

US Environmental Protection Agency (EPA)

Strategic Data Action Plan (SDAP) Version 1.0

March 29, 2011



Table of Contents

Table of Contents.....	i
I. Background and Foundation.....	1
A. Purpose and Approach.....	1
B. Guiding Principles	2
C. Goal and Objectives	3
D. Prioritization of Data Resources.....	3
II. Plan and Schedule.....	4
A. SDAP.....	4
B. Data Inventory	5
C. Data.gov Dashboard	5
D. Expansion of Environmental Dataset Gateway (EDG).....	5
E. Data Policy	5
F. SDAP Transition Plan.....	6
G. Public Participation, Feedback and 3rd Party Use	6
H. Linked Open Data (LOD) Pilots.....	6
I. Repository of Best Practices and Principles.....	6
III. Accountability.....	7
A. Measures.....	7
B. Quarterly Reporting	7
C. Governance.....	7
D. Data.gov Roles and Responsibilities.....	7
E. Emerging Initiatives and Issues.....	8
IV. Data Standards, Metadata, Quality and Security.....	9
A. Data Standards	9
B. Metadata	9
C. Quality	9
D. Data Security and Privacy	10
Appendix A: SDAP Project Team	11
Appendix B: Select, Prepare, and Add a Dataset to Data.gov.....	12
A. Data.gov and Enterprise Data Resource Terms	12
B. Data.gov Process	13
C. Evaluate Public Usefulness.....	13
D. Confirm Characteristics of the Data	14
E. Ensure Quality of the Data	14
F. Protect Sensitive Data.....	14
G. Best Practices for Data.gov Publication	14
H. Open Formats	14

EPA Strategic Data Action Plan (SDAP) Version 1.0 - DRAFT



I. Data.gov Roles and Responsibilities	15
Appendix C: Metadata Services Integration.....	17
Appendix D: Website Links to Resources.....	18

EPA Strategic Data Action Plan (SDAP)

Version 1.0 - DRAFT



The Environmental Protection Agency's (EPA) Strategic Data Action Plan (SDAP) is designed to increase transparency by creating a planning framework to establish and track ways to more systematically provide improved access to and increase the understanding of EPA's data holdings for public and Agency uses. This plan establishes goals, guiding principles, and initial processes and procedures for strategic data management. We also define and set a schedule for actions to establish enterprise policy and governance mechanisms, in addition to implementing improved methods for data management intended to expand and promote the use of EPA's publicly-available data.

The SDAP is also intended to serve as a guide to data owners, managers, program/regional Information Management Officers (IMOs), and others in managing EPA data resources and promoting their access and use as appropriate. The public is also an anticipated user of this document for the purposes of following EPA's data management plans and providing feedback related to the Agency's plan.

An initial focus of this plan is to leverage and streamline the use of EPA's suite of data-related registries to provide better access to and information about the following types of environmental data resources:

- High-value data¹
- Publicly-requested data
- Regulatory compliance data

The SDAP was developed and is being implemented and monitored by the SDAP Project Team, a working group comprised of offices and regions across the Agency. See Appendix A for a listing of the offices and regions on the team.

I. Background and Foundation

The Administration's Open Government (OpenGov) initiative is detailed in [President Obama's January 2009 memorandum](#) and the December 2009 Office of Management and Budget (OMB) [Open Government Directive \(OGD\)](#).

In EPA's April 2010 OpenGov Plan, we committed to develop the SDAP to establish and implement Agency processes to increase transparency by more systematically managing and disseminating EPA's information². In addition, we will leverage the SDAP to help coordinate future data management and access requests such as the January 18, 2011 Presidential Memorandum on Regulatory Compliance directing agencies to publicly release compliance information.³

A. Purpose and Approach

The purpose of the SDAP is to:

- Define EPA's plan for improving data management for transparency, collaboration, and participation

¹ High-value data is any data that can "increase agency accountability and responsiveness; improve public knowledge of the agency and its operations; further the core mission of the agency; create economic opportunity; or respond to need and demand as identified through public consultation."

² Management and quality of financial information adheres to additional requirements as noted in the Agency's [Open Government Data Quality Plan](#). The Agency's Spending Information Quality (SIQ) Workgroup manages the data quality requirements for spending data such as grants and contracts data.

³ Presidential Memorandum of January 18, 2011, on Regulatory Compliance.
<http://www.gpo.gov/fdsys/pkg/FR-2011-01-21/pdf/2011-1386.pdf>.

EPA Strategic Data Action Plan (SDAP)

Version 1.0 - DRAFT



- Mature and foster consistency in the processes and procedures associated with publishing EPA data as introduced in the December 2010 Interim EPA Data.gov Guidance and Procedures⁴
- Identify and track the development of tools and processes to support expanded data access and usability
- Establish measures of success and the method to track and report on progress
- Identify and explore areas of broader data management that need development or improvement
- Pilot new approaches and address recommendations to improve standard practices

EPA is approaching the broader SDAP effort in an iterative manner with distinct phases. As a first step, the SDAP Project Team developed an SDAP outline⁵ to set the process and approach the Agency is taking to address the information management requirements of the OpenGov Directive. One of the most explicit requirements of the OGD instructed agencies to publish their data in Data.gov. In order to provide Agency data owners with more detailed instructions for populating Data.gov, the SDAP Project Team created the Interim EPA Data.gov Guidance and Procedures.

