

Comprehensive Data Inventory¹

I. Executive Summary

This white paper describes Federal information management requirements and how a combined agency data listing, shared with the Commission Director, may assist the Commission in its efforts.

As stated in the Evidence-Based Policymaking Commission Act of 2016, one of the Commission's charges is to "conduct a comprehensive study of the data inventory" to help fulfill its multi-faceted mission, including the evaluation of "what administrative data and survey data are relevant for program evaluation and Federal policy-making and should be included in a potential clearinghouse."² There are a number of Federal information management requirements that can assist the Commission in this effort.

II. Background

In accordance with Executive Order 13642, "Making Open and Machine Readable the New Default for Government Information,"³ the Office of Management and Budget (OMB) Memorandum M-13-13 established a framework to help institutionalize the principles of effective information management at each stage of the information's lifecycle.⁴ Agencies were to apply this framework to all information resources to promote efficiency and produce value. The framework was designed to be consistent with existing requirements in the Paperwork Reduction Act,⁵ the E-Government Act of 2002,⁶ the Privacy Act of 1974,⁷ the Federal Information Security Management Act of 2002 (FISMA),⁸ the Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA),⁹ the Freedom of Information Act (FOIA),¹⁰ the Information Quality Act,¹¹ the Federal Records Act,¹² and existing OMB guidance.¹³

¹ This white paper is intended to provide the Commission on Evidence-Based Policymaking with background information on topics relevant to the Commission's work. The paper was prepared by staff from OMB, with assistance from staff at other Federal agencies.

² Pub. L. 114-140, 130 Stat. 317 (2016).

³ Executive Order 13642, *Making Open and Machine Readable the New Default for Government Information* (May 9, 2013). <https://www.whitehouse.gov/the-press-office/2013/05/09/executive-order-making-open-and-machine-readable-new-default-government->

⁴ OMB Memorandum M-13-13, *Open Data Policy—Managing Information as an Asset* (May 9, 2013). <https://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf>.

⁵ 44 U.S.C. §§ 3501-3521.

⁶ Pub. L. 107-347.

⁷ Pub. L. 93-579, 5 U.S.C. § 552a.

⁸ Pub. L. 107-347, §§ 301-305; 44 U.S.C. § 3541, et seq.

⁹ Pub. L. 107-347, §§ 501-526; 44 U.S.C. § 3501, et seq.

¹⁰ 5 U.S.C. § 552.

¹¹ Pub. L. 106-554, § 515.

¹² 44 U.S.C. §§3101-3107.

¹³ For example, see

OMB Memorandum M-06-02, *Improving Public Access to and Dissemination of Government Information and Using the Federal Enterprise Architecture Data Reference Model* (Dec. 16, 2005), <https://www.whitehouse.gov/sites/default/files/omb/memoranda/fy2006/m06-02.pdf>.

OMB Memorandum M-10-06, *Open Government Directive* (Dec. 8, 2009), https://www.whitehouse.gov/sites/default/files/omb/assets/memoranda_2010/m10-06.pdf.

III. Information Management Policy Requirements

To improve the management of information resources throughout the information lifecycle and reinforce the government’s presumption in favor of openness, M-13-13 requires agencies to:

1. Collect or create information in a way that supports downstream information processing and dissemination activities. This includes using machine-readable and open formats, data standards, and common core and extensible metadata for all new information creation and collection efforts;
2. Build information systems to support interoperability and information accessibility;
3. Strengthen data management and release practices; and
4. Strengthen measures to ensure that privacy and confidentiality are fully protected and that data are properly secured.¹⁴

Most relevant to the Commission, the third policy requirement of M-13-13 mandated that agencies produce an enterprise data inventory and a public data listing.

