

US Environmental Protection Agency (EPA)

# Interim EPA Data.gov Guidance and Procedures

12/21/2010



## Table of Contents

---

Table of Contents .....	2
I. Background.....	4
A. Approach to Guidance and Procedures .....	4
B. Goals.....	5
C. Guiding Principles .....	5
II. Tiered Approach to Populate Data.gov .....	6
A. Data Already in Data.gov – Complete.....	6
B. Publically-Accessible Datasets Not Cataloged in Data.gov – Catalog as a Tool .....	6
C. High-Value or Publically-Requested Datasets Not Cataloged in Data.gov – Consider Adding to Data.gov in Open Format .....	6
1. High-Value Data Defined.....	6
2. Open Format Approach .....	6
III. Formal Process to Evaluate/Add a Dataset to Data.gov.....	7
A. Useful to the Public/Others.....	8
B. Satisfies Physical Characteristics .....	8
C. Meets Data Quality Standards.....	8
D. Does Not Contain Sensitive Data .....	8
IV. Preparing Data for Data.gov.....	9
A. Open Formats .....	9
1. Post the dataset as a CSV file .....	9
2. If the Data Exists in XML, Post the Existing XML.....	9
3. Post the dataset in XML using data standards.....	9
4. Post the dataset in Linked Open Data (LOD) form.....	9
B. Application Programming Interfaces (APIs) .....	10
C. Metadata .....	10
V. Paths into Data.gov: DMS and GDG .....	11
A. Data Management System (DMS) .....	11
B. GeoData Gateway (GDG).....	11
VI. Roles and Responsibilities .....	12
A. Originator (Dataset Owner or Designee within the Program/Region) .....	12
B. Approver from Program Office/Region Staff.....	12



C.	EPA Data.gov POC .....	12
D.	Data.gov PMO (GSA) .....	13
VII.	Parking Lot.....	13
Appendix A:	EPA’s DMS Guidance.....	14
A.	General Guidance .....	14
B.	Welcome to the DMS.....	14
C.	Login .....	15
D.	DMS Home Page .....	16
E.	Tab 1: Profile.....	17
F.	Tab 2: Agency .....	19
G.	Tab 3: Coverage.....	20
H.	Tab 4: Data Quality.....	21
I.	Tab 5: Data/Tool.....	22
J.	Tab 6: Stat Resources (If Applicable) .....	24
K.	Tab 7: OGD.....	25
Appendix B:	OSWER’s Data.gov Checklist.....	26

DRAFT



Data.gov (<http://www.data.gov>), one component of Open Government (OpenGov), is essentially a one-stop shop for government data. The Environmental Protection Agency (EPA) is holistically working to integrate data management for Data.gov into Agency-wide data management practices and policies as part of our overall Strategic Data Action Plan (SDAP).

As the larger effort proceeds, this *Interim EPA Data.gov Guidance and Procedures* document provides current procedures to bring EPA data to Data.gov, including some recommended improvements. The intent of this artifact is to guide dataset owners, managers, program/regional Information Management Officers (IMOs), Data.gov Points of Contact (POCs), and others in selecting and publishing datasets/tools/geodata to Data.gov.

## I. Background

---

The White House OpenGov initiative is detailed in President Obama's January 2009 memorandum (<http://www.whitehouse.gov/the-press-office/transparency-and-open-government>) and the December 2009 Office of Management and Budget (OMB) Open Government Directive ([http://www.whitehouse.gov/sites/default/files/omb/assets/memoranda\\_2010/m10-06.pdf](http://www.whitehouse.gov/sites/default/files/omb/assets/memoranda_2010/m10-06.pdf)).

As noted in EPA's OpenGov Plan, accessible from EPA's Open website (<http://www.epa.gov/open/>), we envision a standardized and streamlined process of selecting and posting high-value datasets and tools to Data.gov for public access. Our goal is to simplify the process of bringing data to Data.gov, while ultimately increasing the quantity, quality and usability of EPA data in a minimally burdensome manner for EPA offices and regions.

### A. Approach to Guidance and Procedures

We are approaching guidance and procedures for Data.gov in an iterative manner with distinct phases. First we are capturing current and recommended processes with some improvements within this document. Using feedback from users and other stakeholders, we will mature the processes and document them in the SDAP v1.0. The SDAP will be updated as needed until the procedures are incorporated into Agency-wide data policies and procedures that address all EPA data sources.

This product will guide parties within EPA that are part of the process, documenting their roles and responsibilities to evaluate, prioritize and process data that will be made available to the public via Data.gov. Our approach is to leverage EPA's previous work where appropriate, including institutional knowledge from EPA's Data.gov team and checklists and procedures developed by various offices within EPA. As previously noted, this product and modifications to procedures will be incorporated into the SDAP v1.0 (and future versions) until the SDAP Project Team's work is fully transitioned into Enterprise data policy and procedures.

#### Purpose of this document

- Provide instructional information on the steps in the process to get datasets/tools/geodata published to Data.gov
- Identify and explore areas of the process that need development or improvement (to be addressed or resolved in the SDAP v1.0 or a future version)
- Pilot new approaches and address recommendations to improve standard practices.



## B. Goals

EPA wants to improve the experience of the public and other stakeholders in finding data via Data.gov to support informed decisions, provide the appropriate level of breadth and depth of information, and streamline and simplify the process for Agency staff. We have identified the following goals to help us in this quest:

- Gather feedback from the public and other stakeholders via IdeaScale
- Formalize the process to select, prepare and publish information on Data.gov, helping to thoughtfully expand our catalog
- Identify any gaps in information we currently provide or plan to provide via Data.gov
- Streamline and simplify metadata capture, minimizing duplicative processes and ultimately improving the usability of our Data.gov catalog.

This Interim guidance is a key initial step in meeting our goals.

## C. Guiding Principles

EPA reviewed and analyzed a collection of data management and data publishing principles from internal and external sources with the goal of setting clear and effective guiding principles that can feed into the forthcoming Enterprise data policy. We are using the Interim document as a vehicle to publish an initial draft of guiding principles, as a means to seek feedback and collaborate with internal and external stakeholders in further developing the principles.

1. **Provide Easy and Timely Access to Quality Data.** Public data will be available, timely, and easy to find through a single, easy to use, online access point (Data.gov) and in keeping with established data quality objectives.
2. **Publish Data in Open Form.** Public data will be published using open, structured, computer readable form following open-based standards. Examples include CSV, XML, KML, RSS, and RDF.
3. **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.
4. **Apply Data Governance.** Data releases will be made within the construct of EPA's data governance. Existing data governance includes the Quality Information Council (QIC) and the QIC's Exchange Network Subcommittee. Programs and data stewards will maintain their responsibility to uphold Agency/Federal guidelines and applicable laws and/or regulations.
5. **Enhance Data Use.** Provide data services, i.e. web services, and tools, i.e. APIs, to enhance public use of data for published data sets. In addition, data should be reported at the lowest analytical unit possible for ease of consumption by the public.
6. **Drive Best Uses/Best Practices.** Lessons learned and best practices should be continually assessed so that the best current technology for high quality data sharing is communicated and implemented across the Agency.
7. **Grow and Improve Through User Feedback.** Apply feedback from the public to help set priorities and identify high value data sets to publish.

Please direct any comments and feedback to Zach Scott at [scott.zachary@epa.gov](mailto:scott.zachary@epa.gov).



## II. Tiered Approach to Populate Data.gov

Programs and regions should take a tiered approach to getting their data into Data.gov to continue populating Data.gov while making the best use of limited resources. In this section we address guidance for programs and regions to review their internal data inventories and take recommended action based on the current status of the dataset or tool.

### A. Data Already in Data.gov – Complete

If your dataset/tool/geodata is already in Data.gov, there is no need at this time to go back and take additional action based upon this guidance. 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. Recommendations for formats are included in Section IV.A Open Formats.

### B. Publically-Accessible Datasets Not Cataloged in Data.gov – Catalog as a Tool

If you have a dataset that is publically-accessible as a site or link with descriptive data but it is not currently included in Data.gov, catalog the dataset as a Tool. You do not need to put the data into an open format as a result of this guidance. Cataloging the data as a tool helps the public discover the information through an additional source – Data.gov.

