



PRESERVATION OF ELECTRONIC GOVERNMENT INFORMATION

**RE: “Leveraging Data as a Strategic Asset Phase 1 Comments” [Docket Number
USBC-2018-0011]**

July 26, 2018

On behalf of the Preservation of Electronic Government Information (PEGI) Project (www.pegiproject.org), we are grateful for the opportunity to comment on the initial draft of the Federal Data Strategy. The PEGI Project 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.

The following items from the Request For Comments for Phase 1 (83 FR 30113) are addressed in this communication:

- 1. Enterprise Data Governance [Best Practices]
- 2. Access, Use, and Augmentation [Best Practices]
- 3. Decision-Making and Accountability [Best Practices]
- 5. Principles
- 7. Stakeholder Engagement

1. Best Practices for Enterprise Data Governance

In establishing governance practices for strategically managing Federal data, an advisory board should be established to make recommendations for data management and stewardship, with substantial representation from academic and non-profit communities.¹ These communities act on behalf of the broad public interest in Federal data investments, and can advise on how Federal data stewards can responsibly leverage emerging best practices for data lifecycle management. For example, the Open Government Data Principles (https://public.resource.org/8_principles.html) developed by public advocates in 2007 articulate a public-first approach to government data to ensure that the investment in these resources is fully realized.

In general, data management practices should incorporate a lifecycle evaluation process that articulates immediate, short-term, and long-term actions, incorporating strategies that address data discoverability, accessibility, usability, and preservation. We note that the FAIR Principles (<https://www.go-fair.org/fair-principles/>) are in widespread adoption as guidance for responsible data lifecycle management, and propose that Federal data governance strategies seek to address these principles.

¹ There is precedence for this with the Depository Library Council (DLC), advising the Director of the U.S. Government Publishing Office, and the National Geospatial Advisory Committee (NGAC), advising the Secretary of the Interior or designee.

Integration with Federal information policy is essential for aligning Federal data practices with public information dissemination practices. To that end, Office of Management & Budget policies, including Circular A-130, should be amended to address public information lifecycle management, including data management, for all information dissemination products.²

2. Best Practices for Access, Use, and Augmentation

1) *Making data available more quickly and in more useful formats*

We note that data curation is an applied professional specialization within the information sciences that applies expertise in information description, access, use, and preservation, to data in all forms that are amenable for research. Key agency personnel responsible for adhering to information access and records management policies should demonstrate proficiency in this field.

Access to public data should take advantage of as many delivery channels as practicable. Versioned data with appropriate documentation and metadata should be available to download directly, with additional tools provided and supported to query, identify, subset, access, and download data from datasets that are too large for a typical desktop computer to process. The Minnesota Population Center (<https://pop.umn.edu/>) and the Missouri Census Data Center (<http://mcdc.missouri.edu/>) demonstrate two models for data delivery, both on modest operating budgets.

(2) *Maximizing the amount of non-sensitive data shared with the public*

Any datasets released in response to at least three FOIA requests should be made publicly available to all, in accordance with FOIA's "frequently requested record" provision enacted as part of the Electronic Freedom of Information Act Amendments of 1996 (E-FOIA).³ Metadata for these data sets should be included in the Federal Data.gov portal. In general, agencies should seek solutions that are scalable and not reliant on a manual or piecemeal process.

(3) *Leveraging new technologies and best practices to increase access to sensitive or restricted data while protecting privacy, security, and confidentiality, and the interests of data providers*

The Federal Statistical Research Data Center (FSRDC) program, operated by the Census Bureau in collaboration with leading research institutions across the US, is a successful partnership model that effectively balances privacy considerations with scholarly research on

² For more on Circular A-130 recommendations, see the submission by *Free Government information*. <https://freegovinfo.info/node/10581> .

³ See US Department of Justice FOIA Counselor Q&A: "Frequently Requested" Records. <https://www.justice.gov/oip/blog/foia-post-2003-foia-counselor-qa-frequently-requested-records>

behalf of the public good. We hope the FSRDC will continue to expand access to administrative and sensitive data through new agency partners and RDC locations.

