U.S. Environmental Protection Agency's Open Data Plan

July 2025

Introduction

The U.S. Environmental Protection Agency values our commitment to providing easy access to the data assets we collect on behalf of the American public, making them broadly available to all. By providing liberal access to our data assets, we empower a diverse array of stakeholders. These include federal partners, state and municipal policymakers, industry leaders, community organizations, and researchers in diverse disciplines. They can leverage EPA managed data and information to promote innovation, improve understanding and mitigation of risks, and ultimately, make more informed decisions. Data access fosters transparency and accountability, ensuring that environmental policies are grounded in evidence that can be scrutinized by the public. It also encourages collaboration, as partners can use EPA data to develop innovative solutions and strategies for tackling environmental and human health challenges. Findable, accessible, and useable data can ultimately help us reduce risks, to human health and the environment.

Access to data is not effective if the data are not understandable and useable. EPA recognizes the importance of delivering data in open formats and with robust descriptive and contextual information to enable a complete understanding of the data. By serving data in this way, EPA enhances usability by allowing stakeholders to interpret and apply it effectively to real world scenarios using freely available tools. As the use of artificial intelligence becomes more routine, additional attention must be paid to robust documentation so users can effectively leverage EPA managed data with AI to analyze ever larger volumes of data and deliver actionable insights. Data clarity, through robust documentation, also facilitates a deeper comprehension of the data's origins, methodologies, and limitations, which are essential for accurate analysis and evidence building and helps to draw meaningful conclusions that improve environmental outcomes.

Background

Title II of the Foundations for Evidence-Based Policy Making Act, or Evidence Act, established new requirements for data governance, data management, and transparent data processes.[1] The position of EPA Chief Data Officer was established in accordance with the Act to oversee the implementation of these requirements.[2] Pursuant to the Act and guidance provided by the U.S. Office of Management and Budget, EPA is publishing this Open Data Plan to describe our proposed activities for fulfillment of the goals regarding data management and publication.[3]

---

[1] Pub. L. No. 115-435 (2019). Title II of the Evidence Act is also referred to as the Open, Public, electronic, and Necessary (OPEN) Government Data Act ("OPEN Government Data Act").

[2] OMB M-19-23 provided guidance on the designation of CDOs and the establishment of a CDO Council. Off. of Mgmt. & Budget, Exec. Off. of the President, OMB M-19-23 at 21-25, *Phase 1 Implementation of the Foundations for Evidence-Based Policymaking Act of 2018: Learning Agendas, Personnel, and Planning Guidance* (July 10, 2019), www.whitehouse.gov/wp-content/uploads/2019/07/M-19-23.pdf.

[3] 44 U.S.C. § 3506(b)(2). OMB M-25-05 provided guidance on the contents of agency Open Data Plans. *See* Off. of Mgmt. & Budget, Exec. Off. of the President, OMB M-25-05, *Phase 2 Implementation of the Foundations for Evidence-Based Policymaking Act of 2018: Open Government Data Access and Management Guidance* (Jan. 15, 2025).

The EPA Open Data Plan addresses the following areas:

1. **Data Collection Processes to Support Open Formats**
2. **Data Usage Information**
3. **Collaboration with Data Users**
4. **Improvement Processes**
5. **Open Data Goal Requirements**
6. **Prioritizing Public Data Asset Review**
7. **Compliance Status (Comprehensive Data Inventory)**

1. Data Collection Processes to Support Open Formats

EPA is committed to ensuring that data assets are stored and accessible by the public in open formats. Currently, most data assets within the EPA Comprehensive Data Inventory, referred to through this plan as the "data inventory", are available in open formats or formats that are readily convertible to such formats. Through increased provisioning of open data, we facilitate improved integration and analysis across diverse platforms and applications, which in turn can foster collaboration and innovation. We acknowledge that there are still challenges to overcome. EPA still collects data in non-open formats, including paper-based submissions. Some of these data asset collections require multistep processes to transform data into open formats. Several factors may challenge our ability to alter data collections to ensure open formats. As we complete our inventory, we need to identify and address each specific challenge across all EPA non-open data asset collections and identify those assets to prioritize for open format collection.

Current Activities

- Continue to populate the EPA Enterprise Data Catalog and maintain the process to submit our metadata entries into Data.gov.
- Identify data assets, especially priority data assets, that are not collected, stored, or shared in open formats as part of our data inventory and governance activities.
- Develop language for contracts, grants and other agreements that specifies data submissions to EPA should be completed in open formats where possible.