### C. High-Value or Publically-Requested Datasets Not Catalogued in Data.gov – Consider Adding to Data.gov in Open Format

For datasets that are high-value, publically-requested, or may be of value to other agencies or non-government stakeholders, consider transforming the dataset into open format.

#### 1. High-Value Data Defined

High-value data is defined in the Open Government Directive as data that:

- a) Increases Agency Accountability/Responsiveness
- b) Improves Public Knowledge of Agency/Operations
- c) Furthers Agency Core Mission
- d) Creates Economic Opportunity
- e) Responds to Need/Demand from Public Consultation

#### 2. Open Format Approach

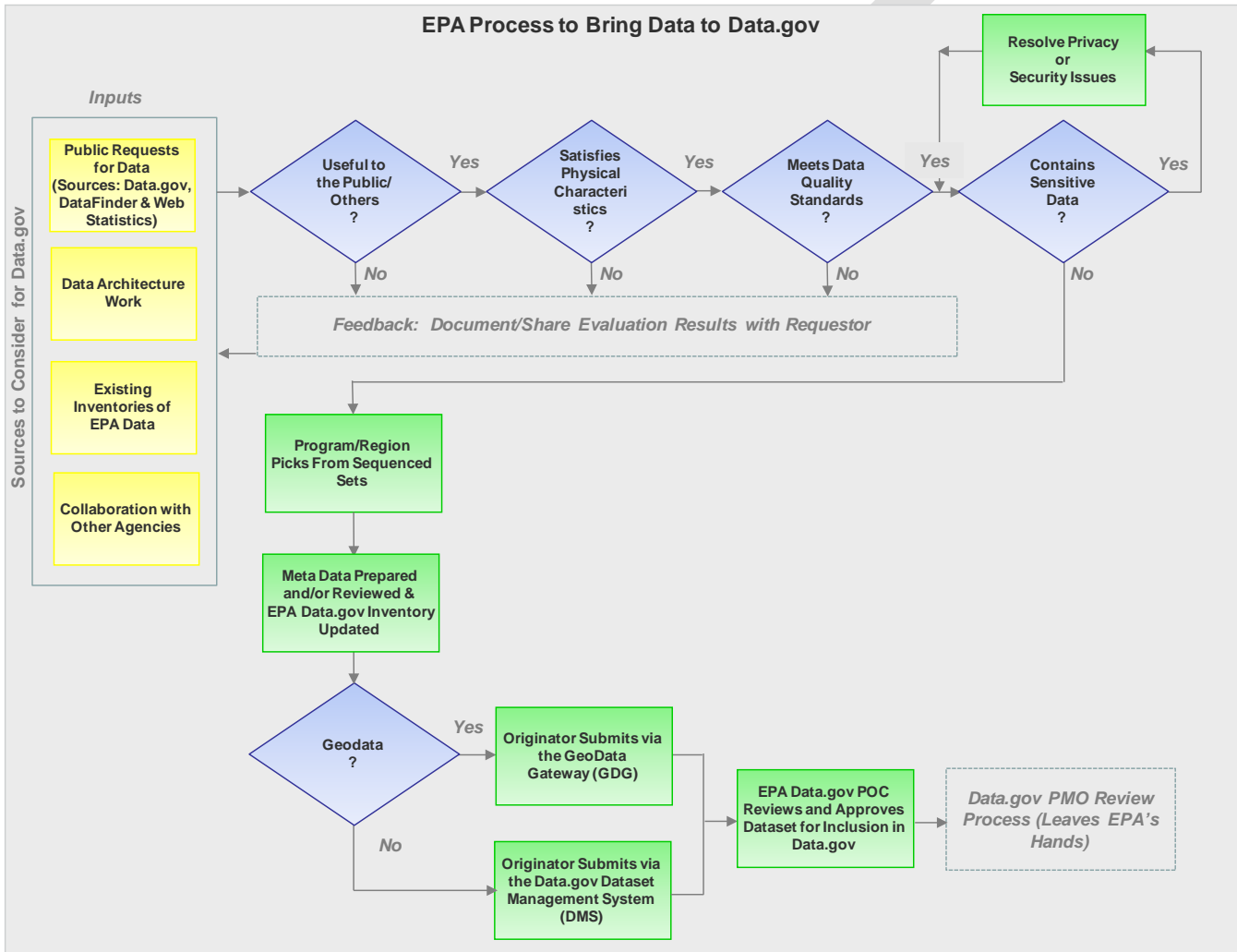
Within EPA, we are recommending a stepped approach to putting information into open format, with the information becoming more usable outside EPA as you progress. Please refer to Section IV.A Open Formats for more information.





### III. Formal Process to Evaluate/Add a Dataset to Data.gov

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 any publication of data or information made available to the public. The diagram below provides a snapshot of the process, with additional detail of the roles and responsibilities of individuals presented in the remainder of this document.



**Figure 1: EPA's Data.gov Process**

Additional detail surrounding the process depicted in Figure 1 is addressed below, in links, and/or as appendices to this document.

EPA has a wide array of information in the form of datasets, tools and GeoData that may be of interest and use to the public. We established these guidelines to help data stewards and other information specialists evaluate and determine the best candidates for Data.gov.



## A. Useful to the Public/Others

EPA is focused on providing data that is useful to the public, businesses, other federal agencies and state/local governments. In determining which data to publish, programs and regions should review and evaluate several 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 stats

## B. Satisfies Physical Characteristics

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

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

## C. Meets Data Quality Standards

EPA's standing data quality procedures and policies apply for data destined for Data.gov. EPA's data quality guidance can be found at: [http://www.epa.gov/quality/qa\\_docs.html](http://www.epa.gov/quality/qa_docs.html).

## D. Does Not Contain 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 published by the Data.gov PMO. 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). We will include a link to the Data.gov guidance once it is complete.





## IV. Preparing Data for Data.gov

### A. Open Formats

Within EPA, we are recommending a stepped approach to putting information into open format, with the information becoming more usable inside and outside EPA the farther you get down the list. At the current time, programs should focus their attention on the first three.

#### 1. Post the dataset as a CSV file

A comma-separated values or character-separated values (CSV) file is a simple text format for a database table. Most software and tools enable you to save as a CSV.

#### 2. 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.

#### 3. Post the dataset in XML using data standards

The standards are provided in hierarchical order with the highest first. If there is a conflict in standards, the standard listed higher in the list supersedes the other.

##### a) EPA's Data Standards (Highest)

[http://iaspub.epa.gov/sor\\_internet/registry/datastds/finddatastandard/epaapproved/](http://iaspub.epa.gov/sor_internet/registry/datastds/finddatastandard/epaapproved/)

##### b) Federal Geographic Data Committee (FGDC) Standards

<http://www.fgdc.gov/standards>

##### c) National Information Exchange Model (NIEM)

<http://it.ojp.gov/default.aspx?area=nationalInitiatives&page=1072>

##### d) Superfund's Staged Electronic Data Deliverable (SEDD) Standards

The Superfund Program's Staged Electronic Data Deliverable (SEDD) standards can be found at <http://www.epa.gov/superfund/programs/clp/sedd-geninfo.htm>.

Additional environmental data standards are being matured as part of the National Environmental Information Exchange Network <http://www.exchangenetwork.net/standards/index.htm>.

#### 4. Post the dataset in Linked Open Data (LOD) form

EPA is currently in the process of putting two key registries in Linked Open Data (LOD) form to enable data mashups of multiple EPA datasets as well as externally with other datasets. Additional guidance and information is forthcoming.



## B. 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 particular software program. If you have APIs or can easily develop them, please consider making them available via Data.gov.

## C. Metadata

To fully embrace OpenGov objectives, EPA must provide the public with adequate information about the published dataset/tool/Geodata. Data about the data is referred to as metadata and is critical to help the public and other stakeholders understand EPA's data. We established the following metadata guidelines for datasets destined for Data.gov:

1. All datasets must be registered in EPA's data set registry ([http://iaspub.epa.gov/sor\\_internet/registry/datareg/home/overview/home.do](http://iaspub.epa.gov/sor_internet/registry/datareg/home/overview/home.do)).
2. If the dataset is derived from an EPA system, there must be a record for that system in EPA's Registry for EPA Applications, (Models) and Databases (READ) ([http://iaspub.epa.gov/sor\\_internet/registry/systmreg/home/overview/home.do](http://iaspub.epa.gov/sor_internet/registry/systmreg/home/overview/home.do)).
3. If the dataset is derived from a model, there must be a record for that model in READ ([http://iaspub.epa.gov/sor\\_internet/registry/systmreg/home/overview/home.do](http://iaspub.epa.gov/sor_internet/registry/systmreg/home/overview/home.do)).