Using feedback from Agency users and other stakeholders, we have updated related processes as documented in this SDAP V1.0. We will revise the SDAP based on the schedule provided in the SDAP Outline, or as otherwise needed. In keeping with the intent and requirements of the OGD, we anticipate that the Agency-wide data policy and related procedures will be developed and put in place to fully address enterprise data holdings.

B. Guiding Principles

EPA reviewed and analyzed a collection of data management and data publishing principles from internal and external sources with the goal of selecting and committing to a set of clear and effective principles that can guide the forthcoming enterprise data policy. We are continually seeking feedback and working collaboratively with internal and external stakeholders to further develop and refine these guiding principles.

- **Provide Easy and Timely Access to Quality Data.** The public will have access to high quality data in a timely manner that is easy to discover and access.
- **Publish Data in Open Form.** Public data will be published using open, structured, computer-readable form and following open standards. Examples include CSV, XML, KML, RSS, and RDF.
- **Help the Public Understand the Data (Metadata).** Public data will have descriptive and informative metadata that is understandable to the general public as well as technical users.
- **Enhance Data Use.** Data services (such as web services) and tools (such as Application Programming Interfaces (APIs)) will be provided to enhance the public's ability to use EPA's data resources. In addition, data should be provided at the lowest level of analytical unit that best supports ease of consumption by the public.
- **Drive Best Uses/Best Practices.** Lessons learned and best practices will be continually evaluated and incorporated so that the best current technologies and methods for high quality data sharing are communicated and implemented across the Agency.

⁴ Interim EPA Data.gov Guidance and Procedures, December 21, 2010.
http://www.epa.gov/open/EPA_DataGov_Guidance_Procedures_2010_12_21.pdf.

⁵ Draft US EPA SDAP Annotated Outline, December 2, 2010.
http://www.epa.gov/open/Draft_EPA_SDAP_Outline_12-02-10_rev.pdf.

EPA Strategic Data Action Plan (SDAP)

Version 1.0 - DRAFT



- **Grow and Improve Through User Feedback.** Feedback from the public will be used to help set priorities and identify high value datasets to publish.

C. Goal and Objectives

EPA's goal is to:

Improve the experiences of the public, other stakeholders, and Agency staff in finding, accessing, and using our data to support informed decisions with the appropriate breadth and depth of information.

This goal is supported by the following objectives:

- Promote, improve, and simplify data discovery and use for internal and external customers
- Continue to gather feedback on EPA data resources from internal and external customers
- Continue to respond to public and Administration requests for specific types of information
- Create the process to select, prepare, and make EPA information publicly available
- Identify and track the development of tools and procedures to support improved data access and usability
- Develop measures of SDAP success and track those measures
- Identify any gaps in information we currently provide or plan to provide to the public
- Streamline and simplify metadata capture
- Continue maturing EPA's data management
- Develop policies and procedures to promote and formalize improvements in data management

We will pursue these objectives while balancing the efficient, effective, and economic use of resources.

D. Prioritization of Data Resources

The supporting policies, procedures, and best practices that are part of the SDAP apply to all Agency data resources. Given resource constraints, our approach is to develop a prioritization process that follows a logical implementation pattern. The following existing classification sequence is a good example:

- Major Capitol Planning and Investment Control (CPIC) systems/resources
- CPIC Lite (non-major) systems/resources
- Small and other

The SDAP Project Team is developing further clarification to help Agency offices and regions prioritize their data resources for publishing in Data.gov. As part of the process, the project team will review and incorporate priorities from enterprise efforts such as the data policy to capture and address the needs of the Agency, along with feedback from the public in helping identify what is most important to them. The Agency's early work on the data policy highlights the importance of additionally considering a data resource's maturity and its support of EPA's mission.

