July 5, 2019

William Hawk
US Census Bureau
4600 Silver Hill Road
Washington, DC 20233

Re: Leveraging Data as a Strategic Asset Phase 3 Comments
[Docket Number USBC-2019-0001]

BSA | The Software Alliance (BSA) appreciates the opportunity to provide feedback in response to the Administration's Request for Comments on the Draft 2019-2020 Federal Data Strategy Action Plan.[1] BSA is the leading advocate for the global software industry.[2] Our members are at the forefront of software-enabled innovation that is fueling global economic growth by helping enterprises in every sector of the economy operate more efficiently. As global leaders in the development of data-driven products and services, BSA members have unique insights into how the Administration can leverage data as a strategic asset to "grow the economy, increase the effectiveness of the Federal Government, facilitate oversight, and promote transparency."[3]

Data is the lifeblood of the modern digital economy – powering innovation and growth across the globe and enabling organizations to create new jobs, boost efficiency, drive quality, and improve output. Unfortunately, the "government lags behind the private sector

[1] 84 Fed. Reg. 25370 (June 4, 2019) [Hereinafter "RFC"].

[2] BSA's members include: Adobe, Akamai, Apple, Autodesk, Bentley Systems, Box, Cadence, CNC/Mastercam, DataStax, DocuSign, IBM, Informatica, Intel, MathWorks, Microsoft, Okta, Oracle, PTC, Salesforce, ServiceNow, Siemens PLM Software, Sitecore, Slack, Splunk, Symantec, Trend Micro, Trimble Solutions Corporation, Twilio, and Workday.

[3] RFC, *supra* n. 1, at 25370.

in its standards for managing and documenting data,"[4] and has therefore been unable to fully leverage existing data resources to improve the delivery of government services or spur innovation by making them available to the public. We are therefore encouraged by the Administration's aim to improve the federal government's data posture.

The Federal Data Strategy outlines a sound vision for coordinating the government's approach to data use and management. We appreciate the thoughtful process by which the Administration has continued to seek stakeholder feedback. The Federal Data Strategy's core Principles and Practices provide a solid foundation for modernizing the Federal Government's approach to data stewardship. Below we offer a series of recommendations that will help align the Draft Action Plan with these Principles and Practices and complement other Federal data initiatives, including implementation of the Foundations for Evidence-based Policymaking Act and the Administration's Executive Order on Maintaining American Leadership in Artificial Intelligence. Our recommendations are organized around four key themes: (1) maximizing opportunities for stakeholder collaboration; (2) leveraging the full range of government data assets; (3) promoting a robust data sharing ecosystem and enhancing privacy protections; and (4) ensuring that agencies are resourced to leverage data as a strategic asset.

1. Maximizing Opportunities for Stakeholder Collaboration

A central tenet of the Federal Data Strategy is that the value of government data can only be maximized through sustained public engagement. To that end, the Strategy seeks to "Promote Transparency" to engender public trust (Principle 3) and "Demonstrate Responsiveness" to ongoing stakeholder input (Principle 7).[5] The Data Strategy Practices likewise prioritize support for "non-federal stakeholders," including through engagement with "industry, academic, and other non-federal users of data to share expert knowledge of data assets, promote wider use, improve usability and quality, and advance innovation and commercialization" (Practice 40).[6] To align the Action Plan with these Principles and Practices, we recommend several clarifications to the Action Plan. As noted below, these recommendations are also intended to complement other Federal data initiatives, including

[4] Comm'n on Evidence-Based Policymaking, The Promise of Evidence-Based Policymaking 5 (2017) at pg. 78 [hereinafter CEP Report].

[5] https://strategy.data.gov/principles/

[6] https://strategy.data.gov/practices/

implementation of the Foundations for Evidence-based Policymaking Act and the Administration's Executive Order on Maintaining American Leadership in Artificial Intelligence.

