's request for feedback on the development of a Federal Data Strategy.[1]

The Center for Data Innovation is the leading think tank studying the intersection of data, technology, and public policy. With staff in Washington, D.C., and Brussels, the Center formulates and promotes pragmatic public policies designed to maximize the benefits of data-driven innovation in the public and private sectors. It educates policymakers and the public about the opportunities and challenges associated with data, as well as technology trends such as predictive analytics, open data, cloud computing, and the Internet of Things. The Center is a non-profit, non-partisan research institute affiliated with the Information Technology and Innovation Foundation.

The goal of the Federal Data Strategy is to develop practices that enable the federal government to leverage data as a strategic asset, which is one of the goals of the President's Management Agenda (PMA), launched in March 2018.[2] We support the goal of the Federal Data Strategy, and we welcome the proposed steps to incorporate public input on the strategy through an iterative process. To that end, we offer several steps the government should take, both as part of and in conjunction with the Federal Data Strategy, to improve how the government manages data, uses it to improve decision making, and makes it available for public use.

[1] "Request for Comments on the Cross-Agency Priority Goal: Leveraging Data as a Strategic Asset," Federal Register, June 27, 2018, https://www.federalregister.gov/documents/2018/06/27/2018-13768/request-for-comments-on-the-cross-agency-priority-goal-leveraging-data-as-a-strategic-asset.
[2] Ibid.

IMPROVING DATA ACCESSIBILITY

First and foremost, the administration should publicly call for the passage of the Open, Permanent, Electronic, and Necessary (OPEN) Government Data Act. The OPEN Government Data Act would require federal agencies to publish government data in machine-readable and open formats and use open licenses. It would also direct agencies to support innovative uses of government data, adopt consistent data practices across government, and develop best practices for open data. Additionally, a firm commitment to providing open data as a public resource would encourage businesses, non-profits, and others to invest in innovative tools that make use of government data. Key parts of the Federal Data Strategy emphasize the need to improve enterprise data governance, increase data access and use, encourage the use of data to improve decision-making and accountability, and make government data available for commercial and public use to spur innovation. Passing the OPEN Government Data Act is likely the single most effective step the administration could take to help achieve these goals, because it would create a legislative mandate for agencies to better manage government data throughout its lifecycle.

There are also several improvements the administration could make to the federal government's open data portal, data.gov, to make federal open data more useful. First, the Office of Management and Budget (OMB) should ensure that all datasets on data.gov have appropriate metadata. In the past, datasets on data.gov have failed to list licensing information indicating that they are freely available under an open license. The absence of complete and accurate metadata can create confusion about whether certain datasets are free to use for commercial and non-commercial purposes, thereby limiting its use. Second, federal agencies submitting datasets to data.gov should clearly designate the person responsible for maintaining a particular dataset and link these identities with government personnel data. Then, when the individuals responsible for maintaining datasets leave their jobs, agencies can update the relevant metadata on these datasets. Should a dataset be faulty or out of date, it can be challenging for users to identify the correct agency contact if this information is not available and up-to-date.

ENCOURAGING INNOVATIVE PUBLIC USE OF GOVERNMENT DATA

There are two important steps the federal government should take to expand innovative public use of government data.

First, the government should consider improving how it receives public input about prioritizing various investments in open data. The federal government has limited resources, and it will always need to make choices about where to allocate these resources to improve data quality, including accuracy, completeness, timeliness, and availability. Agencies mostly make these decisions based on internal deliberations. These deliberations likely prioritize their own needs and goals, which may

not align with other valuable public interests outside each agency's own mission. Similarly, agencies may be unaware of the needs of certain non-government data users since they have limited opportunities to meet with the businesses, researchers, civil society organizations, and members of the public that rely on their data for a wide range of applications.

As a result, as agencies devote resources to improving their open data efforts, they risk wasting resources and limiting the value of government data unless they can proactively identify the needs of open data users and prioritize their efforts accordingly. OMB should therefore consider establishing an Open Data Review Board comprised of both government and non-government stakeholders that could regularly solicit proposals from open data users about specific ways to improve specific government datasets. This board could evaluate these proposals, similar to how the National Science Foundation reviews research grant applications, and award internal government funding to these proposals based on their merits.

Second, the government should consider how to facilitate sharing non-public data, including non-public government data with the private sector and non-public private-sector data with the government. One way to do this is by establishing data trusts—frameworks for securely and efficiently sharing non-public data between federal agencies, the private sector, and academia with a clear set of rights and responsibilities. The United Kingdom is developing a program of data trusts to facilitate the sharing of data that would not otherwise be made publicly available due to its proprietary or sensitive nature but that has high value.[3] OMB may be able to model such an initiative after the United Kingdom's approach, which includes creating a Data Trusts Support Organization to develop tools and guidance for data trusts to make it easy for stakeholders to share data without sacrificing privacy or security.[4]

THE FEDERAL DATA STRATEGY SHOULD SUPPORT FEDERAL IOT AND AI STRATEGIES

The Federal Data Strategy is an important initiative to improve how the federal government manages, uses, and shares its data. OMB should recognize that how successfully the federal government maximizes the value of government data will be directly related to how well it leverages technologies such as the Internet of Things (IoT) and artificial intelligence (AI). The Internet of Things offers an unprecedented opportunity for the federal government to generate valuable datasets about

[3] Dame Wendy Hall and Jérôme Pesenti, "Growing the Artificial Intelligence Industry in the UK," UK Department of Business, Energy, and Industrial Strategy, October 15, 2017, https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/652097/Growing_the_artificial_intelligence_industry_in_the_UK.pdf.
[4] Ibid.

government operations, such as building operations, fleet management, and environmental data, which may have substantial value both within and outside of government. The Federal Data Strategy should consider how to leverage these new data sources, and how to handle challenges related to IoT data, such as processing real-time data. Ultimately, the Federal Data Strategy should support a Federal IoT Strategy that considers how to best leverage the Internet of Things by government.

While IoT offers new opportunities to gather valuable data, AI offers opportunities to use government data in new and innovative ways, particularly by automating and improve routine government functions such as customer support and fraud detection.[5] However, successfully automating government processes requires strong data management. There are also many applications of AI, both within government and in the private sector, that will depend on the availability of large government datasets for training purposes. The Federal Data Strategy should consider how to ensure it is considering future data needs to support AI. In addition, just as the current PMA lists IT modernization as a key driver of transformation, the next PMA should have an overarching goal of "Using AI to Automate Government."

Sincerely,

Daniel Castro
Director
Center for Data Innovation
dcastro@datainnovation.org

Joshua New
Senior Policy Analyst
Center for Data Innovation
jnew@datainnovation.org

[5] Grant Gross, "Government Ventures into AI," *GCN*, September 5, 2017, https://gcn.com/articles/2017/09/05/ai-government.aspx.