Additional services and resources that will be released soon as part of the System of Registries (SoR) will provide additional metadata both within EPA and to our stakeholders.

- Datasets should have a data dictionary and that data dictionary should be registered in the Data Registry for centralized access by the public ([http://iaspub.epa.gov/sor\\_internet/registry/datareg/searchandretrieve/datadictionaries/](http://iaspub.epa.gov/sor_internet/registry/datareg/searchandretrieve/datadictionaries/)). The staff from the Office of Environmental Information (OEI) is available to help programs and regions develop data dictionaries. If you need assistance, please contact John Harman at [harman.john@epa.gov](mailto:harman.john@epa.gov).
- If there is a web service or other tool that can be used with the data set, that web service or tool should be registered in the Reusable Component Services (RCS) ([http://iaspub.epa.gov/sor\\_internet/registry/reusereg/home/overview/home.do](http://iaspub.epa.gov/sor_internet/registry/reusereg/home/overview/home.do))
- In RCS, all associations should be identified between a dataset (harvested from Dataset Registry), the system or model it comes from (harvested from READ), and web services or other tools that can be used with the dataset (registered in RCS).



## V. Paths into Data.gov: DMS and GDG

As depicted in Figure 1, there are two paths to take in preparing and submitting a dataset/tool/geodata for Data.gov. The first path is via the Data.gov Dataset Management System (DMS), developed by the Data.gov Program Management Office (PMO) at the General Services Administration (GSA) to process datasets across the government. The second path is via EPA's GeoData Gateway (GDG), which is used to process geodata datasets.

### A. Data Management System (DMS)

The DMS should be used to process datasets for inclusion in Data.gov that are non-geodata. Users can access the DMS and obtain descriptive and background information at [dms.data.gov](http://dms.data.gov).

There are three defined user functions involved in the process to publish a dataset to Data.gov using the DMS.

- Originator – Typically the dataset owner within EPA
- Agency POC – EPA's Data.gov POC staff who are responsible for Data.gov management within EPA
- PMO/Manager – Data.gov PMO (GSA) staff who are responsible for Data.gov management across the Federal government.

Roles and responsibilities for these user groups are defined in the next section.

EPA employees who need to enter, process, or manage datasets via the DMS need a DMS account and must have the EPA password to request an account. Please contact the EPA Data.gov POC or backup (provided below in Section VI) to obtain the password.

Although users can find Government-wide guidance to the DMS at [www.dms.data.gov](http://www.dms.data.gov), there are some required elements that are of minimal relevance to EPA. Please use EPA's DMS guidance, included as Appendix A, for EPA-specific guidance including values and/or examples for all required elements.

### B. GeoData Gateway (GDG)

Currently the GDG is used to process geodata sets.

Geospatial data is any type of information that is referenced to a location on the Earth's surface. This may include (but is not limited to) data that contains spatial coordinates (e.g., latitude and longitude or X and Y), address information, gridded data, georeferenced imagery, or other content that is referenced to a location on the Earth's surface (e.g., hydrologic networks). Geospatial data may be stored in many formats, including simple spreadsheets, text files or CSV files (with latitude and longitude or X/Y fields), shapefiles, KML/KMZ files, geospatial databases, GeoRSS feeds, or geospatial services. Any data that has a locational aspect and can readily be shown on a map should be considered and classified as geospatial data.

Guidance for submitting geospatial data can be found at [https://geogateway.epa.gov/geoportal/webhelp/en/geoportal\\_931/index.htm#wkg\\_login.htm](https://geogateway.epa.gov/geoportal/webhelp/en/geoportal_931/index.htm#wkg_login.htm).

- GDG EPA POC: Michelle Torreano ([torreano.michelle@epa.gov](mailto:torreano.michelle@epa.gov)).



## VI. Roles and Responsibilities

---

There are several categories of staff both within EPA and the General Services Administration (GSA), which serves as the Data.gov PMO, 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.

Some program offices have developed checklists and internal procedures, and we included one developed by the Office of Solid Waste and Emergency Response (OSWER) as an example. See Appendix B.

In general questions regarding the roles and responsibilities and procedures for the Data.gov process should be directed to:

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

### A. 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 DMS or the GDG and is typically the dataset owner. Within the Program/Region the individual charged with this responsibility may vary, so please consult your Information Management Officer (IMO) or designee for additional guidance.

As previously stated, originators using the DMS must first establish a DMS account. Originators should contact the EPA Data.gov POC or backup (provided above) to obtain the password. Originators using the GDG should contact the GDG Data.gov POC (provided above) to obtain access.

### B. 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:

1. Quality review (Satisfying Agency and Program/Regional requirements)
2. Sign off from the Program/Regional IMO
3. Authorization from a Division Director or above who certifies the data is ready to be published.

### C. EPA Data.gov POC

The Agency POC is a representative from a federal agency appointed by their Chief Information Officer (CIO) who reviews DMS records (or GDG submissions) drafted by Originators and transmits them to the Data.gov PMO for inclusion in Data.gov. Agency POCs are also responsible for ensuring datasets submitted meet the data policy standards for Data.gov.



### D. Data.gov PMO (GSA)

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

## VII. Parking Lot

---

At EPA, we are approaching our work in an iterative fashion. As we develop guidance and evolve, issues will arise that require synchronization with other efforts or more time to fully craft an approach. We want to track such items and inform our stakeholders as to the issues. This approach may also help our internal and external stakeholders identify and raise other items for consideration.

To date, we are aware of and will address or pass along to the appropriate office the following:

- Some datasets/tools/geodata do not have an identified data owner; they are referred to as “orphan data.” EPA needs to address how to assign responsibility for the data.
- There are some potentially high-value data that are not in a publically-accessible environment; the SDAP v1.0 should address this issue and possibly provide guidance and support to data stewards or owners in moving their data to a web-accessible location.
- Some agencies and stakeholders want a full inventory of all data/information the agency maintains, regardless of whether the information can be made public. We are in the process of establishing EPA's position.
- Data originators are responsible for ensuring that data is accurate and current, and each office's IMO is responsible for establishing procedures for internal approval of such updates. When updates are needed, the office IMO must coordinate with the Agency's Data.gov POC.
- Some datasets maybe be co-created by two or more agencies, making it a challenge in cataloging the contributor of the data within Data.gov.





## Appendix A: EPA's DMS Guidance

---

### A. General Guidance

The intent of this guidance is to serve as a point of reference for EPA users when submitting EPA datasets through the Data.gov Dataset Management System (DMS). As previously stated in Section V, there are three major roles involved with publishing a dataset using the DMS. This guidance is focused on assisting Originators in preparing their dataset and associated metadata for Data.gov. For detailed guidance on overall functionality provided by the DMS, please refer to the Dataset Management System User Guide v3.0 ([https://dms.data.gov/dms/webroot/dms\\_user\\_guide%5B3%5D.pdf](https://dms.data.gov/dms/webroot/dms_user_guide%5B3%5D.pdf)).

The following functionality should be kept in mind while completing the data submission process outlined in this guidance:

- View Helpful Hints: Users can hover their cursor over the “i” icon provided next to field headings to view descriptions or “helpful hints” for each field.
- The DMS identifies required fields with a red asterisk.
- Save: It is recommended that the Originator saves their record as they complete each step outlined in the following sections.
- Clone: Cloning the dataset allows the Originator to clone the information provided to reduce duplicative data entry for other similar datasets.
- Return to List: This button returns the Originator to their DMS Home page. Updated information will be lost if the record is not saved before returning to list.
- Upload Metadata: The Upload Metadata button is provided on the DMS Home page. This functionality allows Originators to upload large bulk uploads. Originators should contact their POC for additional guidance on this functionality.

### B. Welcome to the DMS

The DMS Welcome page is displayed in the image below. This page is accessible from the following URL: <https://dms.data.gov/dms/webroot/flex/datagov.html>. EPA users should select the Originator role.





## DATA.GOV

### Welcome to the Data.gov Dataset Management System!

The Dataset Management System (DMS) provides automated manageable process for publishing datasets into the Data.gov catalog by facilitating agencies' efforts to organize and maintain their Data.gov submissions.

