Environmental Protection Agency

EPA's Implementation of the Data.gov Standard: Creating Metadata for EPA Non-geospatial Datasets

February 7, 2011

Prepared by
Office of Information Collection
Office of Environmental Information

CONTENTS

I. INTRODUCTION

Document Scope and Applicability

The primary purpose of this document is to establish guidelines for creating metadata for non-geospatial datasets and services developed by the Environmental Protection Agency (EPA). This document is intended as a reference specifically for metadata that references datasets and services that are *non-geospatial* in nature. It serves as a complementary specification to the EPA's Geospatial Metadata Technical Specification version 1.0 (2007) and is intended as an outline of requirements and guidelines for metadata records that describe datasets that are not geospatial in nature. Individuals wishing to create metadata documents that describe *geospatial* datasets or services should reference EPA's Geospatial Metadata Technical Specification version 1.0. For more information regarding the distinction between geospatial and non-geospatial datasets, please see the **Geospatial Data Definition** provided below.

The intent of establishing this document for EPA's non-geospatial metadata for datasets is two-fold: 1). To ensure that consistent implementation practices are followed for non-geospatial dataset metadata implementation across the Agency and 2) To ensure that sufficient information is provided within metadata published by different Agency authors so that it may serve multiple Agency needs. Consistency in metadata publishing and management practices leads to improved discovery and reuse of resources for internal purposes while also improving Agency support for interagency initiatives, such as Data.gov. In order to support interagency data sharing initiatives, metadata records must adhere to federal standards. The standard used for non-geospatial metadata is EPA's Implementation of the Data.gov Standard. *This document provides users who are developing non-geospatial metadata with clear guidelines for creating content that meets EPA's requirements for adhering to the Data.gov standard.* The specifications set forth in this document should be applied to the creation of all *new* non-geospatial metadata records at EPA.

This document does not provide a detailed overview of the business processes that should be followed for creating and contributing metadata to agency catalogs. The basic steps that can be followed for creating and sharing metadata at EPA are shown in Figure 1 below.

Determining Which Metadata Standard Should be Used for Documenting Datasets

Metadata developers must decide which standard to use to document their data (geospatial or non-geospatial metadata standard). The distinction between geospatial and non-geospatial data may not be clear to all data owners. The *geospatial data definition* provided below is intended to assist users in making the determination as to whether their content should be considered geospatial or non-geospatial. Data owners should review the *geospatial data definition* to determine whether or not their data is considered 'geospatial'. If the data is considered to be 'geospatial data' as per the definition provided below, then use the EPA's Geospatial Metadata Technical Specification to create your metadata records. If your data is not classified as 'geospatial data' as per the definition provided below, then use this document as a reference for documenting your data.

Geospatial Data Definition: "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."