IV. Enterprise Data Inventory (EDI)

To develop a comprehensive understanding of what data assets each Federal agency possesses, M-13-13 required that each agency account for all data assets created or collected, including, but not limited to, data assets used in each agency’s information systems. The enterprise data inventory (EDI) must be enterprise-wide, accounting for data assets across Federal agencies. Inventories should include data assets associated with research, program administration, statistical, and financial activities, as well as data assets produced through agency contracts and cooperative agreements, and in some cases agency-funded grants. It must also use the required Project Open Data metadata schema.¹⁵

After creating the EDI, agencies continually improve the inventory’s usefulness by:

1. Adding additional datasets to the EDI (*expansion*),
2. Improving the quality of metadata describing each data asset (*enrichment*), and
3. Implementing tools and processes that will accelerate the opening of additional valuable data assets by making them public and machine-readable, while ensuring adequate policy, process,

OMB. Digital Government: Building a 21st Century Platform to Better Serve the American People (May 23, 2012). <https://www.whitehouse.gov/sites/default/files/omb/egov/digital-government/digital-government-strategy.pdf>.

OMB Memorandum M-13-13, Open Data Policy—Managing Information as an Asset (May 9, 2013).

OMB Circular A-130, http://www.whitehouse.gov/omb/Circulars_a130_a130trans4/.

¹⁴ OMB Memorandum M-13-13, *Open Data Policy—Managing Information as an Asset* (May 9, 2013). <https://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf>. Policy requirements 1 and 2 apply to all new information collection, creation, and system development efforts, as well as major modernization projects that updated or redesigned existing information systems. National Security Systems, as defined in 40 U.S.C. § 11103, are exempt from the requirements of this policy. Policy requirement 3 applies to the management of all datasets used in an agency’s information systems. Agencies were also encouraged to improve the discoverability and usability of existing datasets by making them “open,” prioritizing those that have already been released to the public or otherwise deemed high-value or high-demand through engagement with customers. Agencies were directed to exercise judgment before publicly distributing data residing in an existing system by weighing the value of openness against the cost of making those data public.

¹⁵ See <https://project-open-data.cio.gov/v1.1/schema/>.

and technical safeguards are in place to prevent against the release of sensitive data (*openness*).¹⁶

Agencies should update their EDIs as necessary, but at a minimum they are required to update and submit them to OMB on a quarterly basis.¹⁷ For the Commission's purposes, it is important to note that EDIs do not include state, local, and other grantee-level information that is collected as a part of administering Federal programs, which does not reside at the Federal level.

Since May 2015, OMB has required agencies to post their entire EDI as a public, machine-readable JavaScript Object Notation (JSON) file¹⁸ on their agency websites in lieu of posting a PDL (see below). Agencies should coordinate with their FOIA offices to determine whether any descriptive information associated with a non-public data asset is subject to a FOIA exemption and should not be disclosed. If this information is determined to be subject to a FOIA exemption, the descriptive information is to be redacted. However, agencies should continue to send complete, unredacted EDIs to OMB on a quarterly basis.

V. Public Data Listing (PDL)

In addition to an EDI, agencies are required to produce a subset of the EDI containing all public data assets called a public data listing (PDL). Agencies publish the PDL on their websites as a machine-readable JSON file. The General Services Administration harvests these data automatically at scheduled intervals to populate Data.gov, the public facing data repository of the U.S. Government's open data. Data.gov contains data, tools, and resources to conduct research, develop web and mobile applications, design data visualizations, and more. In addition, M-13-13 requires agencies to list any datasets in their EDIs that can be made publicly available at [www.\[agency\].gov/data](http://www.[agency].gov/data).¹⁹

Figure 1 shows the universe of Federal data assets covered by EDIs. On the left are the data assets that can be made publicly available. These assets are either already in the public domain and should be listed in agency PDLs, or are not in the public domain but should be moving toward release in the future. On the right side of Figure 1 are the data assets that *cannot* be made publicly available, including assets that are already open to other Federal agencies, assets that are not open but could be,²⁰ and assets that cannot be made available to other Federal agencies.

¹⁶ See <https://project-open-data.cio.gov/implementation-guide/>.