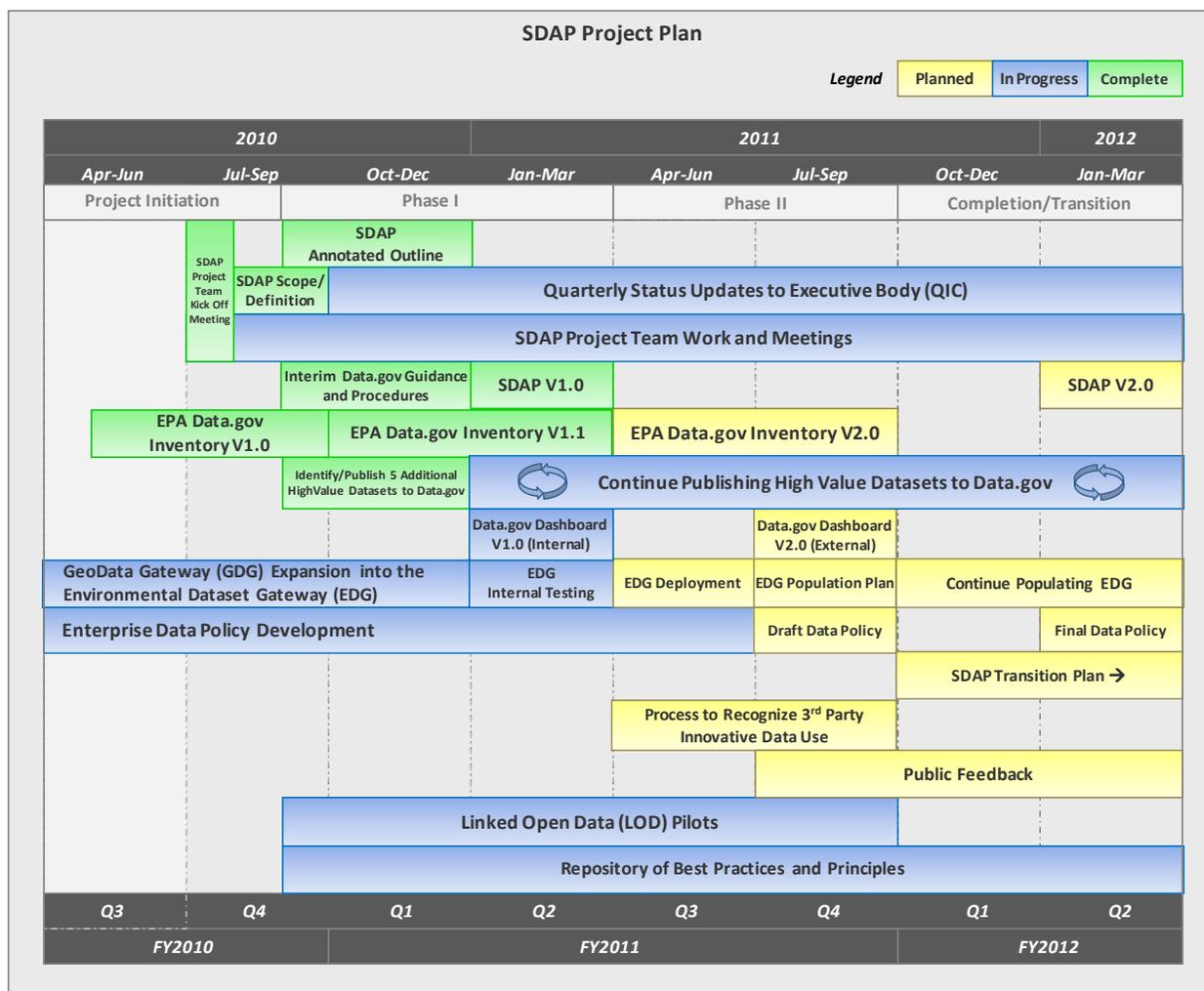
The Agency does not plan to apply SDAP to data resources containing Confidential Business Information (CBI) or sensitive data that are not available for public access.



II. Plan and Schedule

The SDAP is following a phased implementation with milestones and artifacts identified and tracked as complete, in progress, or planned. As part of SDAP monitoring, related enterprise efforts are also being tracked for purposes of reporting. Figure 1 presents the SDAP plan and schedule on a quarterly basis, showing the period of development and end point when a milestone is reached or an artifact is completed.

Figure 1: SDAP Project Plan



The remainder of this section presents the planned and in progress SDAP-related artifacts, along with a brief description and key implementation dates.

A. SDAP

This document, SDAP V1.0, incorporates the Interim Data.gov Guidance and Procedures, thereby superseding the Interim document. The SDAP is designed to be flexible and adapt easily to incorporate feedback from external reviews, as well as any relevant advances in information management. This

EPA Strategic Data Action Plan (SDAP)

Version 1.0 - DRAFT



document can be updated as frequently as needed with minor revisions until V2.0. The SDAP is expected to be superseded by the data policy.

- SDAP V1.0 Q2/2011 (Complete - this document)
- SDAP V2.0 Q2/2012

B. Data Inventory

In order to more easily identify and manage data sets and tools that EPA posts to Data.gov, the Agency created an inventory that is updated and posted on a quarterly basis (<http://www.epa.gov/open/data-gov-inventory2011Q1.xls>). This inventory has been updated to reflect data sets identified by other efforts to evaluate for inclusion on Data.gov. This updated spreadsheet serves as V1.1. We will evaluate existing tools and inventories and determine those best suited and easily adapted to provide the public a user-friendly and automated inventory of Agency data holdings.

- Data.gov Inventory V1.0 & V1.1 Q2/2011 (Complete and posted to EPA's OpenGov website)
- Data.gov Inventory V2.0 Q4/2011

C. Data.gov Dashboard

To help us gauge and measure our success in making data available, EPA is developing an internal dashboard. The initial EPA-only version of the dashboard is intended to measure office-specific contributions to Data.gov with the ability to provide publishing prioritization and sequencing information. We plan to provide a public facing dashboard with the V2.0 version.