#### What's New

1. Metadata template version 2.0 is available for download.



#### Resources:

[Latest Metadata Template](#)

Please select from the following which best describes your role in Data.gov submission process.

#### Originator

The DMS allows Originators to submit new datasets and review the status of previously submitted datasets. Once a dataset suggestion has been added to the DMS, its status can be tracked through the submission lifecycle using the dashboard.

If you do not already have a DMS account, all you need to register is your .gov or .mil e-mail address and your agency's password. Contact your Agency POC to obtain an agency password.

Originator

#### Agency Point of Contact

Agency point of contacts (POCs) can access the DMS to view the entire published catalog, all published datasets and tools submitted by their agency, and a dashboard to track all pending submissions. The DMS is built with the flexibility to allow agency POCs to edit, delete, approve, or return records to the originator.

Agency POC must log in using the Office of Management and Budget's online collaboration system, MAX Community.

Agency POC

#### Program Management Office

The Data.gov Program Management Office (PMO) can access the DMS to view a dashboard illustrating pending submissions, all published records to date, and all validated but not published records. The DMS will allow the PMO to validate records, return records to a POC, and edit, delete, or change published records.

PMO must also log in using the Office of Management and Budget's online collaboration system, MAX Community.

PMO/Manager

[DATA.GOV](#) [Data Policy](#) [Privacy Policy](#) [FAQ's](#) [Glossary](#)

version 1.1.4a

## C. Login

The DMS requires EPA users to have a registered DMS account before being able to access the provided Data.gov dataset tools. New users can register for an account by entering their EPA e-mail address and agency password in the Login page below (Contact Steve Young or Zach Scott to request the EPA agency password).

Upon completing registration, users can log into the DMS using their EPA e-mail address and registered DMS password.

# Interim EPA Data.gov Guidance and Procedures



## DATA.GOV

### Welcome dataset originators!

If you are the Originator and already have a DMS account, please enter your e-mail address and password below to login. If you are a new user, please enter your .gov or .mil email address, and your agency's password to create an account.

Username (e-mail address):

 This field is required.

Password:

 This field is required.

Login

Cancel

Forgot Password

## D. DMS Home Page

The DMS Home page is provided below. Selecting "Add New Record" will initiate the new dataset submission process.

## DATA.GOV @ EPA

[Home](#) [Change Password](#) [What's New](#) [Logout](#)

### Welcome to Data.gov Dataset Management System (DMS)!

Below is a summary and a listing of all records associated with your submissions.

Detailed Instructions are as follows: [Filter List](#) [Add New Record](#) [Upload Metadata](#) [Update Published Records](#)

Draft	SubAgency Review	Agency Review	PMO Review	NSC Review	QC Review	Published	TOTAL
2	0	0	0	0	0	0	2

#### Resources:

- [References](#)
- [DMS User's Guide](#)
- [Latest Metadata Template](#)
- [Agency, Sub-Agency POCs](#)
- [Documents Repository](#)
- [DMS - Report a Bug or Suggest Improvement](#)

**Title and Description** Filter by Status:  Order by:   
Filter by Record Type:  Descending:

Submitted By: EPA; Last Updated on: 2010-11-24 12:41:00

Copy of

Submitted By: EPA; Last Updated on: 2010-11-24 12:52:43

[Add New Record](#) [Upload MetaData](#)

Records 1-2 of 2



## E. Tab 1: Profile

The first step of the dataset submission process is to provide the dataset profile information (see image below).

The screenshot shows the EPA Data.gov submission interface. At the top, there is a navigation bar with 'DATA.GOV @ EPA' and links for 'Home', 'Change Password', 'What's New', and 'Logout'. Below this is a progress bar indicating the current status as 'Draft' and showing steps 1 through 7. Step 1, '1. Profile', is highlighted with a red circle. The form contains several fields: 'User Generated ID' (with a tooltip explaining it as an unambiguous reference), 'Title \*' (with a tooltip explaining it should identify the resource by ISBN, ISSN, URL/URI, etc.), 'Dataset Group Name', 'Description \*', 'Data.gov Catalog Type \*' (set to 'Raw Data Catalog'), 'Specialized Data Category Designation \*' (set to 'Administrative'), 'Subject Area (Taxonomy) \*' (set to 'Section 1. Population'), 'Keywords \*', 'Bibliographic Citation \*', 'Date Released \*', and 'Date Updated \*'. A red asterisk indicates required fields. At the bottom, there are buttons for 'Clone this Record', 'Return to List', and 'Save'.

We established the following guidelines to provide further assistance for completing the displayed Profile fields:

- User Generated ID: Please disregard this field.
- Title: This field should identify the title that will be displayed for the dataset in Data.gov. This should be kept short and descriptive. Please avoid using abbreviations, acronyms, and including EPA in the title (acronyms can be utilized later as keywords).

Example: *New Private Housing Building Permits in 1997*

- Dataset Group Name: Please disregard this field.
- Description: This field should identify the contents of the dataset. This field is important in providing details to Data.gov users on the information captured by the dataset.

Example: *“Ambient ozone concentrations for 2008 from the national ambient air quality monitoring networks stored in the Air Quality System (AQS).”* (Dataset: Air Quality System (AQS) ambient observations: 2008 ozone)



- **Data.gov Catalog Type:** This field should identify whether the dataset should be catalogued as Raw Data or a Tool. Datasets should be identified as raw data if the data file links directly to the raw metadata. EPA defines raw data as “single click downloadable data”. Datasets should be identified as a tool if the data links to a tool landing page.

Raw Data Example : *AirNow Real Time Air Quality KML File*

Tool Example: *ENERGY STAR Find a Product*

- **Specialized Data Category Designation:** EPA datasets should almost always be identified as Administrative. Datasets should only be identified as Statistical if they contain population statistics data (e.g. population means, standard deviations, variances, etc).
- **Subject Area Taxonomy:** This field should identify the subject area in which the dataset falls into. EPA datasets are almost always identified as “6. Geography and Environment” however any can be selected. It is also possible to designate a distinct subject area by providing it in the “Other” text field.
- **Keywords:** This field should identify the important keywords that should be used when searching for the dataset in Data.gov. When providing information for this field, please keep in mind that there are multiple Data.gov stakeholders. It is important to include terms aimed at the lay public in addition to those aimed at scientific academics. We suggest including acronyms/abbreviations in addition to their definition.
- **Bibliographic Citation:** This field should identify how the dataset should be cited in other references. Please provide both the English text and URL for the citation. We recommend the following citation format: “EPA, Name of Office that owns the dataset, Dataset Name, <URL>.”
- **Date Released:** This field should identify when the dataset was first released to the public. The date should follow the following format: mm/dd/yyyy.
- **Date Updated:** This field should identify when the dataset was last updated. We suggest identifying this field as “Continuously as new documents are issued” for dynamic datasets. If the dataset has never been updated, enter “N/A” into this field. The date should follow the following format: mm/dd/yyyy.
- **Comments from Agency and PMO Reviews:** Please disregard this field.



## F. Tab 2: Agency

The second step of the dataset submission process is to provide the dataset agency information (see image below).

The screenshot shows the '2. Agency' step of the submission process. The progress bar at the top indicates the current status is 'Draft' and the current step is 'Step 2'. The form fields are as follows:

- Agency \***: Environmental Protection Agency (EPA)
- Sub-Agency**: N/A
- Agency Program URL \***: (Empty text box)
- Agency Data Series URL \***: (Empty text box)
- Dataset/Tool Contact (Not Public) \***:
  - Contact Name**: (Empty text box)
  - Phone Number**: (Format: ###-###-####)
  - Email Address**: (Empty text box)

At the bottom of the form, there are buttons for 'Clone this Record', 'Return to List', and 'Save'. A red asterisk indicates required fields.

We established the following guidelines to provide further assistance for completing the displayed Agency fields:

- **Agency:** This field should identify the owning government agency. EPA datasets should always be identified as “Environmental Protection Agency (EPA)”.
- **Sub Agency:** Please disregard this field.
- **Agency Program URL:** This field should identify the name of the Program and corresponding URL.  
Example: *AirNow*, *http://airnow.gov/*
- **Agency Data Series URL.** This field should identify the dataset landing page. Both the landing page name and URL should be provided.  
Example: *Clean Air Status and Trends Network (CASTNET):Ozone*,  
*http://www.epa.gov/castnet/data.html*
- **Dataset/Tool Contact (Not Public):** This field should identify the main contact for the dataset. This contact should have familiarity with the data contained in the dataset. Although the contact information will not be released to the general public, feed back provided by the public will be directed to this individual.