- ***Action 1 – OMB Data Council:*** Recognizing that many departments lack senior leadership focused on managing and expanding access to data, the OPEN Government Data Act (Title II of the Foundation for Evidence-based Policymaking Act) requires government agencies to appoint a Chief Data Officer (CDO) to oversee data management and coordination.[7] Because effective management of Federal data requires collaboration, the OPEN Government Data Act also requires the establishment of a Chief Data Officer Council for CDOs to share best practices, promote data sharing, and identify ways to improve access to data. Importantly, the Chief Data Officer Council must also "consult with the public and engage with private users of Government data and other stakeholders on how to improve access to data assets of the Federal Government."[8]

  We seek clarification about whether the OMB Data Council is intended to serve as the mechanism for implementing the Chief Data Officer Council requirement in the OPEN Government Data Act. If so, it is important to clarify within Action 1 that all CDOs will be permitted to participate in the Council. Likewise, we urge the Federal Data Strategy team to include within Action 1 an explicit mechanism for industry to collaborate with the Council on the development of data governance policies, practices, and projects. To that end the Council should convene public meetings and solicit stakeholder feedback on a quarterly basis.
- ***Action 2 – Data Science Training and Credentialing Catalog:*** Action 2 envisions the creation of a curated catalog of federal and non-federal trainings in data science. Given the highly dynamic nature of the service and credential offerings in this space, a one-time data cataloging effort is likely to become quickly outdated. It is therefore critical that the General Services Administration establish a process for receiving input from federal and non-federal stakeholders (including industry, academic, and other non-federal users of data) on a continuing basis. GSA should likewise consider

[7] HR 4174, Foundations for Evidence-Based Policymaking Act of 2018.

[8] HR 4174 – Section 202, Open Government Data.

organizing the catalog along both technical and business tracks to facilitate role-based coordination within an agency.

- ***Action 3 – Data Ethics Framework:*** The development of a Federal Data Ethics Framework is an important undertaking that will help engender public trust and confidence in the government's approach to data collection, management, and use. Currently, Action 3 suggests that the Framework will be developed in collaboration with "academia, professional associations, and federal data stakeholders." To ensure the process is inclusive and informed by a full range of expertise, we recommend *all* "federal and non-federal data stakeholders" be included in the development of the Framework. Given the National Institute of Science and Technology's expertise in data governance practices, we suggest they co-lead this effort in partnership with the GSA. Finally, consistent with Practice 20 (Leverage Data Standards) and OMB Circular A-119, the Data Ethics Framework should leverage existing voluntary standards and best practices in this area with an eye towards ensuring consistency.
- ***Action 5 – Repository of Data Strategy Resources and Tools:*** Action 5 does not currently reference a mechanism for soliciting private sector feedback on relevant tools and resources for implementing the Federal Data Strategy. We encourage the Federal Data Strategy development team to revise Action 5 to ensure that such feedback will be welcomed. Like the Training and Credentialing Catalog, Action 5 should also be revised to clarify that GSA will take an iterative approach to the development of the Repository so that new technologies, tools, and best practices can be added beyond the initial November 2019 deadline.

2. Leveraging the Full Range of Government Data Assets

As the Federal Data Strategy is implemented, agency CDOs should be encouraged to take the broadest possible view about the types of data that may be relevant to the individual Actions. Because of the central role that Data.gov plays in facilitating discovery and access to federal data, it is particularly important that agencies adopt an expansive view of "data" as they create their Data.gov inventories. In passing the Foundations for Evidence-Based Policymaking Act, the Committee on Oversight and Government Reform noted that existing inventories "are of uneven quality and completeness, and [do not] include all datasets useful for evidence building."[9] In an effort to improve the quality of agency data inventories, the Foundations for Evidence-Based Policymaking Act requires them to include the

[9] H. Rept. 115-411, Foundations for Evidence-Based Policymaking Act of 2017.

"maximum amount of metadata about the maximum number of data assets," including "data from applications, devices, networks, and equipment, which is often underutilized and can be useful for strengthening cybersecurity and forensics, improving security, reducing costs, identifying waste, reducing energy consumption, and improving agency operations."[10] We therefore recommend the following:

- ***Action 7 – Automated Inventory Tool for Data.gov:*** Action 7 should clarify that agency inventories must be as inclusive as possible, capturing both structured and unstructured data, including data generated by devices and sensors. The inventory tool should be designed to enable real-time updates to dynamic data sets on Data.gov. To that end, the tool should provide real-time visibility to all data sets and facilitate automated processing, preparation, and posting of data to Data.gov.