- Data.gov Dashboard V1.0 Q2/2011
- Data.gov Dashboard V2.0 Q4/2011

D. Expansion of Environmental Dataset Gateway (EDG)

The Environmental Dataset Gateway (EDG) is where all EPA datasets and geodata will be registered. The EDG registry tool is an expansion of the GeoData Gateway (GDG) and will improve and simplify how EPA creates and manages metadata for individual datasets/geodata. Currently the GDG and Data.gov Dataset Management System (DMS) are used to create or provide metadata for Data.gov. The EDG will become the primary mechanism to publish data resources to Data.gov, and the implementation timeline is provided below:

- EDG Internal Testing Q2/2011
- EDG Deployment Q3/2011
- EDG Population Plan Q4/2011

E. Data Policy

To more clearly communicate existing data management practices and develop points of coordination across the Agency, an effort to create an enterprise-wide data management policy is underway. One goal of this effort is the alignment of very mature practices such as those supporting research and scientific data with other less mature practices for the benefit of consistency and improvements across the board. The data policy is the SDAP-related product within the enterprise-wide data management policy effort.

- Draft Data Policy Q4/2011
- Final Data Policy Q2/2012



F. SDAP Transition Plan

The work of the SDAP Project Team is intended to fold into long-term and sustained enterprise efforts such as the data policy. The SDAP was not intended to be a continuous effort, and the SDAP Project Team will sunset once all components of the SDAP are incorporated into enterprise policy and procedures, and the team is no longer needed.

- Transfer SDAP to Standing Team Q4/2012

G. Public Participation, Feedback and 3rd Party Use

OpenGov is focused on sharing and disseminating information as well as engaging the public, a recognized key consumer of the data. EPA solicited feedback from several audiences to determine the most requested environmental topics and how the public wants to use data (see EPA's National Dialogue on Access to Environmental Information (www.epa.gov/nationaldialogue)). EPA will build on this knowledge to engage the public using a variety of social media tools including IdeaScale (<http://www.openepa.ideascale.com/>), the Data Finder Forum (<http://blog.epa.gov/data>), Greenversations blog (<http://blog.epa.gov/blog/>), Twitter, and Facebook. EPA also encourages third-party use of Agency data to create innovative solutions to environmental problems. EPA will engage and inform the developer community and other private sector parties in what data are available and monitor the use of EPA data by third parties.

- 3rd Party Recognition Process Q4/2011
- Public Feedback Ongoing

H. Linked Open Data (LOD) Pilots

The Administration has made large amounts of Federal data available to the public in one place via Data.gov. Now that many datasets, tools, and geodata are currently available on Data.gov and continue to be published, OMB and the Data.gov PMO are encouraging leading agencies to transform this data into a form that is easily "linked" or mashed with other data. The Agency is currently engaged in an initial pilot to transform datasets into Linked Open Data (LOD) form to improve and expand data discoverability and enable mashups of information from multiple EPA datasets, as well as with non-EPA datasets. EPA is learning and documenting best practices in the approach to transform EPA data into LOD -- a self-describing, machine-readable format with identified relationships. The Agency will evaluate the success of this initial pilot and share the outcome and best practices across the Agency and with other Federal agencies.

- LOD Pilots Q4/2011

I. Repository of Best Practices and Principles

The Agency began collecting best practices and principles at the outset of our OpenGov work, and we will continue to collect these lessons learned throughout the lifecycle of SDAP. Currently these lessons learned are available to Agency employees via EPA's internal Wiki.

- Best Practices and Principles Repository Ongoing



III. Accountability

With competing demands for resources, the Agency needs appropriate accountability to ensure commitments within the SDAP are met. The SDAP Project Team is responsible for tracking and measuring progress, developing quarterly reports, and reporting on progress within the OpenGov governance structure.

A. Measures

As an initial set of measures, the SDAP Project Team will monitor the progress of the milestones as laid out in the SDAP schedule, detailed in Section II, Figure 1 of this document. We will develop measures for each corresponding goal during the planning phase of that goal. These measures will then be tracked and reported quarterly.

B. Quarterly Reporting

In our quarterly reports on OpenGov execution, we describe the qualitative and quantitative impact that OpenGov is having on EPA performance. The quarterly tracking and reporting process is part of an ongoing evaluation of EPA's OpenGov programs and policies. In addition to the standard quarterly reporting required, the SDAP V1.0 lays the foundation for implementing and maintaining additional quarterly tracking and reporting for upcoming and planned SDAP initiatives.

C. Governance

The SDAP and associated activities fall under the OpenGov governance structure, which was developed in response to the OGD to:

- Guide immediate activities for the OpenGov initiative
- Populate Data.gov
- Leverage and enhance existing Agency policies and management controls
- Leverage and better integrate EPA's data management mechanisms to minimize the burden of properly managing data
- Begin to institutionalize open government within EPA's organizational structure

The primary data governance body focused on SDAP is the Open Government Implementation Work Group (OGIWG), and the SDAP Project Team falls under this workgroup. The OGIWG is a working group established to directly execute OpenGov goals and objectives and develop recommendations to be approved by the Quality Information Council (QIC), an executive body.

Additional information on EPA's data governance for OpenGov is described in Section I.D.2 of EPA's Open Government Plan 1.1, dated June 25, 2010 (http://www.epa.gov/open/EPAOpenGovernmentPlan_11.pdf).

D. Data.gov Roles and Responsibilities

There are several categories of staff both within EPA and the General Services Administration (GSA) involved in bringing a dataset/tool/geodata to Data.gov. Appendix B: Section F describes the roles and responsibilities of the staff involved in the preparation, review, and publishing of data resources to Data.gov.



