Leveraging Data as a Strategic Asset Phase 3 Comments
[Docket Number USBC-2019-00001]
July 5, 2019

On behalf of the Preservation of Electronic Government Information (PEGI) Project (www.pegiproject.org), we are grateful for the opportunity to comment on the Draft 2019-2020 Federal Data Strategy Action Plan. The PEGI Project, whose focus is centered around the preservation of government information, is a collaborative effort of library professionals with expertise that includes public data, Federal information policy, public access to Federal information, data curation, and academic research and teaching.

Preservation is a critical part of the lifecycle of data management. We are proposing approaches that will maximize resource use by assuring that the implementation of the Federal Data Strategy will include preservation as a key component.

**Response to Question 1: Identify additional actions needed to implement the Federal Data Strategy that are not included in the draft Action Plan and explain why.**

We applaud the inclusion of Practice 24, “Enhance Data Preservation” in the Strategy. However, we note that none of the Actions identified in the 2019-2020 Action Plan directly address this Practice. Laying cooperative groundwork to minimize the unintentional loss of data assets is a time-sensitive effort, best conducted in cooperation with Federal information policy stakeholders. These stakeholders, including the National Archives & Records Administration, the Library of Congress, and the Government Publishing Office, have expertise on all aspects of data lifecycle management and are closely associated with opportunities for alignment with existing and emerging best practices for digital preservation.

To leverage and extend existing communities, we propose an additional action that initiates and expands cooperative preservation efforts with the existing Federal Agencies Digital Guidelines Initiative (FADGI) [<http://www.digitizationguidelines.gov/>]. FADGI has facilitated over a decade of successful coordination among Federal agencies with administrative support from the Library of Congress. Existing FADGI working groups comprise participants from nearly twenty agencies and offices, including the Government Accountability Office, Department of Justice, and NASA.

In cooperation with Federal information policy stakeholders, including the National Archives & Records Administration (NARA), the Library of Congress, and the Government Publishing Office, a newly-formed FADGI Data Working Group can develop a list of goals pertaining to preservation and long-term access for Federal data assets, and conduct an initial assessment and gap analysis for Federal data preservation and reuse. The FADGI Data Working Group can also develop a template for Federal data management plans; following review and adoption by
the OMB Data Council, agencies would then be able to create and disseminate these plans as a structured public information tool. A Federal data management plan, similar to those required of Federal research grant recipients, would alert stakeholders to how the agency intends to steward its data assets, and provide opportunities for non-governmental stakeholders to engage in preservation activities.


Response to Question 2: Identify additional actions that would align with or complement ongoing Federal data initiatives or the implementation of new legislation, such as the Foundations for Evidence-based Policy Making Act of 2018 and explain why.

The Foundations for Evidence-based Policy Making Act of 2018 requires agencies to develop and maintain a comprehensive inventory accounting for data assets. For the inventory to be verifiably comprehensive, and for the data assets to be effectively utilized by agency stakeholders and the public, appropriate and sufficient metadata is essential. Metadata creation requires resource investment; however, the gains in efficiency by creating accurate and useful metadata carry throughout the data lifecycle. Internal agency requests are filled more quickly and accurately, the public can more easily find and better utilize agency data, and the data assets can be managed more effectively over time.

Response to Question 4: For each action, provide any edits and additional detail to ensure that they accurately and effectively describe needed activities, responsible entities, metrics for assessing progress, and timelines for completion.

Re: Action 1: Create an OMB Data Council

- To ensure continued alignment with Practice 24, “Enhance Data Preservation,” the inaugural OMB Data Council should include representation from the National Archives & Records Administration.
- The OMB Data Council should also include representation from the U.S. Government Publishing Office and the Library of Congress, in their capacities as active stakeholders in the Federal information policy arena.
- The initial charter of the OMB Data Council should account for clear pathways to receive meaningful and actionable input from academic and non-profit stakeholder communities.

Re: Action 2: Develop a Curated Data Science Training and Credentialing Catalog
While the need to embed data management expertise throughout Federal agencies is rightly noted as a motivator for this Action, the Action itself only enumerates data science skill development, not data management or data curation skill development. Data curation and management require a separate and distinct set of applied professional skills pertaining to planning and implementing appropriate steps for data description, access, and storage that maximize the usefulness of the data over time.