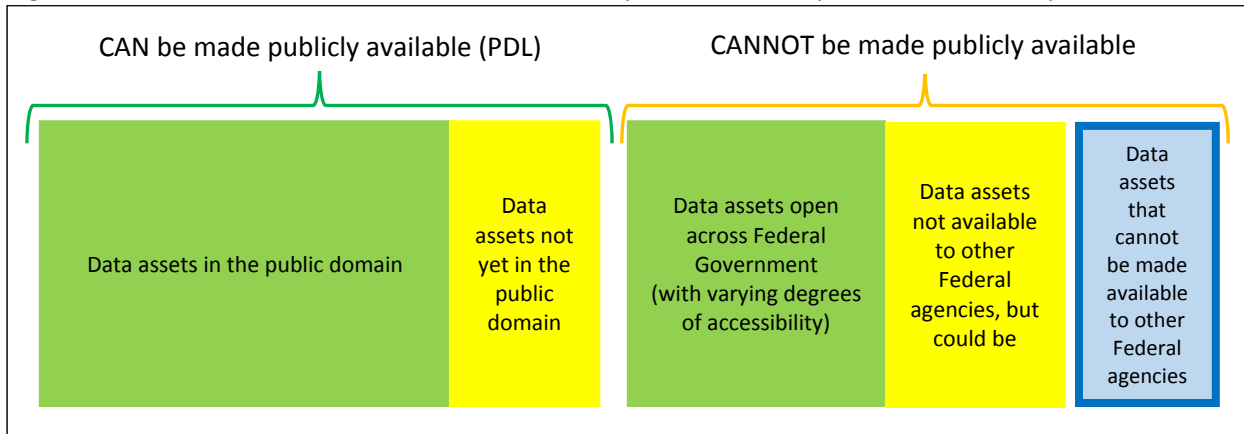
¹⁷ See *ibid.*

¹⁸ JavaScript Object Notation (JSON) is an open-standard format that uses human-readable text to transmit data objects consisting of attribute-value pairs.

¹⁹ OMB Memorandum M-13-13, Section III.3.b.

²⁰ OMB Memorandum M-14-06, *Guidance for Providing and Using Administrative Data for Statistical Purposes*, encourages Federal departments and agencies to promote the use of administrative data for statistical purposes and provides guidance in addressing legal and policy requirements for such uses. The Memorandum also provides some "best practice" tools, including detailed guidance on the interaction of the Privacy Act requirements and the use of administrative data for statistical purposes, as well as a model interagency agreement for departments and agencies to follow when developing their policies and practices for sharing data for statistical purposes to another department or agency.

Figure 1. Universe of Federal Data Assets Covered by M-13-13 Enterprise Data Inventory



Note: Does not include state, local, and other grantee-level information that is caused to be collected as a part of administering Federal programs, but which does not reside at the Federal level.

VI. Combined Agency Public Data Listing

OMB has shared with the Commission Director a file containing a human-readable spreadsheet of the agency PDLs as of February 12, 2016 and can share this with Commissioners individually upon request. It contains 72,757 records and 12 Project Open Data standard fields. Each field is described in Appendix A, accompanied by two case studies to highlight the range of files included in the PDLs. Given the nested structure of JSON files, not all available fields could be included in the file. The excluded fields are listed in Appendix B.

Appendix A: Project Open Data Metadata Fields Included in the Combined Agency PDL