E. Emerging Initiatives and Issues

EPA has identified a number of emerging issues, initiatives, and technologies that may offer opportunities to improve and streamline the life cycle management of, transparent access to, and enhanced usage of EPA data. These areas will be assessed on an ongoing basis and, as appropriate, we will establish milestones and report on progress. This list is not intended to be comprehensive, but reflects areas identified thus far:

- EPA's Freedom of Information Act (FOIA) processes
- EPA's implementation of Information Collection Request (ICR) processes under the Paperwork Reduction Act
- EPA's "OneEPA Web" web modernization effort
- EPA's CPIC and enterprise architecture programs
- EPA's Data Finder (epa.gov/data) and related efforts
- The ongoing evolution of the Data.gov platform and new opportunities it may enable
- Shared datasets that are not solely "owned" by EPA (i.e. some EPA programs are jointly operated with one or more other Federal agencies, some environmental data are collected by states, tribes, and other entities and provided to EPA)
- "Orphan" datasets (Some EPA data, while potentially valuable, do not have a clear owner who can take responsibility for making the data accessible)
- LOD (An approach to making open data connectable, currently in use on the Data.gov.UK site)

These initiatives and issues will be further addressed in SDAP V2.0, as appropriate.



IV. Data Standards, Metadata, Quality and Security

A. Data Standards

Data standards are documented consensus-based agreements on the format and definition of common data. The Federal government requires use of consensus-based standards approved by international and national standards groups whenever they are available and appropriate.

The Agency has adopted some existing consensus standards and created additional standards with our state and tribal partners. EPA data standards are a means to promote the efficient sharing of environmental information among EPA, states, tribes, local governments, the private sector, and other information-trading partners. By using data standards, the public can more quickly and accurately conduct environmental assessments and analyze environmental data, maximize use of resources, and be assured of greater data integrity. The Agency's data element registry will maintain dataset data dictionaries and show how datasets map to data standards.

B. Metadata

As more and more data resources are made available, it is crucial that the data are easily understood by both subject matter experts and the general public. To increase the usability of data resources, best practices encourage creating good metadata. Technology advancements also can help improve the usability of information through providing data in LOD form.

Metadata describing the data asset being published is critical to help the public and other stakeholders understand the meaning and potential value of the Agency's data. A complete data dictionary is fundamental to help our stakeholders better understand our data.

EPA's current set of registries are depicted in Appendix C. Each registry was built to improve and enable internal and external discovery and understanding of EPA's data resources. EPA is working to minimize duplicate data entry for all these metadata, by pulling data from service to service so that the data are entered once and used many times. This effort should improve consistency between metadata stores and simplify actions required by data owners.

C. Quality

EPA is fully committed to providing the highest quality data available for consumption. Over the years, the Agency has built an infrastructure to support shared services, open data standards, and web services that improve data quality. All data resources accessed directly from any public site should be from authoritative sources that meet pre-existing statutory mandates and Agency policy for information dissemination. Data disseminated to the public must adhere to all Information Quality Act requirements. [EPA's Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by the Environmental Protection Agency](#) creates a mechanism that enables the public to seek and obtain, where appropriate, correction of information disseminated by EPA that does not comply with OMB's or EPA's Information Quality Guidelines. More information on Agency guidelines for ensuring quality can be found at [our website](#).

Open Government Data Quality Plan

The OGD also required the completion of a financial data quality plan to ensure that all spending data made available to the public is of the highest integrity. The Open Government Data Quality Plan built upon the Agency's management integrity experience to provide a framework for ensuring the quality of EPA spending information. It specifically assesses the risks to Agency data submitted to USASpending.gov, and recommends improvements as necessary.



D. Data Security and Privacy

Data shared publicly will conform to all applicable security and privacy requirements including the Privacy Act of 1974, the E-Government Act of 2002, applicable Federal security standards including NIST 800-39, and other privacy and confidentiality guidance as issued by OMB. For information bound for Data.gov, the Agency will adhere to the Federal privacy policy (<http://www.data.gov/privacypolicy>).



Appendix A: SDAP Project Team

The SDAP Project Team includes representatives from the following offices and regions across the Agency:

- Regions 2 and 3
- Office of Air and Radiation (OAR)
- Office of Administration and Resource Management (OARM)
- Office of the Chief Financial Officer (OCFO)
- Office of Enforcement and Compliance Assurance (OECA)
- Office of Environmental Information (OEI)
- Office of Research and Development (ORD)
- Office of Chemical Safety and Pollution Prevention (OCSPP)
- Office of Solid Waste and Emergency Response (OSWER)
- Office of Water (OW)



Appendix B: Select, Prepare, and Add a Dataset to Data.gov

EPA continues to increase the information available to the public on Data.gov. As of Quarter 1 in Fiscal Year (FY) 2011, EPA made over 1,800 datasets/tools/geodata available to the public via Data.gov. Information about these and other EPA data are accessible on the updated Data.gov inventory at <http://www.epa.gov/open/data-gov-inventory2011Q1.xls>.

This appendix is geared to EPA staff and guides them in the process of bringing information to the public via Data.gov.