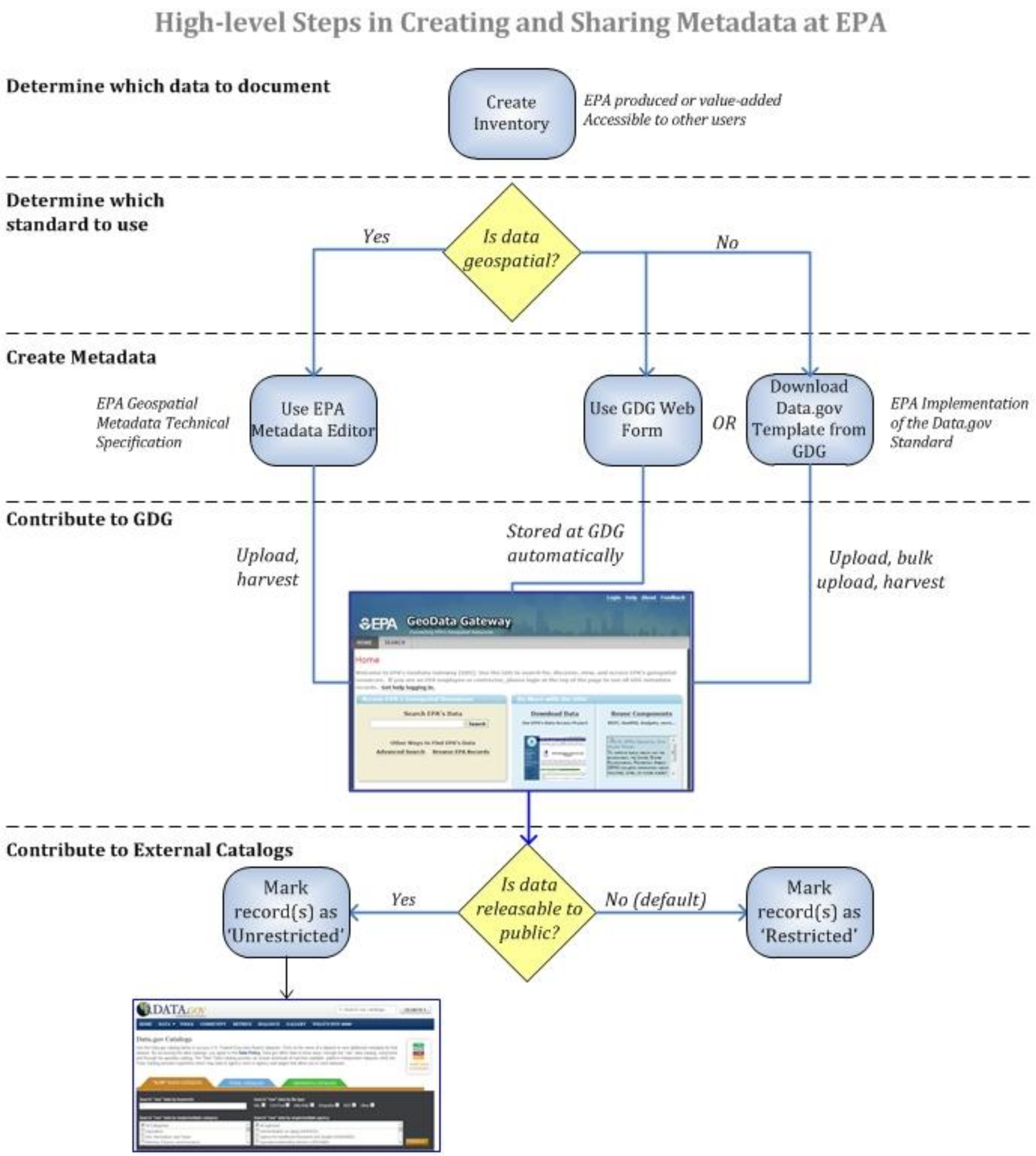
High-level Steps in Creating and Sharing Metadata at EPA
Determine which data to document
Create Inventory
EPA produced or value-added
Accessible to other users
Determine which standard to use
Is data geospatial?
Yes
No
Create Metadata
EPA Geospatial Metadata Technical Specification
Use EPA Metadata Editor
Use GDG Web Form
OR
Download Data.gov Template from GDG
EPA Implementation of the Data.gov Standard
Contribute to GDG
Upload, harvest
Stored at GDG automatically
Upload, bulk upload, harvest
EPA GeoData Gateway
Home
Contribute to External Catalogs
Mark record(s) as 'Unrestricted'
Yes
Is data releasable to public?
No (default)
Mark record(s) as 'Restricted'
DATA.GOV
Data.gov Catalogs

FIGURE 1. METADATA BUSINESS PROCESS STEPS

II. EPA'S IMPLEMENTATION OF THE DATA.GOV STANDARD

About EPA's Implementation of the Data.gov Standard

The EPA Implementation of the Data.gov standard is based on version 2 of the Data.gov Metadata Template for Datasets and Tools which was developed for the Data.gov Dataset Management System (DMS) portal as an initial roadmap to incorporate non-geospatial information and required elements. The EPA Implementation of the Data.gov Standard provides an interpretation of and guidelines for elements outlined in the Data.gov standard. Metadata that is compliant with the EPA Implementation of the Data.gov Standard ensures that it is also compliant with requirements for use across multiple venues.

This section outlines EPA's requirements for Agency-specific non-geospatial metadata, providing guidance on language for specific elements, information on how to make decisions for certain element classification schemes, and the relationship (if any) to Agency standards used or referenced. This document is not intended as a complete explanation of the Dublin Core Metadata Initiative (DCMI) and Data.gov standards; rather, it is intended as a reference for guidance only on those elements that are considered important for the EPA Implementation. Users wishing to obtain additional information about the DCMI or Data.gov should visit http://dublincore.org/about-us/ or http://dms.data.gov.

A detailed overview of EPA's Implementation of the Data.gov Standard is provided below.

EPA'S DATA.GOV IMPLEMENTATION: STANDARD ELEMENTS AND REQUIREMENTS

This section of the document provides an overview of the specific elements, requirements, and guidelines for the EPA's Implementation of the Data.gov Standard. The elements are listed in the table below to enable the user to understand what the fields are, how to document them, and how they relate to the various pieces/parts of the business process and infrastructure.