3. Promoting a Robust Data Sharing Ecosystem and Enhancing Privacy Protections

The Federal Data Strategy will be most effective if it fosters a robust data sharing ecosystem between federal and non-federal stakeholders. To that end, the Strategy development team should consider mechanisms for identifying and mitigating sources of friction that currently inhibit the sharing of government data. Such an effort should include an evaluation of how agencies can leverage emerging technologies and data governance processes to enhance privacy protections, while also making more data available to the public. Consistent with the Executive Order on Maintaining American Leadership in Artificial Intelligence, [11] the Federal Data Strategy Action Plan should direct agencies to consider opportunities to utilize "tiered access" approaches to data protection,[12] differential privacy

[10] *Id.*

[11] Section 5 of the Executive Order calls for agencies to review their data to "identify opportunities to increase access and use by the greater non-Federal AI research community…while protecting safety, security, privacy and confidentiality." To that end, Section 5(a)(iii) directs agencies to "consider methods" for improving "access" to priority data.

[12] A tiered access approach to data governance would enable agencies to make available public versions of otherwise sensitive datasets by stripping out personal information. *See* CEP Report, *supra* note 4, at 38 ("Tiered access is an application of data minimization, a key privacy safeguard for evidence building as embodied in the Fair Information Practice Principles (described in Chapter 3). Data minimization means giving access to the least amount of data needed to complete an approved project. For example, an eligible researcher's project might earn approval for access to confidential information at a highly secure research data center that requires expert review of all output. Another researcher's project may need only access to a data query tool that runs an analysis, checks for disclosure risk without ever showing individual records, and provides group statistics (see the box

frameworks,[13] homomorphic encryption,[14] and other cutting-edge technologies and data governance processes that can facilitate greater access to data while safeguarding user privacy. In addition to promoting the sharing of government data, the 2019 National Artificial Intelligence R&D Strategic Plan highlights the importance of promoting the voluntary sharing of data held by industry.[15] The Federal Data Strategy should therefore incorporate mechanisms for encouraging the voluntary sharing of private sector data in a manner that is respectful of privacy concerns and property interests, including by convening workshops to explore the range of available technical, legal, and contractual tools and best practices for promoting data sharing arrangements.

- ***Action 9 – Data Resources for AI R&D:*** In addition to undertaking internal reviews of their data, we recommend that Action 9 include a requirement for agencies to evaluate opportunities for using privacy-enhancing technologies and/or data governance processes to improve the availability of such data. To promote the voluntary sharing of industry data, the Commerce Department should convene a workshop to explore existing impediments and whether new incentives are necessary. Finally, consistent with the National Artificial Intelligence R&D Strategic Plan, the text of Action 9 should specifically direct agencies to make data resources

"Data Query Tools"). . . .A well-designed and properly implemented data minimization strategy like tiered access can reduce the risk of unauthorized use and unintended harm to individuals.").

[13] *See* Kobbi Nissim, Thomas Steinke, Alexandra Wood, Micah Altman, Aaron Bembenek, Mark Bun, Marco Gaboardi, David O'Brien, and Salil Vadhan, *Differential Privacy: A Primer for a Non-technical Audience* (February 2018), available at https://privacytools.seas.harvard.edu/files/privacytools/files/pedagogical-document-dp_new.pdf ("Differential privacy is a strong, mathematical definition of privacy in the context of statistical and machine learning analysis. It is used to enable the collection, analysis, and sharing of a broad range of statistical estimates, such as averages, contingency tables, and synthetic data, based on personal data while protecting the privacy of the individuals in the data. . . . Computer scientists have developed a robust theory for differential privacy over the last fifteen years, and major commercial and government implementations have now started to emerge.").