A. Data.gov and Enterprise Data Resource Terms

OpenGov encourages agencies to provide access to high-value data via Data.gov, a one-stop shop for federal data. Data.gov categorizes data holdings into three categories – datasets, tools, and geodata. The definition of these categories and the common Agency data resources terms associated with each are provided below:

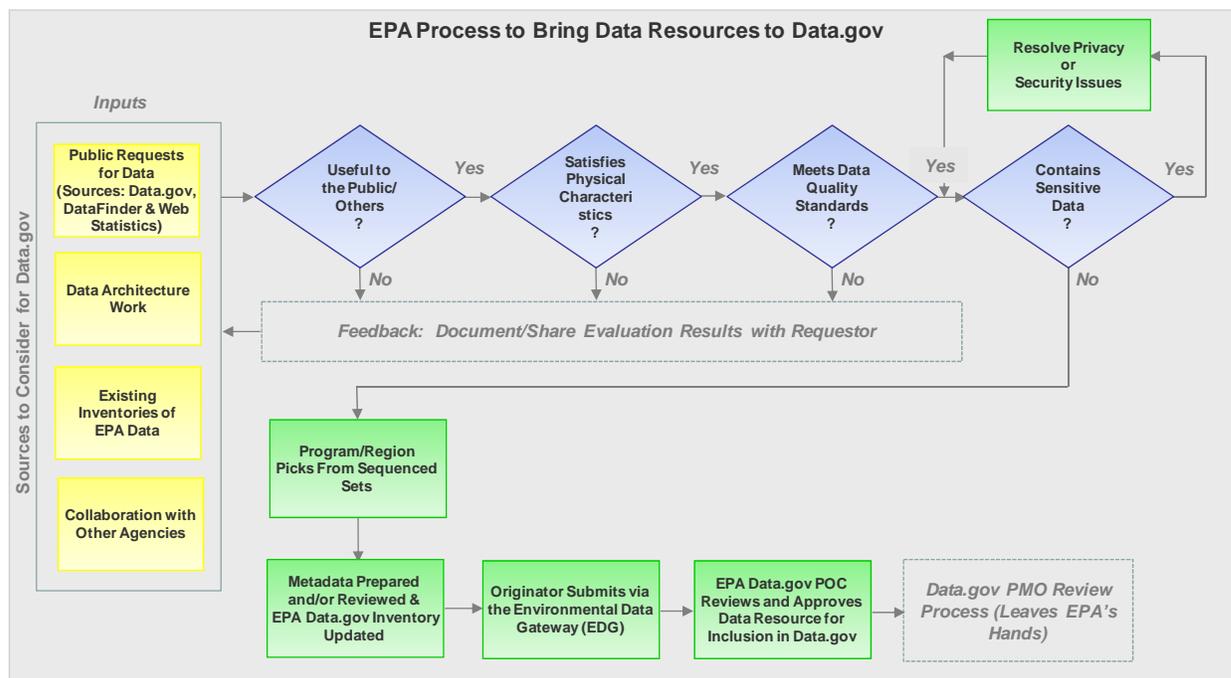
- Datasets – Data available as a single click, machine readable, platform-independent download
 - EPA data resources include: database, dataset
- Tools – Hyperlinks that lead to tools or web pages that allow users to extract datasets
 - EPA data tools include: applications, databases, web services, APIs, and select web pages
- Geodata – Any type of data that is referenced to a location on the Earth's surface
 - EPA geodata include: geodatasets, geodata tools, and geodata models



B. Data.gov Process

There are a number of steps involved in considering, selecting, and processing a dataset/tool/geodata for publication to Data.gov. Most of these steps are required for publication of any data or information made available to the public. The steps are presented in Figure 2 and described in the remainder of this section.

Figure 2: EPA's Data.gov Process



C. Evaluate Public Usefulness

EPA is focused on providing data that is useful to the public, businesses, other federal agencies and state/local governments. In determining what data to publish, programs and regions should review and evaluate the following sources to determine the datasets to consider:

- Interactions with or requests from community or stakeholder groups
- Data.gov suggestions
- IdeaScale or other OpenGov email suggestions
- Web statistics

Other factors to consider in evaluating public usefulness can include information that:

- Supports Agency goals (i.e. Protecting America's Waters)
- Furthers analysis of and conversation about environmental topics in the news (i.e. Greenhouse Gas Emissions)



D. Confirm Characteristics of the Data

To be included in Data.gov, a dataset, tool or geodata should have the following physical characteristics:

- Reside on a publicly accessible platform
- Have unrestricted access - no user identification (ID) required for access
- Be in machine-readable form (preferably in an open format)

E. Ensure Quality of the Data

EPA's standing data quality procedures and policies apply for all EPA data resources destined for the public. See Section IV, C of this document for information on data quality standards.

F. Protect Sensitive Data

In order to protect the privacy and security of the public, businesses, and US Government staff and operations, some types of data may be deemed sensitive and will not be made public or published on Data.gov. Each Agency and the Data.gov PMO must review each dataset against a checklist of restricted types of data including Privacy/Confidentiality (both personal and of a business nature) concerns and National/Homeland security concerns (i.e. Federal continuity of operations/continuity of government (COOP/COG) plans).