#	Element Name	Required (Y/N)	Data.gov Description/Guidelines	EPA Implementation/ Guidance	Domain/Format Restrictions	Example Content
1	**Unique ID**	Y	For Data.Gov PMO Use Only - will be generated when placed in catalog	FOR DATA.GOV PMO USE ONLY	FOR DATA.GOV PMO USE ONLY	
1.1	**User Generated ID**	N	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 Data.gov and will not be published on the catalog.	Follow Data.gov guidelines for the implementation of this field. If there is a meaningful ID that may track this record to your system and then enable your users to easily find that content at the EDG, then it is recommended that your organization make use of the unique ID field.	Free text	
2	**Title**	Y	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.	Provide a meaningful title that includes the date, area of coverage, such as the state name, or other regional breakdown (where applicable); full name of data and acronym of data if possible. Including the term "EPA" and your organization in the title can also be useful to others outside the agency. Please note that all submissions that are intended for Data.gov should have unique titles.	Free text	EPA Facility Registry System (FRS) 2010 Facilities Data for the State of Nebraska

#	Element Name	Required (Y/N)	Data.gov Description/Guidelines	EPA Implementation/ Guidance	Domain/Format Restrictions	Example Content
3	**Dataset Group Name**	N	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.	If the data belongs to a larger group, then include the name of the group consistently across all entries so that a search for the content will retrieve all values. In general, it is recommended that if data is part of a larger group, that a single record be provided with multiple links to the data rather than multiple individual records.	Free text	2006 Toxics Release Inventory data
4	**Description**	Y	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.	Follow Data.gov guidelines for the implementation of this field.	Free text	The Toxics Release Inventory (TRI) is a publicly available EPA database that contains information on toxic chemical releases and waste management activities reported annually by facilities in certain industries as well as federal facilities.
5	**Agency Name**	Y	Department or Independent Agency name.	Use the Term "Environmental Protection Agency"	Environmental Protection Agency	Environmental Protection Agency
5.1	**Agency Short Name**	Y	Acronym or short name corresponding to the Agency name (e.g. DOC, DoD, NASA, GSA).	Use the Term "EPA"	EPA	
6	**Sub-Agency Name**	Y	Bureau or Sub-Agency or operating unit name.	Specify EPA Organization Name. List the AA-ship, Office, Division, and Branch if possible. Include as much organizational content as possible so that the information can be searched and found for your organization.	Please choose from Sub-agency Name List shown in Appendix A	Office of Environmental Information, Office of Information Collection, Information Exchange & Services Division, Information Services and Support Branch
6.1	**Sub-Agency Short Name**	Y	Acronym or short name corresponding to the Sub-Agency name (e.g. IRS, FBI, BIA).	Specify EPA Organization Acronym. List the AA-ship, Office, Division, and Branch if possible. Include as much organizational content as possible so that the information can be searched and found for your organization.	Please choose from Sub-agency Short Name List shown in Appendix A	OEI, OIC, IESD, ISSB

#	Element Name	Required (Y/N)	Data.gov Description/Guidelines	EPA Implementation/ Guidance	Domain/Format Restrictions	Example Content
7	**Contact Name**	Y	Contact person's name (first, then last) where questions from the Data.gov PMO should be sent. ** Note that these "contact" elements (7, 7.1 & 7.2) will not be published on the Data.gov website but may receive public email comments made specifically about this dataset or tool.**	Follow Data.gov guidelines for the implementation of this field. Please note that contact information **is** displayed at the EDG web interface.	Free text	Michelle Torreano
7.1	**Contact Phone Number**	Y	Contact person's phone number.	Follow Data.gov guidelines for the implementation of this field. Please note that contact information **is** displayed at the EDG web interface.	Free text	202-566-2141
7.2	**Contact Email Address**	Y	Contact person's email address.	Follow Data.gov guidelines for the implementation of this field. Please note that contact information **is** displayed at the EDG web interface.	Free text	torreano.michelle@epa.gov
8	**Agency responsible for Information Quality**	Y	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.	This should typically be entered as "Environmental Protection Agency". Refer to EPA's data quality guidance for additional details. http://www.epa.gov/quality/qa_docs.html.	Environmental Protection Agency	Environmental Protection Agency
8.1	**Compliance with Agency's Information Quality Guidelines**	Y	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.	Enter "Yes" for this field, unless the content does not meet EPA's Information Quality Guidelines. If the content does not meet EPA's Information Quality Guidelines, then the information should not be shared.	Yes, No	Yes

