U.S. Environmental Protection Agency
Open Data Policy Implementation Plan

I.   Introduction and Background

This report documents the U.S. Environmental Protection Agency's (EPA) implementation plan and current progress in meeting the requirements set forth in the White House "Open Data Policy – Managing Information as an Asset" Memorandum M-13-13. The Agency understands that the information it manages is of significant interest and value to the public and reaffirms its commitment to share uncontrolled information, be responsive to public requests and promote the proper use of all information as a standard Agency practice. Many of the activities identified in this plan were initiated as part of the Agency's Open Government initiative. Progress made on these efforts underscores the significant investment EPA has made in promoting its data for public use.

The Agency maintains publicly available registries[1] with associated metadata for EPA data systems and EPA-generated datasets. In 2011, EPA expanded the Environmental Dataset Gateway (EDG)[2], a metadata catalog that was initially limited to geographic data, to accommodate metadata for non-geospatial datasets as well. The EDG team also maintains an open-source desktop metadata editor that simplifies and streamlines the metadata authorship process. Users around the world have downloaded this tool approximately 8,000 times. In addition to the EDG, EPA maintains six additional registries, collectively known as the System of Registries (SOR). SOR provides environmental program managers and users of environmental information with automated services to enable better understanding of environmental terminology and data used by the Agency as well as taxonomies to help cross reference similar terms used in the environmental community.

The Agency also provides resources for developers interested in building applications using EPA data and Web services through Developer Central. Developer Central is a resource for developers who are interested in building applications using EPA data and Web services. It highlights priority data sources found in the EDG, Application Program Interfaces (API) and external resources located within EPA's publicly available registries.

The Agency has a variety of additional efforts under way to identify priority datasets wanted by stakeholders, make EPA data more open to the public and increase public engagement in Open Government activities. EPA aims to make the data most frequently requested by the public easily available. When the Agency determined that a large percentage of Freedom of Information Act (FOIA) requests were related to EPA's property-specific data, it developed MyProperty. This tool allows real estate agents, mortgage banks, engineering and environmental consulting firms and the public to determine whether EPA databases have records on a specific property without filing a FOIA request. EPA supports the Green Button initiative via its voluntary ENERGY STAR Program, which identifies and promotes energy efficient practices in residential and commercial buildings. Green Button is an industry-led effort that began in January 2012 in response to a White House directive to provide utility customers electronic access to their energy data in a standard format. Where feasible, EPA has

---

[1] For more information on EPA's System of Registries, please visit: www.epa.gov/sor

[2] For more information on the Environmental Dataset Gateway, please visit: https://edg.epa.gov

worked to integrate ENERGY STAR tools targeted at residential and commercial stakeholders with Green Button efforts.

EPA will report its progress in promoting data for public use by updating its dataset inventory and in meeting other aspects of the Open Data Policy on quarterly basis. Updates will be posted on EPA's Digital Strategy website.

Inventory Schedule Requirement	Due Date	Status
Develop tool to create inventory (EDG)	11/30/13	Complete
Confirm EDG accommodates Open Data Policy metadata	11/30/13	Complete
Develop capability to create JSON file	11/30/13	Complete
Post link on EPA's Digital Strategy website	11/30/13	Complete
Include datasets posted on Data.gov before 8/1/13	11/30/13	Complete
Upload snapshot to MAX	11/30/13	Complete
Confirm schema follows Project Open Data (POD)	11/30/13	Schema 1.0 Complete, Schema 1.1 In Progress
Provide reports to programs to identify metadata holes	Quarterly	Ongoing
Create dashboard to measure progress in filling metadata holes	Quarterly	In Progress
Require programs to ID all remaining datasets/systems to determine if the data can be made public	Quarterly	In Progress

II. Enterprise Data Inventory – Environmental Dataset Gateway (EDG)

The EDG serves as the primary tool for generating both the Agency's internal and public-facing inventory that comprises the Open Data Policy requirement for an enterprise data inventory.

Enterprise Data Inventory Requirement	Due Date	Status
Develop an inventory schedule	11/30/13	Complete
Publish inventory schedule on EPA's Digital Strategy website (via this Plan)	Ongoing	Ongoing
Create a comprehensive Enterprise Data Inventory (EDG)	11/30/14	Complete
Submit inventory snapshot to MAX (JSON format)	Quarterly	Ongoing
Maintain inventory: Expand, Enrich, Open	Quarterly	Ongoing
Upload snapshot quarterly in JSON format on MAX	Quarterly	Ongoing
Update Schedule on EPA's Digital Strategy website	Quarterly	Ongoing

III.   Public Data Listing

The purpose of creating and maintaining a public data listing (PDL) is to improve the discoverability and promote use of uncontrolled data assets. EDG serves as the EPA's PDL, providing the public with access to the Agency's uncontrolled data assets. EPA requires publication of datasets in the EDG before they are published to Data.gov. EPA has adopted enhanced metadata requirements, Project Open Data (POD) Metadata Schema v1.1, to adequately describe datasets registered in the EDG and fulfill Open Data Policy metadata requirements. It also conducts reviews of EDG records on an ongoing basis and takes action as needed to ensure each is sufficiently described. The EDG is available to public for viewing uncontrolled EPA information at [Environmental Dataset Gateway](). EPA will provide a summary of updates to the EDG on a quarterly basis at [EPA's Digital Strategy website]().