G. Best Practices for Data.gov Publication

EPA encourages programs and regions to take a thoughtful approach to publish their data publicly. In this section, we describe how to approach the following categories of data population:

- Data that already exists in Data.gov
 - For dataset owners who have already published their data via Data.gov, EPA recommends that the owner(s) of the dataset(s) provide automatic updates (if appropriate) or periodically check the accuracy and currency of information in the datasets. If your Data.gov Technical POC gets feedback from the public requesting the data in a different format, please evaluate and consider providing the data in the requested format.
- Data that is publicly accessible and not catalogued in Data.gov
 - If you have a dataset that is publicly-accessible through a web application or link with descriptive data, but it is not currently accessible via Data.gov, you should catalog the dataset as a tool. Cataloging the data as a tool helps the public discover the information through an additional source – Data.gov.
- High-Value data not yet catalogued in Data.gov
 - For datasets that are high-value, publicly-requested, or may be of value to other agencies or non-government stakeholders, consider transforming the dataset into an open format.

H. Open Formats

Within EPA, we are recommending a progressive approach to publishing information on Data.gov and providing the information in open formats. Below are five open formats, and a data resource becomes more usable inside and outside EPA the farther you get down the list of formats. A dataset earns an EPA data check (√) for each way you provide the information. A maximum of five data checks (√√√√√) is possible for a data resource. At the current time, programs should focus their attention on the first four formats.



1. Post/Catalog the Data Resource in Data.gov

Making the data available in any form on Data.gov earns the data resource one check (√).

2. Post the Dataset as a CSV File

A comma-separated values or character-separated values (CSV) file format is a simple text format for a database table. Most software and tools enable you to save a file as a CSV, and the data resource earns an additional check (√) if the data is provided in CSV.

3. If the Data Exists in XML, Post the Existing XML

Extensible Markup Language (XML) is a set of rules for encoding documents in machine-readable form. If you already have the data in XML format, post the file you have and earn an additional data check (√).

4. Post and/or Create Application Programming Interfaces (APIs)

Application Programming Interfaces (APIs) enable a software program to access and make use of the services and resources provided by another software program. If you have APIs or can easily develop them, please consider making them available via Data.gov thereby earning an additional data check (√).

5. Post the Dataset in Linked Open Data (LOD) Form

EPA is in the early stages of exploring how to best provide data resources in LOD form such as RDF so that EPA data is “findable,” can be pointed to, and can be more easily mashed with other EPA or non-EPA data. Data resources that provide information in LOD form earn an additional data check (√).

I. Data.gov Roles and Responsibilities

There are several categories of staff both within EPA and the General Services Administration (GSA) involved in bringing a dataset/tool/geodata to Data.gov. Within EPA, the Originator and Approver reside in the program/regional office, and the EPA Data.gov POC is in the Office of Environmental Information (OEI). The roles and responsibilities of each category of participant in the process of preparing and approving a dataset are included in this section. In general, questions regarding the roles and responsibilities and procedures for the Data.gov process should be directed to the EPA Data.gov POC (see below):

▪ Originator (Dataset Owner or Designee within the Program/Region)

The Originator is responsible for preparing a dataset and associated metadata for Data.gov. The Originator will enter the information into the EDG (or other source system that is appropriate for the data resource to be shared via Data.gov). Originators may be supported by contractors or program office coordinators who prepare the metadata and provide instructions/guidance, but the Originator needs to be an EPA employee. Within the Program/Region the individual charged with this responsibility may vary, so please consult your IMO or designee for additional guidance.

- Contact the EDG Data.gov POC (Michelle Torreano torreano.michelle@epa.gov) for access.

▪ Approver from Program Office/Region Staff

Each office may have their own process to release datasets to Data.gov. For example, data quality and appropriateness concerns should be reviewed and addressed prior to proposing a dataset be considered or processed for Data.gov. We recommend the following approval/authorization/sign off chain:

- Quality review satisfying Agency and Program/Regional requirements
- Sign off from the Program/Regional IMO
- Authorization from a Division Director or above who certifies the data is ready to be published

EPA Strategic Data Action Plan (SDAP) Version 1.0 - DRAFT



▪ EPA Data.gov POC

The EPA Data.gov POC is a representative from EPA appointed by the Chief Information Officer (CIO) who reviews EDG submissions drafted by Originators and transmits them to the Data.gov PMO for inclusion in Data.gov. Agency Data.gov POCs are also responsible for ensuring that datasets submitted meet the data policy standards for Data.gov. EPA's Data.gov POC's are as follows:

- EPA Data.gov POC: Steve Young (young.steve@epa.gov)
- Backup: Zach Scott (scott.zachary@epa.gov)

▪ Data.gov PMO (GSA)

The Program Management Office/Manager (PMO) is a group of individuals at the government-wide and agency level who collectively manage Data.gov. Members of the PMO validate approved Agency submissions from the POCs.