#	Element Name	Required (Y/N)	Data.gov Description/Guidelines	EPA Implementation/ Guidance	Domain/Format Restrictions	Example Content
8.2	**Privacy and Confidentiality**	Y	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.	Information that can be shared with Data.gov should be marked as "Yes" in this field. Information that can be shared within EPA only via the Environmental Dataset Gateway (EDG) should be marked as "No" in this field. This field should align with the Environmental Dataset Gateway's Access Level classification. Records with a "No" in this field should be classified as "Restricted" at the EDG, and those with a "Yes" in this field should be classified as "Unrestricted" at the EDG.	Yes, No	Yes
9	**Data.gov catalog type**	Y	Select the appropriate catalog on Data.gov that this submission should be listed (i.e. Raw Data Catalog or Tool Catalog)	Downloadable data should be classified as "Raw Data" and Applications should be classified as "Tool Catalog"	Raw Data Catalog, Tool Catalog	Raw Data Catalog
10	**Subject area (Taxonomy)**	Y	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.	Follow Data.gov guidelines for the implementation of this field. For many EPA developed datasets and resources, this classification will be "Geography and Environment".	Please choose 1 from Subject Area list (Appendix A)	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 http://geodata.gov and you do not need to fill out any additional information for this submission. We would,	Please Select the appropriate category. Data that is identified as Geospatial should be documented using the EPA's Geospatial Metadata Technical Specification.	Administrative, Statistical, Surveillance, Research, Other	Administrative

#	Element Name	Required (Y/N)	Data.gov Description/Guidelines	EPA Implementation/ Guidance	Domain/Format Restrictions	Example Content
			however request that you submit the elements 1 through 7 to datagov.submit@gsa.gov so that we can synchronize our efforts with geodata.gov. If you select "statistical", please note the section "SM" that follows.			
12	**Keywords**	Y	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.	Please use Keywords as defined for your organization that meet the specifics of your dataset. Please refer to EPA's Web Taxonomy Guidelines for more information. http://yosemite.epa.gov/OEI/webguide.nsf/resources/webtaxonomy	N/A	OSWER, CEPPO, Chemicals, Emergency Planning and Community Right-To-Know Act, EPCRA, Superfund Amendments and Reauthorization Act, SARA, Title III, CERCLA, reporting, Clean Air Act, CAA, 302, 304, 313, Section 112®
13	**Date released**	Y	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.	Follow Data.gov guidelines for the implementation of this field	Date Format (m/dd/yyyy)	8/12/2010
14	**Date updated**	Y	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).	Follow Data.gov guidelines for the implementation of this field	Date Format (m/dd/yyyy)	8/12/2010
15	**Agency Program URL**	Y	URL that is closest to the program that is responsible for this dataset or tool.	Provide a URL to the organizational web page that is responsible for providing the data.	Free text	http://www.epa.gov/tri/

#	Element Name	Required (Y/N)	Data.gov Description/Guidelines	EPA Implementation/ Guidance	Domain/Format Restrictions	Example Content
16	**Agency Data Series URL**	Y	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).	Provide a website that describes the data/information specifically.	Free text	http://www.epa.gov/tri/tridata/current_data/index.html
17	**Collection mode**	Y	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.	Follow Data.gov guidelines for the implementation of this field	Free text	Computer
18	**Frequency**	Y	Frequency of data collection (one-time, annual, hourly, etc.).	Follow Data.gov guidelines for the implementation of this field	Free text	Annual
19	**Period of Coverage**	Y	Dates or time interval(s) covered by the data. Please use commas to separate multiple periods of Coverage.	Follow Data.gov guidelines for the implementation of this field. This can be a date or a time frame as specified in text.	Free text	Calendar Year 2006
20	**Unit of analysis**	Y	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.	Follow Data.gov guidelines for the implementation of this field	Free text	Facility, Name of Chemical, A single TRI chemical released or managed by a qualifying facility
21	**Geographic scope**	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 simple leave it blank. You may use commas to separate multiple entries.	Enter the US State Name(s), or the Territories that are covered by the metadata. Including as many states and place names as possible will maximize the value of your content at the EDG and Data.gov	Free text	California, USA
21.1	**Geographic Granularity**	N	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).	If the locational information is provided in latitude and longitude coordinates, then please use the EPA's Geospatial Metadata Technical Specification for metadata.	Free text	Zip Code

#	Element Name	Required (Y/N)	Data.gov Description/Guidelines	EPA Implementation/ Guidance	Domain/Format Restrictions	Example Content
22	**Reference for Technical Documentation**	Y	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.	Follow Data.gov guidelines for the implementation of this field. If your variable names and content is described at a web accessible location, please include that information here.	Free text	http://www.epa.gov/tri/tridata/current_data/basic/TRI_Basic_Data_File_Format_v09.pdf
23	**Data dictionary/variable list**	Y	URL to resource containing variable names, descriptions, standard vocabularies and taxonomies, units, multipliers, etc. May be identical to element 22.	Follow Data.gov guidelines for the implementation of this field. If your variable names and content is described at a web accessible location, please include that information here.	N/A	http://www.epa.gov/tri/tridata/current_data/basic/TRI_Basic_Data_File_Format_v09.pdf
24	**Data collection instrument**	Y	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.	Follow Data.gov guidelines for the implementation of this field	Free text	http://www.epa.gov/tri/tridata/current_data/basic/TRI_Basic_Data_File_Format_v09.pdf
25	**Bibliographic citation for dataset**	[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.	Follow Data.gov guidelines for the implementation of this field	Free text	http://www.epa.gov/tri/tridata/current_data/basic/TRI_2006_TX_v09.csv
26	**Number of Datasets Represented by this Submission**	[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. **	Follow Data.gov guidelines for the implementation of this field	Numeric (whole)	1
27	**Additional Metadata**	[0,1]	This is an optional field. Please provide a URL to any additional metadata for the dataset or tool.	Follow Data.gov guidelines for the implementation of this field	Free text	