Public Data Listing Requirement	Due Date	Status
Create and publish PDL in JSON format at epa.gov/data	11/30/13	Complete
Maintain the PDL	Ongoing	Ongoing
Include assets where "accessLevel = public" in the EDG	11/30/13	Ongoing
Include "accessURL" to link the data asset's metadata for all data assets in the PDL	11/30/13	Ongoing

IV.   Customer Engagement

EPA understands the importance of engaging with the public to help prioritize the Agency's data publication efforts. [E.O. 13642]() requires all new non-sensitive data collections and information assets created after May 9, 2013, to follow Open Data practices and be promoted actively for public use. To help direct the Agency's focus and prioritize which datasets are of greatest interest to the public, EPA will consider feedback from the public as it prioritizes data publication actions.

The public is encouraged to view the Agency's PDL, follow EPA updates and provide any feedback from links contained at [EPA's Digital Strategy website](). In addition to direct public feedback and requests, EPA will look at similar requests made during previous efforts, along with information we can surmise from the Agency's website metrics, as indicators of what is of greatest interest to the public. The Agency will also continue to monitor FOIA requests and look for additional ways to make the data most frequently requested more easily available to the public.

Through Developer Central, EPA promotes the use of EPA data and APIs by working closely with universities and by attending software-coding events to promote the development of applications that help the public better understand and protect the environment. Developer Central also provides mechanisms for the public and developers to provide feedback about priority datasets, APIs and application ideas.

U.S. Environmental Protection Agency
Open Data Policy Implementation Plan

Customer Engagement Requirement	Due Date	Status
Establish Customer Feedback Mechanisms	11/30/13	Complete
Describe customer feedback processes	11/30/13	Complete
Refine Customer Feedback Mechanisms	FY 2015	In Progress
Include point of contact in EDG for each data asset	11/30/13	Complete
Update EPA's Digital Strategy website to describe process to engage with customers	FY 2015 Q3	In Progress
Follow and update process as necessary	Ongoing	Ongoing
Participate in two hackathons to promote EPA data and APIs	FY 2014	Complete
Participate in at least two hackathons to promote EPA data and APIs	FY 2015	In Progress
Integrate the EPA Home Energy Yardstick Tool with Green Button	08/30/13	Complete
Integrate the EPA Portfolio Manager Tool, a component of the commercial side of the ENERGY STAR Program, with Green Button	2/11/15	Complete

V. Data Publication Process (and Exceptions)

The Open Data Policy requires agencies to develop and strengthen policies and processes to ensure that only appropriate data are released to the public and made available online. EPA must designate one of three "access levels" for each data asset (public, restricted public and non-public). Exceptions to publicizing data may result from law, regulation or policy, which address privacy, confidentiality, security or other valid restrictions.

EPA's data registration, evaluation and publication processes are being developed as part of the implementation of the Agency's Environmental Information Management Policy (EIMP) signed by the EPA Chief Information Officer (CIO) on February 27, 2015. A mandatory dataset cataloguing procedure is part of the EIMP package. Additional procedures and standard operating procedures, such as defining the details of the data publication process will follow.

The registration and classification of data assets will be addressed by the following:
- Issue and implement the EIMP, Cataloging EPA Data Resources Procedure and associated Standard Operating Procedures (SOPs) that require the registration of data assets.
- Create comprehensive processes to fill metadata gaps for existing records and ensure compliance with the registration of non-listed assets.
- Modify the EPA Metadata Editor to include a list of options for justifying why a dataset is classified as non-public or restricted public based on the Controlled Unclassified Information (CUI) Registry maintained by the National Archives and Records Administration.

U.S. Environmental Protection Agency
Open Data Policy Implementation Plan

EPA considers all information to be releasable, unless a statue, regulation or other federal government order requires its release to be controlled. EPA currently conducts reviews to evaluate the appropriateness of automatic non-release of information to the public under these conditions. To address the anticipated increase in demand for information, the Agency is developing a more formal process to document controlled data determinations. These controlled determinations are still subject to FOIA and may need additional evaluations by EPA's Office of General Counsel (OGC), which will be a part of the more formal process to determine which data can and cannot be released.

Data Publication Process/Exceptions Requirement	Due Date	Status
Develop Data Publication Process	11/30/13	Geospatial Data: Complete Non Geospatial: In Progress
Publish an overview of Data Publication Process on EPA's Digital Strategy website	11/30/13	In Progress
Update Publication Process	Ongoing	Ongoing

VI.   Roles and Responsibilities

EPA understands the need to clarify roles and responsibilities for promoting efficient release and management of data as an asset. EPA has identified leads for the following roles:

Role and Responsibility	POC and Contact Information
Communicating the strategic value of open data to internal stakeholders and the public	Ann Dunkin, Chief Information Officer (CIO), (202) 564-6404, dunkin.ann@epa.gov
Ensuring data released to the public are as open as appropriate	EPA's Draft Enterprise Information Management Policy includes processes to ensure Agency registration and publication of information assets are compliant and consistent with Open Data principles
Assisting with the use of open data and responding to EPA's implementation of open data requirements	EPA's Office of Environmental Information
Engaging entrepreneurs and innovators in the private and nonprofit sectors to encourage and facilitate the use of agency data to build applications and services	The Agency plans to use its Developer Central Program to engage and encourage entrepreneurs, innovators, and the nonprofit sector to use Agency data as they build applications and services.
Working with agency components to scale best practices that excel open data practices across the enterprise	Ann Dunkin, Chief Information Officer (CIO), (202) 564-6404, dunkin.ann@epa.gov
Update POC and contact information for each listed responsibility in MAX collect as part of the quarterly E-Gov IDC	Updated as needed