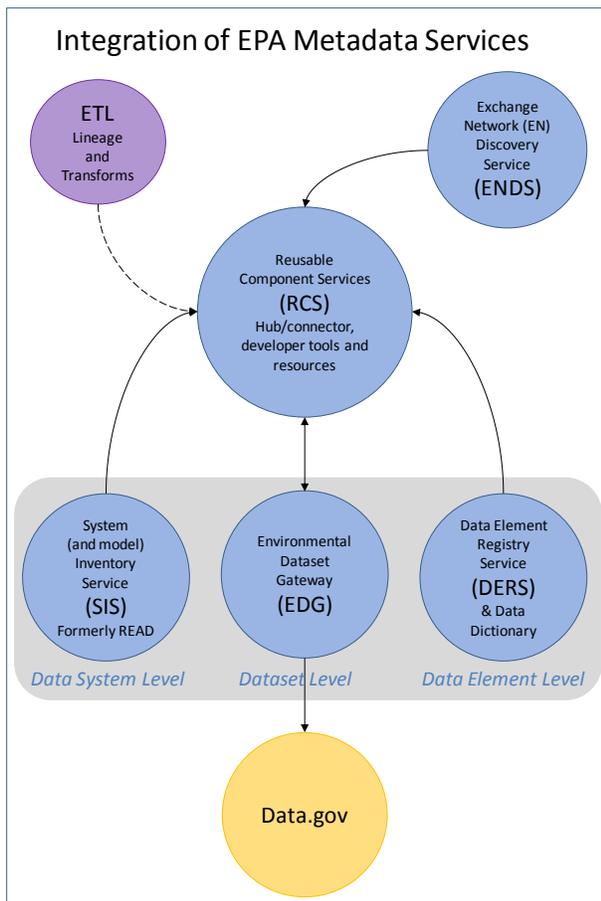


Appendix C: Metadata Services Integration

The diagram below shows the interrelationships between EPA's various metadata system components and shows at a high level how the components will feed into Data.gov. At the center, the Reusable Component Services (RCS) is the hub connection for developer tools and other resources.

Below is a brief listing and description of EPA's relevant data registries. For more information on EPA's System of Registries, see http://iaspub.epa.gov/sor_internet/registry/sysofreg/home/overview/home.do.

Figure 3: EPA Metadata Services - Planned Integration



- Each dataset that will be made available to the public, such as those reported in Data.gov, should be registered in the Environmental Dataset Gateway (EDG) (www.epa.gov/edg). As EPA publishes more tools like APIs, web services, etc., EDG can also become the mechanism for pulling the metadata for those components from the systems where they are stored (e.g., RCS mentioned below) and transferring the metadata for those components to Data.gov. The EDG will replace the Data Management System (DMS), which currently is used to upload non geo-dataset information to Data.gov.
- If the dataset is derived from an EPA system or model, there should be a record for that system or model in EPA's System and model Inventory System (SIS) (formerly known as the Registry for EPA Applications and Databases (READ) (<http://www.epa.gov/read>).
- The dataset should have a data dictionary that is registered in the Data Element Registry Services (DERS) to enable understanding of the data fields in the dataset.
- If there is a web service or other tool that can be used with the dataset it should be registered in the Reusable Component Services (RCS) (<http://www.epa.gov/rcs>).

- EPA is considering tools that can automatically pull metadata from datasets hosted at EPA.



Appendix D: Website Links to Resources

- President Obama's January 2009 Memorandum: <http://www.whitehouse.gov/the-press-office/transparency-and-open-government>
- Presidential Memorandum of January 18, 2011, on Regulatory Compliance: <http://www.gpo.gov/fdsys/pkg/FR-2011-01-21/pdf/2011-1386.pdf>
- Open Government Directive (OGD): http://www.whitehouse.gov/sites/default/files/omb/assets/memoranda_2010/m10-06.pdf
- EPA Open Government Plan: http://www.epa.gov/open/EPAOpenGovernmentPlan_11.pdf
- Interim EPA Data.gov Guidance and Procedures, December 21, 2010: http://www.epa.gov/open/EPA_DataGov_Guidance_Procedures_2010_12_21.pdf
- Draft US EPA SDAP Annotated Outline, December 2, 2010: http://www.epa.gov/open/Draft_EPA_SDAP_Outline_12-02-10_rev.pdf
- EPA Open Government Data Quality Plan V 1.0, May 18, 2010: http://www.epa.gov/open/opengov_data_quality_plan.pdf
- EPA's System of Registries: http://iaspub.epa.gov/sor_internet/registry/sysofreg/home/overview/home.do
- EPA's Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by the Environmental Protection Agency: http://www.epa.gov/QUALITY/informationguidelines/documents/EPA_InfoQualityGuidelines.pdf
- Federal privacy policy: <http://www.data.gov/privacypolicy>
- EPA's System Inventory Services (SIS) or Registry of EPA Applications and Databases (READ): http://iaspub.epa.gov/sor_internet/registry/systemreg/home/overview/home.do
- Reusable Component Services (RCS): <http://www.epa.gov/rcs>
- Environmental Dataset Gateway: <http://geogateway.epa.gov>
- Quarterly Inventory Spreadsheet: <http://www.epa.gov/open/data-gov-inventory2011Q1.xls>
- EPA's OpenGov IdeaScale: <http://openepa.ideascale.com/a/panel.do>
- EPA's Data Finder Forum: <http://blog.epa.gov/data/>