We encourage the inclusion of data management and data curation training and credentialing in this catalog, with appropriate incentives for the continued development of associated expertise throughout the statistical offices.

Re: Action 5: Develop a Repository of Federal Data Strategy Resources and Tools
- Priorities should include data documentation and metadata; standards for discovery and interoperability; data curation; and resources for identifying appropriate preservation measures throughout the data lifecycle.

Re: Action 8: Pilot Standard Data Catalogs for Data.gov
- Ensure that all actions pertaining to data.gov have a preservation component included that identifies what preservation commitments are in place for the data. Ideally, all data described in data.gov should have a preservation strategy in place.
- Datasets and other data.gov content should be open and machine harvestable for preservation and reuse.

Re: Action 12: Constitute a Diverse Data Governance Body
- To ensure compliance with FERMI and other Federal records management requirements, we concur with the Public Comment (Tracking Number: 1k3-9atq-pth8) that advises the inclusion of Records Management Officers in the data governance process.

Re: Action 15: Identify Data Needs to Answer Key Agency Questions
- The Foundations for Evidence-based Policy Making Act of 2018 requires agencies to develop and maintain a comprehensive inventory accounting for data assets. Accurate and useful metadata is critical to usage, access and long-term preservation of data across agencies. We recommend that the Action Plan align with the Foundations for Evidence-based Policy Making Act of 2018 by requiring agencies to create robust metadata and follow metadata standards set by the Data Documentation Initiative (DDI) Alliance <https://www.ddialliance.org> for Federal datasets.

Re: Action 16: Identify Priority Datasets for Agency Open Data Plans
- Agencies should develop open, transparent, and consistent mechanisms to encourage stakeholders outside of the agency to identify priority datasets. These mechanisms
should be designed to reach Federally-funded researchers who are data users, and others working in the public’s interest.

- Priority datasets should receive an initial preservation assessment for long-term access purposes, with remediation needs documented and a plan developed to address these needs.
- Agency-created Federal data management plans, ideally developed using consistent guidance from a cross-agency working group, would inform stakeholders of how the agency will steward open data assets.

**Response to Question 5: For each action, provide information about the implementation resources necessary to ensure success of the action steps.**

**Re: Action 8: Pilot Standard Data Catalogs for Data.gov**

- Including preservation status in the standard data catalog metadata set will support appropriate preservation and lifecycle management. Creation of this metadata and the systems needed to meet preservation needs will require additional funding.
- One approach to efficiently address preservation needs may be to collaborate with agencies administering existing, Federally-mandated content repositories such as the National Archives & Records Administration and the Government Publishing Office.

**Re: Action 15: Identify Data Needs to Answer Key Agency Questions**

- To support the identification of data needed to answer key agency questions, sufficient and appropriate descriptive metadata needs to be applied to existing data resources. Additional resources for cataloging and description may be needed to support this action.
- Note that additional investment in metadata creation at the agency level also supports implementation of the Strategy related to other identified Actions, including 6, 8-11, 13, and 16.
Respondents from the PEGI Project (institutions for identification only):

Shari Laster (Lead Contact)
Head, Open Stack Collections
Arizona State University Library
Arizona State University
P.O. Box 871006
Tempe, AZ 85287
shari.laster@asu.edu

James R. Jacobs
US Government Information Librarian and FDLP Coordinator
Stanford University Libraries
Stanford University
Stanford, CA 94305
650.862.9871
jrjacobs@stanford.edu

Scott Matheson
Associate Librarian for Technical Services and Lecturer in Legal Research
Lillian Goldman Law Library
Yale Law School
Yale University
127 Wall Street
New Haven, CT 06511
Scott.Matheson@Yale.edu

Roberta Sittel
Department Head, Government Information Connection/Eagle Commons Library
University of North Texas Libraries
1155 Union Circle #305190, Denton TX 76203-5017
940.565.4150
roberta.sittel@unt.edu