#	Element Name	Required (Y/N)	Data.gov Description/Guidelines	EPA Implementation/ Guidance	Domain/Format Restrictions	Example Content
28	**Dataset use requires a license agreement**	[1,1]	This is a required field to ensure that license agreements are not bypassed during the one-click download interface on the website.	If data are restricted by license agreements, make sure that the accessibility of the metadata is appropriate for internal/external parties.	Yes, No	No
29	**Dataset license agreement URL**	N	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.	Follow Data.gov guidelines for the implementation of this field	Free text	

EPA's Data.gov Implementation: Data Download Elements and Requirements

This section of the Data.gov standard may be repeated to include multiple download links.

#	Element Name	Required (Y/N)	Data.gov Description/Guidelines	EPA Implementation/ Guidance	Domain/Format Restrictions	Example Content
D1	**Access point**	Y	If the dataset is downloadable, enter the URL for instant access to the downloadable data file. This is the URL for access to the dataset via a "one-click download". Please enter a URL only.	Provide a URL that directly links users to the dataset(s) being described by this record. This field serves as the primary link to your dataset/resource.	Free text	http://www.epa.gov/tri/tridata/current_data/basic/TRI_2006_TX_v09.csv
D2	**Media Format**	Y	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.	Follow Data.gov guidelines for the implementation of this field. In general ensure that the media format field corresponds appropriately with content entered for the file format field. If data are stored in KML/KMZ or Shapefile form, you should be using the EPA's Geospatial Metadata Technical Specification. EPA non-geospatial content should be limited to RSS, XML, CSV/TXT, and/or XLS.	RSS, XML, CSV/TXT, XLS	CSV/TXT
D3	**File size**	Y	If downloadable, please enter the size of file in MB. Should be limited to 15 characters.	Follow Data.gov guidelines for the implementation of this field	Numeric (float)	15.5

#	Element Name	Required (Y/N)	Data.gov Description/Guidelines	EPA Implementation/ Guidance	Domain/Format Restrictions	Example Content
D4	**File format**	Y	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.	If your data are stored in KML/KMZ, Map, Shapefile, or map form, you should be using the EPA's Geospatial Metadata Technical Specification. Please limit this form to RSS, XML, TXT (CSV), XLS, and Data Extraction Tool or Widget.	RSS, XML, TXT (CSV), XLS. For Tool Catalog, select either Data Extraction Tool or Widget.	TXT (CSV)

EPA'S DATA.GOV IMPLEMENTATION: STATISTICAL ELEMENTS AND REQUIREMENTS

This section of the Data.gov standard applies to resources that contain statistical information.

#	Element Name	Required (Y/N)	Data.gov Description/Guidelines	EPA Implementation/ Guidance	Domain/Format Restrictions	Example Content
SM	**Statistical methodology**	N	Components identifying statistical information/properties of data.	Follow Data.gov guidelines for the implementation of this field	Free Text	
SM.01	**Sampling**	[0,1]	Description or URL of resource containing more detailed information.	Follow Data.gov guidelines for the implementation of this field	Free Text	
SM.02	**Estimation**	[0,1]	Description or URL of resource containing more detailed information.	Follow Data.gov guidelines for the implementation of this field	Free Text	
SM.03	**Weighting**	[0,1]	Description or URL of resource containing more detailed information.	Follow Data.gov guidelines for the implementation of this field	Free Text	
SM.04	**Disclosure avoidance**	[0,1]	Description or URL of resource containing more detailed information.	Follow Data.gov guidelines for the implementation of this field	Free Text	
SM.05	**Questionnaire design**	[0,1]	Description or URL of resource containing more detailed information.	Follow Data.gov guidelines for the implementation of this field	Free Text	
SM.06	**Series breaks**	[0,1]	Description or URL of resource containing more detailed information.	Follow Data.gov guidelines for the implementation of this field	Free Text	
SM.07	**Non-response adjustment**	[0,1]	Description or URL of resource containing more detailed information.	Follow Data.gov guidelines for the implementation of this field	Free Text	
SM.08	**Seasonal adjustment**	[0,1]	Description or URL of resource containing more detailed information.	Follow Data.gov guidelines for the implementation of this field	Free Text	
SM.09	**Data quality (variances, CVs, CIs, etc)**	[0,1]	Description or URL of resource containing more detailed information.	Follow Data.gov guidelines for the implementation of this field	Free Text	