3. Best Practices for Decision-Making and Accountability

(1) Providing high quality and timely information to inform decision-making and learning

We encourage agencies to coordinate data dissemination practices across offices and departments in order to build enhanced datasets for public use, while also ensuring that data linkages do *not* enable de-anonymization. For example, Economic Census data might be linked with Population & Housing Census data for better visualization of potential markets. Interfaces that enable access through geographical information systems are of increasing utility and value in policy, research, and practice, and their use frequently bridges traditional disciplinary boundaries.

(2) Facilitating external research on the effectiveness of government programs and policies which will inform future policymaking

Research depends on long-term, predictable access to detailed historical data with appropriate documentation and versioning. If data are slated for removal from a public portal, a suitable notification period with notice provided in the *Federal Register* should be required, along with a justification. Additionally, data providers should adhere to a notification period for suspension or cancellation of any ongoing data collection and dissemination programs.

(3) Fostering public accountability and transparency by providing accurate and timely spending information, performance metrics, and other administrative data

All public policy analysis and reporting should include citations to relevant data used in the making of policies and programs; and public data interfaces should include easy-to-use citation tools. The *Joint Declaration of Data Citation Principles* (<https://doi.org/10.25490/a97f-egykh>) lays out key considerations for data citations, and DataCite (<https://www.datacite.org/cite-your-data.html>) is a highly regarded initiative to improve and standardize data citation practices.

5. Principles

Proposed Core Principle: Public Use and Reuse

Data use policies can facilitate and enable commercial use and innovation, but should require that any such commercial use provide clear, explicit, prominent links back to the original, freely-available source data in order to ensure full, continued public access to Federal data. Any commercialization or privatization that removes data from the public domain results in inefficiencies and added expense to nearly all potential users, including Federal agencies, researchers, and the general public.

All metadata about federal data sets should be made available in the central Data.gov data repository. Clear licensing terms must be available for public data that allow use and reuse through both programmatic access, such as an API, and direct download by members of the public.

Stewardship

Responsible data lifecycle management demands an articulated—and funded—preservation strategy. In most cases, data and information that are not adequately preserved cannot later be authoritatively recreated or rediscovered, leading to loss of this investment.

The Federal Agencies Digital Guidelines Initiative (FADGI) demonstrates that Federal activity conducted in alignment with strategy can lead the development of best practices across broad communities of practice. For responsible stewardship of the investment in public data resources, the Federal Data Strategy must address enabling long-term access to data through the application of emerging best practices in digital preservation.

Quality

Interoperability and reuse of data are fundamentally dependent on data curation practices, including documentation, metadata, and version control. These present sufficient high-level challenges as to require dedicated resource investment, to ensure that data are useful throughout their lifecycle.

7. Stakeholder Engagement

Engagement with networks of library professionals is an effective approach to reaching experts in public use of Federal data. We recommend engaging with the Federal Depository Library Program (FDLP), operated by the US Government Publishing Office (GPO), and the State Data Center (SDC) Program, operated by the Census Bureau. Library and information professionals typically belong to additional information networks and are in an especially good position to share updates and further calls for comments.

Respondents from the PEGI Project:

Roberta Sittel (Lead Contact)
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

**Respondents from the PEGI Project
(con't):**

Marie Concannon
Head, Government Information and Data
Archives
106-B Ellis Library
University of Missouri
Columbia, MO 65201
Phone (573) 882-0748
concannonm@missouri.edu

Martin Halbert
Dean of University Libraries
University of North Carolina - Greensboro
PO Box 26170, Greensboro NC
27402-6170
336.334.3418
martin.halbert@uncg.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

Lynda Kellam
Social Sciences Data Librarian & Associate
Professor
University Libraries
University of North Carolina at Greensboro
PO Box 26170
Greensboro NC 27402
lmkellam@uncg.edu

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