



**United States Department of Agriculture
Dataset Publication Process
Version 2.0**

USDA Dataset Publication Process

The ultimate goal for publishing datasets is to provide increased access to the vast amounts of data and information produced by the United States Department of Agriculture's (USDA) nineteen agencies research and program areas. Much of the data is a direct result of USDA's Research, Economics, and Education mission area where there are currently more than 230 datasets, application programming interface (API), maps and charts available to the public. As USDA continues to improve its dataset publication process, it will streamline and automate, incorporate best practices, analyze dataset usage, and employ a process to collect customer feedback to enrich the dataset quality and quantity for key stakeholders and American public..

Since November 2013, USDA and the General Services Administration (GSA) have improved the dataset publishing process, and will continue to implement a multi-step approach to ensure private or sensitive information is not released to the public. USDA has established an Open Data Council (ODC) to: 1) Implement the Open Data Policy; 2) Reinforce Departmental Priorities; and 3) Advocate Information Sharing. The multi-step process involves stakeholders such as: Program Managers and Scientists, Agency and Staff Office Data Managers, Agency and Staff Office Privacy Officers, Office of the General Counsel (OGC), Agency and Staff Office Data Stewards, Enterprise Architects, and Web site Specialists. The high level multi-organizational process is outlined below. Figure 1: USDA Dataset Publication Process is a graphic of the multi-organizational process. The information below outlines the defined roles and responsibilities for each of the stakeholders involved in the dataset publication process.

Open Data Council

The USDA ODC establishes departmental priorities and oversees implementation of the Federal Agency requirements outlined in the President's Open Data Policy, Office of Management and Budget (OMB), Memorandum [M-13-13](#), *Open Data Policy – Managing Information as an Asset*. The ODC coordinates with other USDA organizations to ensure that deliverables and milestones are well-aligned with the overall digital government strategies, activities, and priorities of USDA.

The ODC actively disseminates open data policies and procedures that promote internal data asset management practices for:

- Collecting or creating information in a way that supports downstream processing and dissemination activities;
- Building information systems that support interoperability and information accessibility;
- Strengthening data management and release practices;
- Reinforce measures ensuring that privacy and confidentiality are fully protected and that data are properly secured; and
- Incorporate new interoperability and openness requirements into core agency and staff office processes.

The first step in the data publishing process is to select *Dataset Priorities for Open Data*. The ODC publishes quarterly priorities to help guide and focus agency and staff office activity. To accomplish this task, the ODC receives and approves / disapproves recommendations from the Open Data Working Group (ODWG). The ODWG uses a variety of sources to formulate recommendations to the ODC, including but not limited to:

- President's Management Agenda
- OMB Cross-Agency Goals
- Secretary's Priorities and Key Initiatives
- Key Stakeholder feedback

Agency and Staff Offices

Agencies and staff offices are responsible for:

- Identifying datasets to be published;
- Initiating and conducting privacy, legal, and security review;
 - Performing security reviews to ensure the data being released does not pose a threat to national security.
 - Agency Privacy Officers use Federal regulations, USDA Regulations, National Institute of Standards and Technology (NIST) Special Publications (SP), the USDA developed guidance and templates to review the datasets and prepare a justification. A sampling of reference materials and resources used in making the determination are included in Appendix A.
 - Agency Privacy Officers in conjunction with the Information System Security Program Manager, and the Agency Freedom of Information Officer prepares justification when an agency recommends a dataset not be published for public release.
 - A multi- agency subcommittee will convene to review the agency's justification. As appropriate, the USDA Privacy Office will consult with the OGC to make a final determination on withholding a dataset. The OGC will review the datasets determined not for public consumption, and provide counsel to the USDA Privacy Office on other issues, such as the mosaic effect.
- Preparing datasets for submission and submitting datasets to Comprehensive Knowledge Archive Network (CKAN); a data catalog on an open source data management system used by the data portals of the United Kingdom (U.K.), Germany, Brazil, and a number of other governments around the world;

- Generating the agency or staff office Public Data Listing (PDL) containing each dataset; Data stewards will ensure that no error files are produced. Should an error file be produced, then the steward will remediate the error and correct the metadata;
- At a minimum, the Data Steward includes with the submission the mandatory metadata for all datasets, and a description of the dataset and its contextual use. The Common Core Metadata Schema includes the following mandatory metadata information:
 - Title
 - Description
 - Tags (Keywords)
 - Last Update
 - Publisher
 - Contact Name
 - Contact Email
 - Unique Identifier
 - Public Access Level (Public, Restricted, Private)
 - Bureau Code
 - Program Codes;

Additional required metadata items, if applicable, include but are not limited to:

- Access Level Comment (for Restricted and Private only)
- Download Uniform Resource Locator (URL) (for all Public)
- Endpoint (for data collections with API capability)
- Format (i.e., CSV, XLSX, API)
- License
- Spatial (geographic or location name)
- Temporal (start and end dates of data applicability).