## G. Tab 3: Coverage

The third step of the dataset submission process is to provide the dataset coverage information (see image below).

The screenshot shows the EPA Data.gov submission interface. At the top, there is a navigation bar with 'DATA.GOV @ EPA' and links for 'Home', 'Change Password', 'What's New', and 'Logout'. Below this is a progress bar showing the current status as 'Draft' and the current step as 'Step 3'. The '3. Coverage' tab is selected and highlighted. The form contains the following fields:

- Unit of Analysis \*
- Geographic Scope \*
- Period of Coverage \*
- Geographic Granularity
- Frequency \*

A red asterisk indicates that the fields marked with an asterisk are required. At the bottom of the form, there are buttons for 'Clone this Record', 'Return to List', and 'Save'.

We established the following guidelines to provide further assistance for completing the displayed Coverage fields:

- **Unit of Analysis:** This field should identify the measuring unit that is represented in each record of the dataset.  
Examples: Facility, Chemical, Vehicle, etc
- **Period of Coverage:** This field should identify the time period captured by the dataset. This should be kept short and simple. If the dataset captures a specific year then provide the year. If the dataset does not then enter "N/A".
- **Frequency:** This field should identify the frequency in which the data is captured.  
Example: Annual, Bi-Annually, Monthly, Daily, Once, etc
- **Geographic Scope:** This field should identify the geographic coverage captured by the dataset. If the dataset is nationwide, enter "United States". Datasets related to Toxicity should be identified as "N/A".
- **Geographic Granularity:** This field should identify further details on the geography captured by the dataset.  
Example: Latitude/Longitude, City within a Zip Code, By State within the U.S., etc





## H. Tab 4: Data Quality

The fourth step of the dataset submission process is to provide data quality information for the dataset (see image below).

The screenshot shows the EPA Data.gov submission interface. At the top, there are navigation links: Home, Change Password, What's New, and Logout. Below that is the 'DATA.GOV @ EPA' logo. A progress bar indicates the current status is 'Draft' and shows steps 1 through 7. Step 4, 'Data Quality', is highlighted with a red circle. Below the progress bar are seven tabs: 1. Profile, 2. Agency, 3. Coverage, 4. Data Quality (highlighted), 5. Data/Tool, 6. Stat Resources, and 7. OGD. The '4. Data Quality' tab contains several form fields:

- Reference for Technical Documentation \* (required field)
- Additional Metadata (info icon)
- Collection Mode (info icon)
- Data Collection Instrument (info icon)
- Data dictionary/variable list (info icon)
- Agency responsible for Information Quality \* (required field)
- Compliance with Agency's Information Quality Guidelines \* (required field) with radio buttons for Yes, No, and Not Relevant.
- Privacy and Confidentiality \* (required field) with radio buttons for Yes, No, and Not Relevant.

A red asterisk at the bottom right of the form indicates that fields marked with an asterisk are required. At the bottom of the form, there are three buttons: 'Clone this Record', 'Return to List', and 'Save'.

We established the following guidelines to provide further assistance for completing the displayed Data Quality fields:

- Reference for Technical Documentation: This field should identify an additional reference where Data.gov users can view further information on the dataset. We suggest providing a URL describing the Program who owns the dataset.
- Additional Metadata: This field should identify the URL (if any) that points to additional dataset metadata.
- Collection Mode: Please disregard this field.
- Data Collection Instrument: This field should identify the URL that points to the collection instrument (if any) used for the dataset.
- Data Dictionary/Variable List: This field should identify the Data Dictionary (if available) for the dataset.
- Agency Responsible for Information Quality: This field should identify the government agency that the dataset is in compliance with. EPA dataset should always be identified as "United States Environmental Protection Agency (EPA)".



- **Compliance with Agency’s Information Quality Guidelines:** This field should identify whether or not the dataset meets the EPA’s Information Quality Guidelines. This field should always be identified as “Yes” otherwise the dataset is not appropriate for publication on Data.gov.
- **Privacy and Confidentiality:** This field should identify whether or not the dataset meets the EPA’s responsibilities under the Privacy Act. This field should always be identified as “Yes” otherwise the dataset is not appropriate for publication on Data.gov.

## I. Tab 5: Data/Tool

The fifth step of the dataset submission process is to provide data/tool information (see image below).

**DATA.GOV @ EPA** Home Change Password What's New Logout

Current Status: Draft | SubAgency Review | Agency Review | PMO Review | NSC Review | QC Review | Published

Step 1 | Step 2 | Step 3 | Step 4 | **Step 5** | Step 6 | Step 7

1. Profile | 2. Agency | 3. Coverage | 4. Data Quality | **5. Data/Tool** | 6. Stat Resources | 7. OGD

Create | Edit | Delete

#	Access Point	Media Format	File Size (MB)	File Format

Number of Datasets Represented by this Submission <sup>?</sup>

Dataset use requires a license agreement \* <sup>?</sup>  
 ▾

Dataset license agreement URL <sup>?</sup>

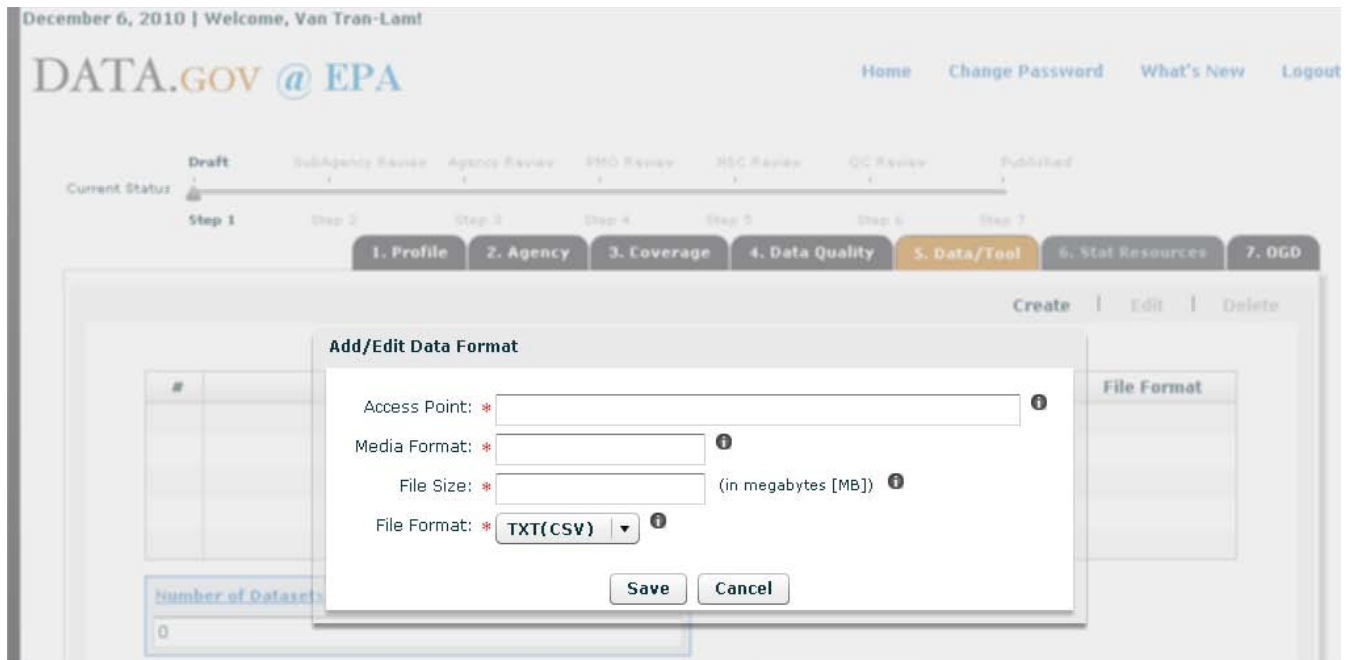
Clone this Record | Return to List | Save

We established the following guidelines to provide further assistance for completing the displayed Data/Tool fields:

- **Number of Datasets Represented by this Submission:** Please disregard this field.
- **Data set use requires a license agreement:** This field should identify whether or not the dataset required a license agreement. This field should be identified as “No” for the majority of EPA datasets.
- **Dataset license agreement URL:** This field should identify the URL for the license agreement if the user selects “Yes” in the previous field.