[14] Homomorphic encryption is a form of encryption that allows a computational analysis of encrypted data, ensuring that the data remains confidential. Use of homomorphic encryption could, for instance, enable the sharing of aggregated medical data to facilitate AI research without risking patient confidentiality. *See* Jean Louis Raisaro, Jeffrey Klann, Kavishwar Wagholikar, Hossein Estiri, Jean-Pierre Hubaux, and Shawn Murphy, *Feasibility of Homomorphic Encryption for Sharing I2B2 Aggregate-Level Data in the Cloud*, AMIA Jt Summits Translational Science (May 2018), available at https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5961814/#

[15] https://www.whitehouse.gov/wp-content/uploads/2019/06/National-AI-Research-and-Development-Strategic-Plan-2019-Update-June-2019.pdf

available for both "commercial and public interests."[16] To that end, agencies should seek to make data resources available to "non-Federal AI research and commercial communities."

4. Ensuring Agencies are Resourced to Leverage Data as a Strategic Asset

The ability for agencies to deliver on the promise of the Federal Data Strategy will ultimately depend on whether they have the resources needed to truly leverage data as a strategic asset. In this regard, the government's "unwieldy and out-of-date Federal IT infrastructure" will continue to be a major impediment.[17] Reliance on aging physical data centers can result in data siloes that prevent an agency from making use of data outside of the application the agency used to generate it, and make it exceedingly difficult to share with colleagues in other agencies.

Following through on the Administration's commitment to modernizing the government's aging IT infrastructure through the adoption of commercial cloud offerings will therefore be critical to the success of the Federal Data Strategy.[18] IT modernization is necessary to ensure agencies have the systems in place to leverage large scale data assets for AI and machine learning. Unfortunately, we have seen inconsistent progress due to funding issues and inefficiencies within the Federal Risk and Authorization Management Program (FedRAMP) accreditation program. We urge the Administration to address the following issues:

- **Modernizing Government Technology Act – Working Capital Funds:** As a strong supporter of the Modernizing Government Technology Act, we have been disappointed that few agencies have made use of the authorization to establish working capital funds for IT modernization. The legislation authorized important flexibility that would allow both programs and CIO offices to modernize systems over time. To date, only three agencies have established the authorized working capital funds that allow for the reinvestment of savings to fund IT modernization efforts.

---

[16] *Id.*

[17] American Technology Council, *Report to the President on Federal IT Modernization* (December 2017), available at https://itmodernization.cio.gov/assets/report/Report%20to%20the%20President%20on%20IT%20Modernization%20-%20Final.pdf.

[18] *See* Federal Cloud Smart Strategy, https://cloud.cio.gov/strategy/.

During a recent House Committee on Oversight and Reform hearing, witnesses testified that the lack of agency uptake is due to uncertainty about whether they have the authorization needed to both establish working capital funds and to transfer funds thereto.[19] Members of the Committee, however, explained that they intended to provide such authority and requested that OMB provide additional guidance to address agency questions and/or explain whether additional legislative authorization is necessary to effectuate the original intention of the Modernizing Government Technology Act. We urge the Administration to work closely with agency CIOs, OMB, and Congressional appropriators to determine how agencies can use the working capital funds to finance cloud adoption. The Office of the Federal CIO is well positioned to issue guidance on the establishment and use of these funds.

- **FedRAMP Accreditation:** The process for cloud providers to get accredited by FedRAMP remains too slow and costly, and ultimately fails to provide the presumption of adequacy that was originally intended. Industry appreciates the program's evolution over time, but FedRAMP's "authorize once, reuse many times" objective has not been realized. FedRAMP's value proposition is further undermined because agencies continue to impose bespoke requirements on their vendors and are unwilling to rely on an Authorization to Operate issued by other agencies or a provisional-ATO by FedRAMP. As a result, cloud service providers must get multiple certifications to provide services across the government. We urge the Administration to continue to focus on efforts to streamline the FedRAMP process and to encourage agencies to treat previously granted ATOs as presumptively adequate. Ensuring FedRAMP meets these objectives is critical to the Federal Data Strategy's goal of ensuring that the government is leveraging data as a strategic asset.

* * * *

Thank you again for the opportunity to share our views on these important issues.

Sincerely,

Christian Troncoso
Director, Policy

[19] *See* FITARA 8.0, Hearing Before the House Committee on Oversight and Reform (June 26, 2019).