Office of the Chief Information Officer

Enterprise Architects within the Office of the Chief Information Officer (OCIO) are responsible for:

- Coordinating with Agencies and Staff Offices to assist agency data stewards with the CKAN tool as they collect dataset metadata information;
- Updating the Enterprise Data Inventory (EDI) in Max Collect for OMB;
- Sending consolidated PDL to Office of Communications (OC) for final publication to the public;

Office of Communications

The OC is responsible for:

- Converting the OCIO provided JavaScript Object Notation (JSON) file, data.json, to Extensible Markup Language (XML) file, data.xml;
- Reformatting the XML file with the 'dataset' tags, it will be used for usda.gov/data page;
- Generating the HyperText Markup Language (HTML) file;
- Creating the Web page html file, data.html;
- Adding the Agency and Staff Office alpha tags and back to top to the file to help navigation;
- Publishing the datasets (data.json, data.xml and data.html).

Data Consumers

It is the intention of USDA that our data consumers will actively participate in a process to improve the quality of data delivered to them to fuel entrepreneurship, innovation, and scientific discovery-all of which improve Americans' lives and contribute significantly to job creation.

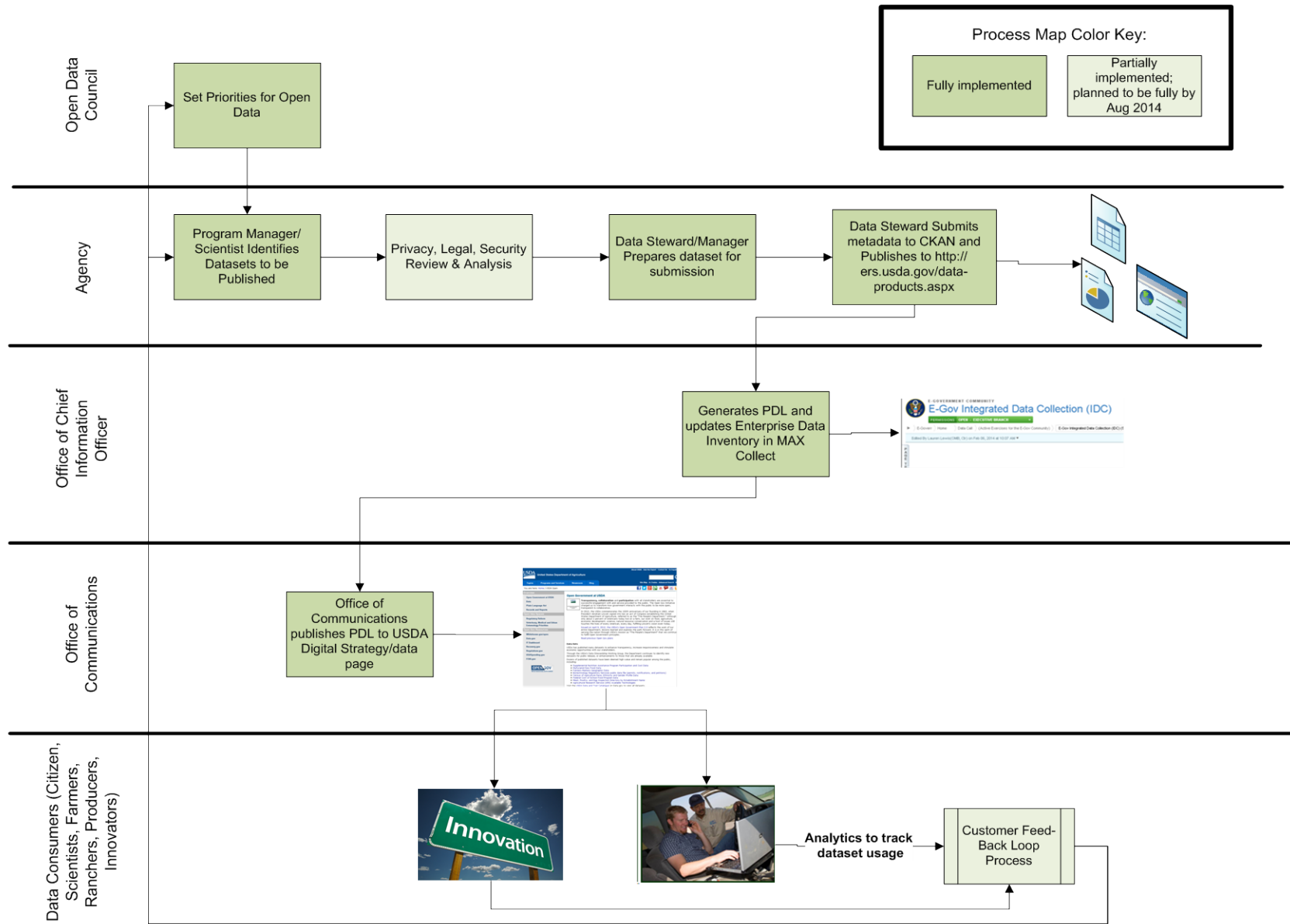


Figure 1: USDA Dataset Publication Process

Appendix A: USDA Dataset Publication Process References

Appendix A contains a sampling of reference material used to make the determination as whether data should be made public.

- DM3515-000 [[HTM](#)] [[PDF](#)] (02/17/05) *Privacy Requirements*
- DM3515-002 [[HTM](#)] [[PDF](#)] (02/17/05) *Privacy Impact Assessment*
- Office of Management Budget Memorandum OMB Memo M-11-02, *Sharing Data While Protecting Privacy, November 3, 2010.*
- M-03-22, *OMB Guidance for Implementing the Privacy Provisions of the E-Government Act of 2002*
- OMB M-07-16, *Safeguarding Against and Responding to the Breach of Personally Identifiable Information*