To provide information for the #, Access Point, Media Format, File Size (MB) and File Format, Originators must click the “Create” link above the table.



We established the following guidelines to provide further assistance when adding/editing the data format seen in the image above:

- **Access Point:** This field should identify the URL for immediate access to the downloadable dataset. In the case where the dataset is a tool, this field should identify the URL for the tool's landing page.

Raw Data Example: *Green Vehicle Guide Data Download:*  
[http://www.epa.gov/greenvehicles/download/all\\_alpha\\_10.txt](http://www.epa.gov/greenvehicles/download/all_alpha_10.txt)

Tool Example: *Energy Star Find a Product:*  
[http://www.energystar.gov/index.cfm?fuseaction=find\\_a\\_product](http://www.energystar.gov/index.cfm?fuseaction=find_a_product).

- **Media Format:** This field is used to identify the downloadable media type for the data contained in the dataset. It is possible for datasets to have multiple media formats. If this is the case, please enter a separate row in the table for each media format.

Example: PDF, XML, CSV, etc

- **File Size:** This field should identify the size of the downloadable dataset file. In the case that the file size varies depending on selected values, enter “Varies” into this field. In the case that the file size can be identified with a specific number value, please enter the number only.
- **File Format:** This field should identify the file format in which the dataset can be downloaded. The options displayed in the dropdown are based on whether the dataset has been identified as raw data or a tool. In the case that the file format is a Zip file, please identify the file format of the Zip file contents.

# Interim EPA Data.gov Guidance and Procedures



## J. Tab 6: Stat Resources (If Applicable)

The Stat Resources tab seen below is only accessible if the dataset is identified as “Statistical” during selection of the Specialized Data Category Designation in step 1. As previously stated, datasets should only be identified as “Statistical” if they contain population statistics such as standard deviations, variances, means, survey data, etc. Originators should contact their POC if they believe their dataset is in fact a statistical resource.

The screenshot shows the EPA Data.gov interface. At the top, the logo 'DATA.GOV @ EPA' is on the left, and navigation links 'Home', 'Change Password', 'What's New', and 'Logout' are on the right. Below the logo is a progress bar with 'Current Status' and a dropdown menu. The progress bar shows steps from 'Step 1' to 'Step 7'. Below the progress bar is a row of tabs: '1. Profile', '2. Agency', '3. Coverage', '4. Data Quality', '5. Data/Tool', '6. Stat Resources', and '7. OGD'. The '6. Stat Resources' tab is highlighted with a red circle. Below the tabs is a form with several sections, each with a title and a text input field: 'Statistical Methodology', 'Sampling', 'Questionnaire Design', 'Estimation', 'Series Breaks', 'Weighting', 'Non-Response Adjustment', 'Disclosure Avoidance', 'Seasonal Adjustment', and 'Data Quality'. A red asterisk at the bottom right of the form indicates '\* denotes required field'. At the bottom of the form are three buttons: 'Clone this Record', 'Submit for Approval', and 'Return to List', followed by a 'Save' button.



## K. Tab 7: OGD

The final step of the dataset submission process is to provide the dataset OGD information (see image below).

The screenshot shows the EPA Data.gov submission interface. At the top, it says "DATA.GOV @ EPA" and has navigation links for "Home", "Change Password", "What's New", and "Logout". Below this is a progress bar with steps: Draft, SubAgency Review, Agency Review, PMO Review, NSC Review, QC Review, and Published. Underneath, there are seven numbered steps: 1. Profile, 2. Agency, 3. Coverage, 4. Data Quality, 5. Data/Tool, 6. Stat Resources, and 7. OGD. Step 7 is highlighted with a red circle. The main form area contains several questions with dropdown menus and text input fields:

- Is this record a part of OGD submissions?** (Dropdown menu with "No" selected)
- Is this record a high-value data?** (Dropdown menu with "No" selected)
- Is this record listed in your agency open government plan?** (Dropdown menu with "No" selected)
- What makes this a high value dataset?** (Dropdown menu with "Increases agency accountability and responsiveness" selected)
- How is this new?** (Text input field)

At the bottom of the form, there are buttons for "Clone this Record", "Submit for Approval", "Return to List", and "Save".

We established the following guidelines to provide further assistance for completing the displayed OGD fields:

- Is this record a part of OGD submissions?: This field should be identified as “No”.
- Is this record listed in your agency open government plan?: This field should be identified as “No”.
- Is this record a high-value data?: This field should be identified as “Yes”
- What makes this a high-value data set?: This field should identify why the data contained in the dataset is considered high value data. Originators should select the best answer based on the selections provided. If possible, please do not select “Other”.
- How is this new?: This field should identify how the dataset is new. We suggest identifying the dataset as not previously catalogued in Data.gov or as new to the web.

The Originator has now reached the end of their dataset submission process. At this point, the Originator can submit their dataset for approval by clicking the “Submit for Approval” button. Originators should only click this button if they are sure their provided information is ready for POC review and approval. Datasets should not be submitted to the POC if they have not been approved by the Division Director or higher management.



## Appendix B: OSWER's Data.gov Checklist

CHECKLIST FOR PUBLISHING DATASETS ON DATA.GOV			
NAME OF DATASET(s): _____		<b>KEY</b>	
DATASET OWNER: _____		Publishing	<b>P</b>
ORIGINATOR: _____		Discovery	<b>D</b>
		Feedback	<b>F</b>
This checklist guides you through the process of publishing datasets on Data.gov, which only involves 3 overarching steps:			
<ol style="list-style-type: none"> <li>Determine the data you want to publish</li> <li>Load the metadata into a spreadsheet</li> <li>Upload to Data.gov</li> </ol>			
Process	Status	Notes/Comments	
<b>STEP 1: Determine the data you want to publish</b>			
P1		Determine if data set or tool is already listed as a high-value data on EPA's Open Government Data Inventory. The latest version is currently available at: <a href="http://www.epa.gov/open/Data_Set_Inventory0625.pdf">http://www.epa.gov/open/Data_Set_Inventory0625.pdf</a>	
P2		Determine if additional sets/tools may be valuable to the public. Track the data sets and tools with NO public access restrictions (from the previous steps).	
P2.1		Research if your data set or tool has public access restrictions in READ. Look under the column titled "Data Accessible to the General Public?" Verify that the value is "Yes" for your data set.	TIP: READ generates a report titled READXMLExport.xls, which we have available for you upon request.
P2.2		Verify that no restrictions to public access are specified. Additional narrative information regarding data restrictions may be available in CPIC, CPIC Lite, and/or Exhibit 53s, in the fields titled 'I. Justification' and '2. Who are the customers, stakeholders, partners and/or collaborators of this project?'	Upon request, we can check CPICs, CPIC Lites, and Exhibit 53s for your data sets/tools.
P2.3		Create a list called 'Recommended List of Data Sets/Tools for publishing to Data.gov,' and populate it with the names of data sets and tools with NO public access restrictions (from the previous steps).	
P2.4		Meet with your management to discuss approval to publish each data set/tool to Data.gov on your 'Recommended List of Data Sets/Tools for publishing to Data.gov.' Next to each data set/tool, track whether Mgt Approval was 'Yes,' 'No,' or 'Needs Additional Research or Stakeholder Approval.'	
P2.5		For 'Needs Additional Research or Stakeholder Approval,' conduct the appropriate additional research or stakeholder meetings, and update 'Mgt Approval' to Yes or No as appropriate.	
P2.6		If your management has identified concerns to explore regarding Personal Privacy, research EPA's related data privacy policies at: <a href="http://www.epa.gov/privacy/policy/2151/index.htm">http://www.epa.gov/privacy/policy/2151/index.htm</a>	
P2.7		Discuss any Security Concerns with management and security stakeholders as appropriate, to ensure that any security concerns that arise are resolved.	
P3		Gather and document the URL by which Data.gov's DMS can access the tool or raw data set	
<b>Administrative Step: Prepare for Publishing Datasets</b>			
Process	Status	Notes/Comments	
P4		List the data sets/tools approved by management for publishing to Data.gov as your next Publishing Release.	
P5		Prepare metadata spreadsheet to upload to DMS.data.gov. (Originator uploads metadata file with data sets/tools, Data.gov PMO handles remainder of the process)	
P6		Ensure approvals were obtained for all people that need to approve publishing your data to Data.gov.	
P7		Obtain an active Data.gov DMS login by visiting <a href="https://dms.data.gov/dms/webroot/flex/datagov.html">https://dms.data.gov/dms/webroot/flex/datagov.html</a> , selecting 'Originator,' and inputting your .gov or .mil email address with a password.	
P8		Sign up for Data.gov DMS training by notifying your Agency PMO Lead of the need for training. Steve Young serves as EPA's Agency PMO for Data.gov.	



