August 9, 2019

Russell T. Vought
The Office of Management and Budget
725 17th Street NW
Washington, DC 20503

Re: Identifying Priority Access or Quality Improvements for Federal data and models for Artificial Intelligence Research and Development (R&D), and Testing

Dear Mr. Vought:

BSA | The Software Alliance (BSA) appreciates the opportunity to provide feedback in response to the Office of Management and Budget's (OMB) Request for Information on Identifying Priority Access or Quality Improvements for Federal data and models for Artificial Intelligence Research and Development (R&D), and Testing.[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 greater access and quality improvements to Federal government data assets could help propel the development of Artificial Intelligence.

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 in its standards for managing and documenting data,"[3] and has therefore been unable to fully maximize the value of government data by consistently making it available to the public in a manner that is conducive to AI development. We are therefore encouraged by the Administration's aim to improve the federal government's data posture. Below we offer three recommendations that will help ensure that federal data assets are collected, maintained, and made available in a manner that will advance the development of AI.

---

[1] 84 Fed. Reg. 32962 (July 10, 2019) [Hereinafter "RFI"]
https://www.govinfo.gov/content/pkg/FR-2019-07-10/pdf/2019-14618.pdf.

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

Foster an Open Data Culture that Extends to Federally Funded Research

We appreciate OMB's effort to identify specific federal data assets that – if made available to the public – could help advance the development of AI. In addition to identifying individual data assets, OMB can help stimulate long-term AI R&D by fostering an open data culture within the federal agencies. To that end, we are encouraged by OMB's recent release of agency guidance for the implementation of the Foundations of Evidence-based Policymaking Act (FEBP).[4] Title II of the FEBP – the OPEN Government Data Act – requires government agencies to appoint a Chief Data Officer (CDO) to oversee data management and coordination.[5] Consistent with the goals of the Executive Order on Maintaining American Leadership in AI (and of this RFI), agency CDOs will play a critical role in maximizing the value of government data by ensuring that it is collected, maintained, and made available in ways that are most useful for AI R&D. We therefore appreciate that OMB's implementation guidance reinforces the importance of agency CDOs by empowering them with the authority to oversee the "data governance and lifecycle data management" practices within their agencies.

The implementation guide indicates that OMB will soon prepare Phase 2 guidance on the development of agency "Open Data Plans." Because these Open Data Plans will inform the dissemination improvement priorities for agencies, OMB's Phase 2 guidance should set ambitious and measurable targets that reflect the Administration's commitment to fostering a culture of open government data. 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. The Phase 2 guidance should make clear that OMB will evaluate agencies' progress in improving the quality of their data inventories and the volume of high-value data made available via the federal data catalog (Data.gov). OMB should clarify that agency inventories must be as inclusive as possible, capturing both structured and unstructured data, including data generated by devices and sensors. OMB should likewise direct agencies to focus on providing real-time updates to dynamic data sets on Data.gov and improving the availability of data provenance information.

The Administration's commitment to fostering a culture of open data should also be extended to include technical data generated as part of federally funded research. To that end, OMB should partner with the National Science Foundation, Department of Energy, the National Institute of Health, and other research funding agencies, to explore options for making more research data available to the public. To the greatest extent permissible, and without prejudice to any applicable intellectual property protections, data generated during the course of federally funded research should be made available to the public so that it can be leveraged for additional research and development. In partnership with the research agencies, OMB should ensure that federally-funded agreements require award recipients to make available to the funding agency any raw data assets (i.e., those that are unencumbered by IP protections) generated during the course of the funded research project. Such data assets should, in turn, be made available for use by the broader AI community.

---

[4] Phase 1 Implementation of the Foundations for Evidence-Based Policymaking Act of 2018: Learning Agendas, Personnel, and Planning Guidance, https://www.whitehouse.gov/wp-content/uploads/2019/07/M-19-23.pdf.

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

Develop a Consistent Approach to Data Licensing

One major source of friction currently impeding the private sector's use of government data are the inconsistent licensing practices that agencies use to make their data available to the public. Agencies currently make their data available under a variety of terms, conditions, and use restrictions that can undermine efforts to make robust use of government data. The particular practices tend to vary wildly by agency. For instance, data made available by the Social Security Administration via Data.gov tends to be distributed via a Creative Commons Zero License, while data made available by US AID tends to be distributed via a Creative Commons "Attribution-No Derivatives" license. In some cases, government data is made available pursuant to license agreements that explicitly forbid commercial uses. For instance, the Smithsonian Institute recently made available 3D printable models of the space suit worn by Neil Armstrong during the Apollo 11 mission.[6] When users attempt to download the models, they are confronted with a popup license that stipulates that the files "may be downloaded…only for non-commercial, educational, and personal uses."

These heterogeneous licensing practices can make it difficult to use government data. The problems are compounded when a project requires use of data from multiple agencies that may be subject to a thicket of varying terms and conditions. To aid the development of AI that relies on government data, OMB should explore options for adopting standardized data sharing agreements for use across government agencies. The value of a standardized government data sharing agreement will be enhanced if it is consistent with emerging best practices and industry standards. OMB should therefore consider the adoption of a data license model already in use by the AI developer community, such as the Linux Foundation's Community Data License Agreement[7] and/or the Open Use Data Agreement.[8] Regardless of the model, OMB should ensure that government data assets are never subject to "no commercial use" restrictions.

Ensure Agencies are Resourced to Collect, Maintain, and Make Available High-Quality Data

The ability for agencies to deliver on the open data vision of the Executive Order on Maintaining American Leadership in AI 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.[9] 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 the public.

---

[6] See https://3d.si.edu/armstrong.

[7] The Linux Foundation Projects, Community Data License Agreement, https://cdla.io/.

[8] Microsoft, The Open Use of Data Agreement, https://github.com/microsoft/Open-Use-of-Data-Agreement.

[9] 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.

August 9, 2019
Page 4

Following through on the Administration's commitment to modernizing the government's aging IT infrastructure through the adoption of commercial cloud offerings[10] will therefore be critical to the success of the Executive Order on Maintaining American Leadership on AI. 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. 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.[11] 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.

---

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

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