EPA'S DATA.GOV IMPLEMENTATION: OPEN GOVERNMENT DIRECTIVE (OGD) ELEMENTS AND REQUIREMENTS

#	Element Name	Required (Y/N)	Data.gov Description/Guidelines	EPA Implementation/ Guidance	Domain/Format Restrictions	Example Content
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.	Most EPA datasets will be listed as "No" for this field, unless the data is specifically called out in the EPA's Open Government Plan (http://www.epa.gov/open/)	Yes, No	No
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 Government plan.	Most EPA datasets will be listed as "No" for this field, unless the data is specifically called out in the EPA's Open Government Plan (http://www.epa.gov/open/)	Yes, No	No
OGD.02	**Is this a high value dataset**	[1,1]	Answer yes/no to indicate if this dataset submission is considered "high value". High-value information is information that can be used to 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.	Follow Data.gov guidelines for the implementation of this field. The majority of EPA datasets will contain 'yes' for this field unless there is justification for classifying the content as 'no'.	Yes, No	Yes
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".	One or more of the options below should be selected. 1. Increase agency accountability and responsiveness. 2. Improve public knowledge of the agency and its operations. 3. Further the core mission of the agency. 4. Create economic opportunity. 5. Respond to need and demand as identified through public consultation.	Increases agency accountability and responsiveness; Improves public knowledge of the agency and its operations; Furthers the core mission of the agency; Creates or expands economic opportunity; Responds to need and demand as identified through public consultation	Furthers the core mission of the agency

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.	Suggested guidance for considering a dataset to be 'new': 1. The data were not available before 2. This is a new version of a previously, or available dataset that has had significant enhancements, or 3. The data were only available internally at EPA before and have recently been made available to the public.	Free text	Created a newly available dataset based on user requests.

III. REFERENCES

- Data.Gov http://www.data.gov/
- Dublin Core Metadata Initiative (DCMI) http://dublincore.org/about-us/
- EPA’s Geospatial Metadata Technical Specification version 1.0 (2007)

  http://epa.gov/geospatial/docs/EPA_Geospatial_Metadata_Technical_Specification_v1_11_2_2007.pdf
- Federal Geographic Data Committee (FGDC), Content Standard for Digital Geospatial Metadata (CSDGM), Version 2 - FGDC-STD-001-1998 http://www.fgdc.gov/metadata/contstan.html
- U.S. EPA Controlled Vocabulary of Subject Terms http://www.epa.gov/webguide/metadata/

IV. APPENDICES

Appendix A: EPA Data.gov Implementation Controlled Vocabularies

Sub-Agency Name List

- Office of Administration and Resources Management
- Office of Air and Radiation
- Office of Chemical Safety and Pollution Prevention
- Office of Enforcement and Compliance Assurance
- Office of Environmental Information
- Office of General Counsel
- Office of Inspector General
- Office of International and Tribal Affairs
- Office of Solid Waste and Emergency Response
- Office of the Chief Financial Officer
- Office of Water
- Office Research and Development
- Region 1
- Region 2
- Region 3
- Region 4
- Region 5
- Region 6
- Region 7
- Region 8
- Region 9
- Region 10

Sub-Agency Short Name List

- OAR
- OARM
- OCFO
- OCSPP
- OECA
- OEI
- OGC
- OIG
- OITA
- ORD
- OSWER
- OW
- Region 1
- Region 10
- Region 2
- Region 3
- Region 4
- Region 5
- Region 6
- Region 7
- Region 8
- Region 9

Subject Area List

1. Population
2. Births, Deaths, Marriages, and Divorces
3. Health and Nutrition
4. Education
5. Law Enforcement, Courts, and Prisons
6. Geography and Environment
7. Elections
8. State and Local Government Finances and Employment
9. Federal Government Finances and Employment
10. National Security and Veterans Affairs
11. Social Insurance and Human Services
12. Labor Force, Employment, and Earnings
13. Income, Expenditures, Poverty, and Wealth
14. Prices
15. Business Enterprise
16. Science and Technology
17. Agriculture
18. Natural Resources
19. Energy and Utilities
20. Construction and Housing
21. Manufactures
22. Wholesale and Retail Trade
23. Transportation
24. Information and Communications
25. Banking, Finance, and Insurance
26. Arts, Recreation, and Travel
27. Accommodation, Food Services, and Other Services
28. Foreign Commerce and Aid
29. Puerto Rico and the Island Areas
30. International Statistics
31. Other