# Interim EPA Data.gov Guidance and Procedures



<b>STEP 2: Load the Metadata into a Spreadsheet</b>			
	<b>Process</b>	<b>Status</b>	<b>Notes/Comments</b>
P9	Download the current version of Data.gov's metadata template at <a href="https://dms.data.gov/dms/webroot/flex/datagov.html">https://dms.data.gov/dms/webroot/flex/datagov.html</a> .		The current metadata template version is: data_gov_metadata_template_v2.xls)
P10	Populate the metadata fields in the Data.gov metadata template. Some of the fields may be populated with metadata in READ (revisit step 1.1 for obtaining READ information). Some of the fields may be populated with information from CPIC, CPIC Lite, and Exhibit 53 information.		FYI: We have an OSWER Data Catalog available that has already collected metadata from these sources, along with an instructions document that details how the information was collected. Our help inputting data and/or resources are available upon request).
P11	For publishing raw data sets, ensure that your data is available in one of the formats supported by Data.gov, which include: XML, CSV/Text, KML/KMZ Shapefile, RSS, XLS. Document the supported formats of your dataset.		Presenting data in standardized formats enhances its usability and improves discovery
P12	Verify that the URL to access your data set or tool is listed in the metadata record.		
<b>STEP 3: Upload data onto Data.gov</b>			
	<b>Process</b>	<b>Status</b>	<b>Notes/Comments</b>
P13	Ensure all mandatory fields have been populated in the Data.gov metadata template		The Data.gov metadata template specifies mandatory vs. optional fields.
P14	Log into dms.data.gov, select originator, and upload the metadata template containing your data sets/tools.		
P15	Log into DMS to monitor status on whether your data set/tool has been published, as determined by the Agency PMO Lead.		

DRAFT



## CHECKLIST FOR IMPROVING DISCOVERY FOR DATA PUBLISHED ON DATA.GOV

NAME OF DATASET(s): _____		KEY	Code
DATASET OWNER: _____		Discovery	D
ORIGINATOR: _____			
Process	Status	Notes/Comments	
D1		Inventory all description documents that describe structure, relationships, and intended usage of your data in order to improve discovery. In the metadata record, include URLs to all reference documentation.	
D2		Ensure data is in a Machine Readable format and provide URL/URLs.	
D3		Include any content that accompanies the data set or tool. At a minimum, each data set needs an accompanying Data Dictionary and Description Document. The URLs to the additional documents must be filled out in the metadata template that is uploaded to DMS.data.gov.	
D4		Indicate any data standards to which your data set or tool adheres.	
D5		Indicate any taxonomies to which your data set or tool conforms.	
D6		Indicate any ontologies domains to which your data set or tool conforms.	
D7		Specify URL/URI's for any Linked Open data contained in your set/tool in models / descriptive documents that accompany your data set/tools.	
		Identify any models that describe relationship of linked open data to other data concepts in structure.	
		Identify RDF/XML resources available to accompany your data set/tool.	
D9		List any applicable standards in the metadata record	
D10		Prepare data dictionary for publishing (ensuring at least X fields are completed)	
D12		Integrate metadata with APIs on Data.gov (if applicable)	
D13		Integrate with Search Widget Boxes on Data.gov (if applicable)	
D14		Enable Advanced Search for XML, RDF, etc. on DMS.data.gov	
D15		Enable Advanced Query for Geographic Searching on DMS.data.gov	

## CHECKLIST FOR HANDLING FEEDBACK FOR DATA PUBLISHED ON DATA.GOV

NAME OF DATASET(s): _____		KEY	Code
DATASET OWNER: _____		Feedback	F
ORIGINATOR: _____			
Process	Status	Notes/Comments	
F1		Collect, monitor, and review public feedback on Data.gov's dialog site, which can be accessed directly at <a href="http://datagov.ideascale.com/">http://datagov.ideascale.com/</a> or by selecting the dialogue tab on Data.gov.	
F2		Work with Agency PMO Lead when public requests a data set, tool or resource from your area at <a href="http://www.data.gov/suggestdataset">http://www.data.gov/suggestdataset</a> .	
F3		If you are monitoring public dialog on a site other than Ideascale, enter the address of the mechanism being used to monitor	
F4		Track action items associated with your data set/tool in public feedback.	
F5		Assign individual(s) to handle action items associated with feedback	
F6		Categorize action items with low, medium and high levels of difficulty / urgency.	
F7		Discuss action items with related stakeholders until action item can be closed.	
F8		Post response as a comment below the public comment on Ideascale and/or other site as appropriate.	
F9		When posting data assets in response to a public request, post a comment beneath the public request to announce when the data asset will be published.	

# Interim EPA Data.gov Guidance and Procedures



2.0 DataGov Metadata Template for Datasets and Tools				
Element Number	Element	Occurrence [min,max]	Description	Enter Metadata for Data Single Dataset (Use Other Worksheet for Multiple Dataset Loading)
1	Unique ID	[1,1]	For DataGov PMO Use Only - will be generated when placed in catalog	FOR DATAGOV PMO USE ONLY
1.1	User Generated ID	[0,1]	This field may be used by the agency to track the submissions on their internal systems. The field is optional as it will not typically be used by the Datagov and will not be published on the catalog.	(READ identifier)
2	Title	[1,1]	Unique name of the dataset or Tool. (e.g., Current Population Survey, Consumer Price Index, FBI Ten Most Wanted Widget). This field will be used to populate the data catalog; the catalog will be sorted on this field. Note: if the title is not unique within the entire data.gov catalog, you will be asked to change it.	
3	Dataset Group Name	[0,1]	This is an optional field. This field allows agencies to provide a Group Name to multiple datasets and/or tools in order to show that they may be presented as a group or a set.	
4	Description	[1,1]	Please provide a detailed description of the dataset or tool (e.g., an abstract) such that the user would be able to determine the nature and purpose of the data.	
5	Agency Name	[1,1]	Department or Independent Agency name.	Environmental Protection Agency
5.1	Agency Short Name	[1,1]	Acronym or short name corresponding to the Agency name (e.g. DOC, DoD, NASA, GSA).	EPA
6	Sub-Agency Name	[1,1]	Bureau or Sub-Agency or operating unit. name.	
6.1	Sub-Agency Short Name	[1,1]	Acronym or short name corresponding to the Sub-Agency name (e.g. IRS, FBI, BIA ).	
7	Contact Name	[1,1]	Contact person's name (first, then last) where questions from the Data.gov PMO should be sent. <b>** Note that these "contact" elements (7, 7.1 &amp; 7.2) will not be published on the Data.gov website but may receive public email comments made specifically about this dataset or tool.**</b>	
7.1	Contact Phone Number	[1,1]	Contact person's phone number.	
7.2	Contact Email Address	[1,1]	Contact person's email address.	
8	Agency responsible for Information Quality	[1,1]	Enter the Agency name corresponding to the applicable Information Quality Guidelines for the dataset. In some cases, this may be a parent organization, such as a Department.	
8.1	Compliance with Agency's Information Quality Guidelines	[1,1]	Confirm that the dataset meets the Agency's (as identified in Element 6) Information Quality Guidelines (Yes, No). If the dataset is not in compliance with your Agency's Information Quality Guidelines it will not be posted on Data.gov.	
8.2	Privacy and Confidentiality	[1,1]	Confirm that dissemination of all data submitted is consistent with the agency's responsibilities under the Privacy Act and, if applicable, CIPSEA or other relevant statute. Enter "Yes" or "No." or "Not Relevant." If the answer is "No", this dataset will not be posted on Data.gov.	
9	Data.gov catalog type	[1,1]	Select the appropriate catalog on Data.gov that this submission should be listed (i.e. Raw Data Catalog or Tool Catalog)	
10	Subject area (Taxonomy)	[1,1]	Please choose the category from the drop-down menu that best describes your dataset. If more than one category applies, choose the category that you think most people would use. We realize that this taxonomy, which is based on the Statistical Abstract, is not perfect for all datasets. If the item being described is a 'data mining and/or extraction tool', then please provide the category that describes the underlying dataset.	(example 6. Geography and Environment )
11	Specialized data category designation	[1,1]	Identify the type of dataset (i.e., administrative, statistical, geospatial, surveillance or research). Please choose the category that best fits the dataset, understanding that there is some potential for overlap among these categories. Note that some types of data have additional metadata requirements (e.g., see statistical datasets). If the item being described is a 'data mining and/or extraction tool,' then please provide the designation that describes the underlying dataset from the pre-populated dropdown list. Note: If you select Geospatial, you should submit your dataset to <a href="http://geodata.gov">http://geodata.gov</a> and you do not need to fill out any additional information for this submission. We would, however request that you submit the elements 1 through 7 to <a href="mailto:datagov.submit@gsa.gov">datagov.submit@gsa.gov</a> so that we can synchronize our efforts with geodata.gov. If you select "statistical", please note the section "SM" that follows.	