- OMB M 99-05, *Attachment B (Privacy and Personal Information in Federal Records)*
- [OMB Circular A-130, Appendix I](#) (marked)
- NIST SP 800-122, *Guide to Protecting the Confidentiality of Personally Identifiable Information (PII)*
- Freedom of Information Act Exemptions
- Executive Order 13556 -- Controlled Unclassified Information: <http://www.whitehouse.gov/the-press-office/2010/11/04/executive-order-controlled-unclassified-information>
- DM3540-002 [[HTM](#)] [[PDF](#)] (08/19/04) *Risk Assessment and Security Checklists*
- DM3550-002 [[HTM](#)] [[PDF](#)] (02/17/05) *Sensitive but Unclassified (SBU) Information Protection*
- [DR3080-001](#) *Records Management* (05/23/13)
- DR3440-002 [[HTM](#)] [[PDF](#)] (01/30/03) *Control and Protection of “Sensitive Security Information*
- DR3450-002 [[HTM](#)] [[PDF](#)] (02/07/03) *FOIA Implementing Regulations*
- [Executive Order 13556](#), *Controlled Unclassified Information*
- *Freedom of Information Act (FOIA)*
- *Freedom of Information Act Exemptions*
- NIST, [SP 800-122](#), *Guide to Protecting the Confidentiality of Personally Identifiable Information (PII)*, April 2010
- OMB, M 99-05, *Privacy and Personal Information in Federal Records*, [Appendix B](#), *Instructions for Complying with the President’s Memorandum*, January 7, 1999
- OMB, [M-03-22](#), *OMB Guidance for Implementing the Privacy Provisions of the E-Government Act of 2002*, September 26, 2003
- OMB, [M-07-16](#), *Safeguarding Against and Responding to the Breach of Personally Identifiable Information*, May 22, 2007
- OMB, [M-11-02](#), *Sharing Data While Protecting Privacy*, November 3, 2010

- OMB, M-13-13, *Open Data Policy – Managing Information as an Asset*
- OMB, Circular A-130, *Management of Federal Information Resources, Appendix I, Federal Agency Responsibilities for Maintaining Records About Individuals*, (marked), November 28, 2000
- *Privacy Act of 1974*, 5 U.S.C. Section 552a