Specialized Data Category Designation List

- Administrative
- Statistical
- Surveillance
- Research
- Other

Open Government Directive List

- Increases agency accountability and responsiveness
- Improves public knowledge of the agency and its operations
- Furthers the core mission of the agency
- Creates or expands economic opportunity
- Responds to need and demand as identified through public consultation
- Other

APPENDIX B: DATA.GOV GLOSSARY OF TERMS (HTTP://WWW.DATA.GOV/GLOSSARY)

Term	Definition
Catalog ***(source: Data.gov)***	A catalog is a collection of datasets. **Data.gov** has three types of searchable data catalogs: The "Raw Data Catalog" features instant view/download of datasets; the "Tool Catalog" contains simple, application-driven access to federal data; and the "Geodata Catalog" contains federal geospatial data.
Category ***(source: Data.gov)***	The category identifies the type of dataset (e.g., administrative, geospatial, research, statistical). Some types of data have additional metadata requirements.
CSV ***(source: Wikipedia)***	A comma separated values (CSV) file is a computer data file used for implementing the tried and true organizational tool, the Comma Separated List. The CSV file is used for the digital storage of data structured in a table of lists form. Each line in the CSV file corresponds to a row in the table. Within a line, fields are separated by commas, and each field belongs to one table column. CSV files are often used for moving tabular data between two different computer programs (like moving between a database program and a spreadsheet program).
Data ***(source: Federal Enterprise Architecture: Data Reference Model)***	A value or set of values representing a specific concept or concepts. Data become "information" when analyzed and possibly combined with other data in order to extract meaning, and to provide context. The meaning of data can vary depending on its context.
Data Extraction Tool ***(source: Data.gov)***	Data extraction tools allow a user to select a data basket full of variables and then recode those variables into a form that the user desires. The user can then develop customized displays of any selected data.
Dataset ***(adapted from: Wikipedia)***	A dataset is an organized collection of data. The most basic representation of a dataset is data elements presented in tabular form. Each column represents a particular variable. Each row corresponds to a given value of that column's variable. A dataset may also present information in a variety of non-tabular formats, such as an extended mark-up language (XML) file, a geospatial data file, or an image file, etc.
KML ***(source: Wikipedia)***	Keyhole Markup Language (KML) is an XML-based language schema for expressing geographic annotation and visualization on existing or future Web-based, two-dimensional maps and three-dimensional Earth browsers.
KMZ ***(source: Wikipedia)***	KML files are very often distributed in KMZ files, which are zipped files with a ".KMZ" extension. When a KMZ file is unzipped, a single "doc.kml" is found along with any overlay and icon images referenced in the KML as well as any network-linked KML files.
Metadata ***(source: Federal Enterprise Architecture: Data Reference Model)***	Metadata describes a number of characteristics, or attributes, of data; that is, "data that describes data". (ISO 11179-3). For any particular datum, the metadata may describe how the datum is represented, ranges of acceptable values, it should be labeled, as well as its relationship to other

Term	Definition
	data. Metadata also may provide other relevant information, such as the responsible steward, associated laws and regulations, and access management policy. The metadata for structured data objects describes the structure, data elements, interrelationships, and other characteristics of information, including its creation, disposition, access and handling controls, formats, content, and context, as well as related audit trails.
Shapefile ***(source: ESRI Shapefile Technical Description)***	A shapefile stores nontopological geometry and attribute information for the spatial features in a dataset. The geometry for a feature is stored as a shape comprising a set of vector coordinates. Shapefiles can support point, line, and area features.
XML ***(source: Wikipedia)***	XML (Extensible Markup Language) is a general-purpose specification for creating custom markup languages. It is classified as an extensible language, because it allows the user to define the mark-up elements. XML's purpose is to aid information systems in sharing structured data, especially via the Internet, to encode documents, and to serialize data.
From MetaData Page	
Date Released ***(source: Data.gov)***	The date that the dataset was originated.
Date Updated ***(source: Data.gov)***	The date that the dataset was last modified.
Time Period ***(source: Data.gov)***	Date or time interval(s) for which the dataset provides data.
Frequency ***(source: Data.gov)***	Frequency of data collection (one-time, annual, hourly, etc.).
Data.gov Data Category Type ***(source: Data.gov)***	The category designation for the entry as either an instantly downloadable raw data file or tool (i.e., data extraction and mining or widget).
Specialized Data Category Designation ***(source: Data.gov)***	The type of dataset (e.g., administrative, geospatial, research, or statistical). Some types of data have additional metadata requirements.
Keywords ***(source: Dublin Core)***	Used to describe the content of the resource. The element may use controlled vocabularies or words or phrases that describe the subject or content of the resource.
Unique ID ***(source: Data.gov)***	An unambiguous reference to the resource within a given context. **Dublin Core** defines best practice for this field as identifying the resource by a unique number (e.g., ISBN, ISSN, URL/URI, etc.). The Unique ID is intended for **Data.gov** internal reference only.
Citation ***(source: FGDC-STD-001-1998)***	The recommended reference citation to be used to cite the dataset.