# Interim EPA Data.gov Guidance and Procedures



12	<b>Keywords</b>	[1,n]	Searchable keywords help users discover your datasets from different perspectives. They also provide ways of identifying other datasets that are similar to yours. Please include terms that would be used by both technical and non-technical users. If the item being described is a 'data mining and/or extraction tool,' then please provide keywords that describe the underlying dataset. Agencies are encouraged to include as many keywords as possible. Please use commas to separate keywords.
13	<b>Date released</b>	[1,1]	Date when the dataset was first made available to the public. This date should not be confused with when the data is being entered into Data.gov as it could have already been published on your website.
14	<b>Date updated</b>	[1,1]	Date of last change to dataset or tool. Note that this could be the same as the date released if the data has not changed since first being published. For example, data could have been released on 06/03/2006 and published to your Agency website, but later found to contain data errors. If the dataset was corrected in July 2006, the updated date would be whenever this corrected update was applied to the dataset (e.g. 07/12/2006).
15	<b>Agency Program URL</b>	[1,1]	URL that is closest to the program that is responsible for this dataset or tool.
16	<b>Agency Data Series URL</b>	[1,1]	URL to the high level repository in which this dataset or tool resides at the agency (e.g., the HTML page that displays the link to the dataset).
17	<b>Collection mode</b>	[1,n]	Data collection mode (e.g., phone/paper, phone/computer, person/paper, person/computer, web, fax, type of monitor, other). Multiple collection modes should be separated by commas.
18	<b>Frequency</b>	[1,1]	Frequency of data collection (one-time, annual, hourly, etc.).
19	<b>Period of Coverage</b>	[1,n]	Dates or time interval(s) covered by the data. Please use commas to separate multiple periods of Coverage.
20	<b>Unit of analysis</b>	[1,1]	The unit of analysis is the major entity that you are analyzing in your study (e.g., person, household, forest, county, establishment). If the item being described is a 'data mining and/or extraction tool,' then please provide the category that describes the underlying dataset.
21	<b>Geographic scope</b>	[0,n]	Please indicate the geographical extent covered by this dataset. In the case of multiple locations, please delimit with commas. Some datasets are not earth-based, thus will not use this field and should simply leave it blank. You may use commas to separate multiple entries.
21.1	<b>Geographic Granularity</b>	[0,1]	This is an optional field. Please indicate the most detailed level at which the geography is defined (e.g. City vs. zip code vs. longitude/latitude pair).
22	<b>Reference for Technical Documentation</b>	[1,n]	URL or bibliographic citation for the technical documentation for this dataset. This may include description to the study design, instrumentation, implementation, limitations, and appropriate use of the dataset or tool. In the case of multiple documents or URLs, please delimit with commas.
23	<b>Data dictionary/variable list</b>	[1,1]	URL to resource containing variable names, descriptions, standard vocabularies and taxonomies, units, multipliers, etc. May be identical to element 22.
24	<b>Data collection instrument</b>	[1,n]	URL for resource containing a copy of, or detailed descriptions of, the data collection instrument for each listed mode. May be identical to element 22. Multiple URLs should be separated by commas.
25	<b>Bibliographic citation for dataset</b>	[1,1]	This field may be used when others make reference to the data, as in a bibliographic citation or source reference. If the agency does not have a standard reference for this dataset, simply provide the URL for the dataset.
26	<b>Number of Datasets Represented by this Submission</b>	[1,1]	If this submission is a compressed file, data extraction tool or mining tool, please enter the total number of datasets represented by this submission. ** Please note that this field is now required whereas with prior to versions of the metadata template it was optional. **
27	<b>Additional Metadata</b>	[0,1]	This is an optional field. Please provide a URL to any additional metadata for the dataset or tool.
28	<b>Dataset use requires a license agreement</b>	[1,1]	This is a required field to ensure that license agreements are not bypassed during the one-click download interface on the website.
29	<b>Dataset license agreement URL</b>	[1,1]	URL to the license agreement page for the dataset or tool. This is a required field if Element 28 above is answered yes. This field is conditionally required. If Element 28 above is "Yes", please provide the URL to the dataset license agreement.

# Interim EPA Data.gov Guidance and Procedures



Downloadable file specific metadata (repeat this section for multiple download formats using multiple columns in the following rows.)			
D1	Access point	[1,1]	If the data set is downloadable, enter the URL for instant access to the downloadable data file. This is the URL for access to the data set via a "one-click download". Please enter a URL only.
D2	Media Format	[1,1]	In some cases files are downloaded in a compressed file (e.g. zip). Please enter the media type for information contained within the compressed file (RSS, XML, CSV/TXT, XLS, Shapefile, KML/KMZ) If not compressed, enter the file suffix of the downloadable file.
D3	File size	[1,1]	If downloadable, please enter the size of file in MB. Should be limited to 15 characters.
D4	File format	[1,1]	If downloadable, enter the format in which file may be downloaded. For Raw Data Catalog, select from the following options: RSS, XML, TXT (CSV), XLS, KML/KMZ, Shapefile, or map. For Tool Catalog, select either Data Extraction Tool or Widget.
Additional Statistical Metadata required for Statistical Data Sets (i.e., above Element 2 = Statistical)			
SM	Statistical methodology	[0,n]	Components identifying statistical information/properties of data.
SM.01	Sampling	[0,1]	Description or URL of resource containing more detailed information.
SM.02	Estimation	[0,1]	Description or URL of resource containing more detailed information.
SM.03	Weighting	[0,1]	Description or URL of resource containing more detailed information.
SM.04	Disclosure avoidance	[0,1]	Description or URL of resource containing more detailed information.
SM.05	Questionnaire design	[0,1]	Description or URL of resource containing more detailed information.
SM.06	Series breaks	[0,1]	Description or URL of resource containing more detailed information.
SM.07	Non-response adjustment	[0,1]	Description or URL of resource containing more detailed information.
SM.08	Seasonal adjustment	[0,1]	Description or URL of resource containing more detailed information.
SM.09	Data quality (variances, CVs, CIs, etc)	[0,1]	Description or URL of resource containing more detailed information.
OGD Response specific metadata			
OGD	Is this record a part of OGD submissions	[1,1]	Answer yes/no to indicate if this dataset submission is in response to the OGD announcement.
OGD.01	Is this record listed in your agency open government plan	[1,1]	Answer yes/no to indicate if this dataset submission is listed in your Agency's Open Governemtn plan.
OGD.02	Is this a high value dataset	[1,1]	Answer yes/no to indicate if this dataset submission is considered "high value".
OGD.03	What makes this a high value dataset	[1,1]	Make a selection that best describes your validation for defining the dataset as "high value". This field is conditionally required: if OGD.02 above is "Yes". (Example: Improves public knowledge of the agency and its operations )
OGD.04	How is this new	[1,1]	Please provide a brief explanation how this data is new. This field is conditionally required. If Element OGD.02 above is "Yes", please provide a brief explanation why this dataset is considered new.