Field	Definition	Required?	Completeness (out of 72,757)	Case Study #1 - CFPB Database	Case Study #2 - FNS Memorandum
Agency <i>(created field)</i>	Agency name, based on OMB Circular A-11. Created for the Commission.	N/A	100%	Consumer Financial Protection Bureau	Department of Agriculture
Publisher	The publishing entity and optionally their parent organization(s).	Always	100%	Consumer Financial Protection Bureau	Food and Nutrition Service
Title	Human-readable name of the asset. Should be in plain English and include sufficient detail to facilitate search and discovery.	Always	100%	Consumer Complaint Database	School Food Authorities (SFAs) Purchasing Produce from U.S. Department of Defense (DoD) Vendors using Section 4 and 11 for Fresh Fruit and Vegetable Program (FFVP) Funds Memo
Description	Human-readable description (e.g., an abstract) with sufficient detail to enable a user to quickly understand whether the asset is of interest.	Always	99.9%	These are complaints we've received about financial products and services.	This memorandum clarifies how SFAs may use funds provided under Sections 4 and 11 or 19 of the National School Lunch Act (NSLA) to purchase fresh fruits and vegetables from DoD Fresh Fruit and Vegetable Program (DoD) vendors.
Modified	Most recent date on which the dataset was changed, updated or modified.	Always	99.9%	11/6/2013	3/6/2014

Field	Definition	Required?	Completeness (out of 72,757)	Case Study #1 - CFPB Database	Case Study #2 - FNS Memorandum
Identifier	A unique identifier for the dataset or application program interface (API) as maintained within an Agency catalog or database.	Always	99.9%	CCDB	USDA-FNS-84
License	The license or non-license (i.e., Public Domain) status with which the dataset or API has been published. See Open Licenses for more information.	If Applicable	69.1%	(empty)	https://creativecommons.org/publicdomain/zero/1.0/
ContactPoint	Contact person's name and email for the asset.	Always	99.9%	mailto:devops@cfpb.gov	mailto:Kevin.Russ@fns.usda.gov
DescribedBy	URL to the data dictionary for the dataset. Note that documentation other than a data dictionary can be referenced in the field, "References."	No	7.5%	http://www.consumerfinance.gov/complaintdatabase/technical-documentation/#field-reference	(empty)
BureauCode	Federal agencies, combined agency and bureau code from OMB Circular A-11, Appendix C (PDF ¹ , CSV ²) in the format of 015:11.	Always	99.9%	581:00:00	005:84

¹ https://www.whitehouse.gov/sites/default/files/omb/assets/a11_current_year/app_c.pdf

² https://project-open-data.cio.gov/data/omb_bureau_codes.csv

Field	Definition	Required?	Completeness (out of 72,757)	Case Study #1 - CFPB Database	Case Study #2 - FNS Memorandum
AccessLevel	The degree to which this dataset could be made publicly-available, regardless of whether it has been made available. Choices: public (Data asset is or could be made publicly available to all without restrictions), restricted public (Data asset is available under certain use restrictions), or non-public (Data asset is not available to members of the public).	Always	99.9%	public	public

Appendix B: Nested Project Open Data Metadata Fields Not Included in the Combined Agency PDL

Field	Definition	Required?
Keyword	Tags (or keywords) help users discover your dataset; please include terms that would be used by technical and non-technical users.	Always
ProgramCode	Federal agencies, list the primary program related to this data asset, from the Federal Program Inventory. Use the format of 015:001.	Always
Rights	This may include information regarding access or restrictions based on privacy, security, or other policies. This should also serve as an explanation for the selected "accessLevel" including instructions for how to access a restricted file, if applicable, or explanation for why a "non-public" or "restricted public" data asset is not "public," if applicable. Text, 255 characters.	If Applicable
Spatial	The range of spatial applicability of a dataset. Could include a spatial region like a bounding box or a named place.	If Applicable
Temporal	The range of temporal applicability of a dataset (i.e., a start and end date of applicability for the data).	If Applicable
AccrualPeriodicity	The frequency with which the dataset is published.	No
IsPartOf	The collection of which the dataset is a subset.	No
References	Related documents such as technical information about a dataset, developer documentation, etc.	No
DataQuality	Whether the dataset meets the agency's Information Quality Guidelines (true/false).	No
Language	The language of the dataset.	No
Theme	Main thematic category of the dataset.	No
AccessURL	URL providing indirect access to a dataset, for example via API or a graphical interface.	If Applicable
DownloadURL	URL providing direct access to a downloadable file of a dataset.	If Applicable