Term	Definition
Agency Program Page ***(source: Data.gov)***	The URL link (and name, if applicable) to the home page of the agency or program that is the dataset owner.
Agency Data Series Page ***(source: Data.gov)***	The URL link (and name, if applicable) to the agency web page where the link to the dataset is located. This is different from the URL for the actual dataset.
Unit of Analysis ***(source: Data.gov)***	The level of granularity or aggregation which is represented by a single record or observation in a dataset (e.g. person, household, production workers, establishment, city, country).
Granularity ***(source: Dublin Core)***	The level of detail at which an information object or resource is viewed or described.
Geographic Coverage ***(source: Dublin Core)***	Used to designate the extent or scope of the content of the resource and typically includes spatial location (a place name or geographic co-ordinates).
Collection Mode ***(source: Data.gov)***	Identifies the modality of the instrument used to gather data for the dataset (e.g., phone/paper, phone/computer, person/paper, person/computer, web, fax, other).
Data Collection Instrument ***(source: Data.gov)***	Identifies the specific instrument or tool (e.g., form, survey questionnaire) used to collect the data in the dataset corresponding to the collection mode.
Data Dictionary/Variable List ***(source: Federal Enterprise Architecture: Data Reference Model)***	A database used for data that refers to the use and structure of other data; that is, a database for the storage of metadata [ANSI X3.172-1990].
Data Quality ***(source: OMB Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies, 67 FR 8452)***	"Quality" is an encompassing term comprising objectivity, utility, and integrity. Sometimes these terms are referred to collectively as "quality." Any agency contributing a dataset to **Data.gov** must certify that the dataset conforms to the agency's information quality guidelines.
Privacy and Confidentiality ***(source: 44 U.S.C. 3542)***	Preserving authorized restrictions on information access and disclosure, including means for protecting personally identifiable and proprietary information. Any agency contributing a dataset to **Data.gov** must certify that dissemination of the data is consistent with the agency's responsibilities under the Privacy Act and, if applicable, the Confidential Information Protection and Statistical Efficiency Act of 2002.
Technical Documentation ***(source: Data.gov)***	Additional documentation that describes a dataset and its intended use.
Additional Metadata ***(source: Data.gov)***	Additional metadata that may be available for a dataset. Such metadata may conform to an existing standard (e.g., FGDC Metadata Standard).

Term	Definition
Statistical Methodology *(source: Data.gov)*	A description of the overall approach used for statistical design, sampling, data collection, statistical analysis, and estimation.
Sampling ***(source: Box, Hunter, and Hunter, Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building, 1978)***	The procedure used to define the total number of statistical observations (i.e., samples) from an overall population size.
Estimation ***(source: Box, Hunter, and Hunter, Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building, 1978)***	The approach used to compute statistical quantities based on the observations (e.g., mean, mode, standard deviation).
Weighting ***(source: Box, Hunter, and Hunter, Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building, 1978)***	An approach for applying a scaling factor to observations from one or more combined data series in order to normalize or otherwise adjust the observations.
Disclosure avoidance *(source: Federal Committee on Statistical Methodology)*	Techniques (e.g., aggregation) that are applied to statistical data to ensure published data cannot be used to attribute a specific value to an individual.
Questionnaire design *(source: Data.gov)*	A structured approach used to develop a questionnaire or survey that describes the structure and content of the survey instrument and the approach intended to be used for analyzing the survey results.
Series breaks *(source: Data.gov)*	A discrete event or changes to the sample, the population, their environment, or the survey instrument occurring within a data collection that may affect statistical estimates or inferences.
Non-response adjustment ***(source: Box, Hunter, and Hunter, Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building, 1978)***	The approach for adjusting observations to account for missing or incomplete data within a series.
Seasonal adjustment ***(source: Wikipedia)***	A statistical method for removing the effects of seasonal variation of a time series that is used when analyzing non-seasonal trends.
Statistical Characteristics (CV, CI, variance, etc.) ***(source: Box, Hunter, and Hunter, Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building, 1978)***	Summary of statistical characteristics that reflect the overall accuracy and correlation of a statistical data sample relative to the overall population including coefficients of variation, confidence intervals, and variance.