Additional Planned Actions to Support Open Formats

- Identify different classes of collections that result in the collection of data in non-open formats, and develop appropriate strategies to collect, convert and serve data in open formats.
- Identify requirements, some unique to individual data assets, to promote collection, storage, and sharing of data in open formats.
- Assess the cost of altering existing processes to collect or convert data in non-open formats versus the value of providing EPA data assets in open formats.
- Prioritize data assets for conversion based on factors such as demand, potential impact, regulatory requirements, and ease of conversion. Focus on high-value data that will benefit the most from being in open formats.

- Develop and implement a strategy for modifying or developing processes (updating contract language, directives, etc.) that ensure targeted data collections are made in or converted to open formats.
- Develop comprehensive rules and guidelines to define processes to ensure open data format standards are applied and best practices are adopted.

2. Data Usage Information

EPA gathers information related to the access of many of our data assets. By understanding user interactions with EPA's data stores, the agency can gain valuable insights into which datasets are most in demand and adjust data provisioning strategies accordingly. This approach will not only enhance the relevance and accessibility of EPA's data offerings but also provide a framework for measuring progress in open data initiatives.

Current Activities

- Continue and enhance ability to track data views and downloads from EPA web areas.
- Ensure that metadata is assigned based on the latest EPA data standards at the earliest opportunity after data collection.

Additional Planned Actions to Support Understanding of Data Usage

- Leverage existing resources to track views and downloads of EPA-managed data assets, particularly the measurements made by the U.S. General Services Administration via Data.gov and the Digital Analytics Program.
- Develop additional mechanisms to track downloads and API usage by both internal and external parties.
- Develop a reporting mechanism to communicate which data assets are most frequently viewed and accessed over time.
- Identify opportunities to collect data citations, especially in the scientific research space, to further identify downstream uses of EPA collected research data.

3. Collaboration with Data Users

EPA regularly collaborates with a wide variety of external stakeholders. These collaborative efforts help foster direct information exchange, which improves how EPA serves our data users. These efforts include partnerships with state, municipal, and tribal governments, which help to ensure that we cooperatively achieve our statutory requirements. EPA also works with other federal agencies to combine data assets and address crosscutting issues such as the effects of wildfires and toxic releases on natural and human dominated environments. EPA also routinely collaborates with academic institutions to leverage outside expertise on using EPA managed data assets to address complex environmental and human health issues.

Current Activities

- Partnership in across-federal cooperative effort between the Environmental Protection Agency, the Department of Interior, the Department of Agriculture, and the Centers for Disease Control to leverage data to help protect communities from the impacts of wildfire smoke and promote land management practices that reduce the risk of large, severe fires.

- Collaboration with the Environmental Council of States, which provides engagement and exchange of views to enhance coordination between EPA and our state partners.
- Collaboration with academic institutions to address a wide variety of environmental challenges through targeted research.

Additional Planned Actions to Support Collaboration with Data Users

- Build an inventory of existing and planned collaboration with external partners to better capture the breadth or our collaboration.
- Expand efforts to build partnerships based on existing and future EPA priorities.
- Mature the EPA open data point of contact, which leverages EPA library and information management professionals, to serve as a conduit for identifying collaboration opportunities.

4. Improvement Processes

EPA has several initiatives in progress across the agency to improve the Findability, Accessibility, Interoperability and Reusability properties of EPA-managed data. Many of our governance processes exist within individual programs across the agency. We are actively working to standardize our governance by developing data and system lifecycle managment process to be used across EPA. As we develop and deploy more uniform data governance, the EPA Data Governance Advisory Committee will socialize them across EPA to enhance our data governance culture. This will help promote continual improvement of the open data principles surrounding the data we manage. As technology evolves in the AI space and beyond, we must also be mindful of how we may need to alter data management practices to take best advantage of new technologies.

Current Activities

- Developing a well-defined set of governance roles to help manage priority data assets across the agency.
- Developing an enterprise data model, starting with priority data assets across EPA, to better manage data assets for reuse across EPA and provide environments to manage data throughout the data lifecycle.

Additional Planned Actions to Support Improvement

- Develop and document standard requirements for data assets entries in the inventory.
- Develop routine processes to ensure metadata entry into the EPA data inventory.
- Revise existing and, where necessary, create standardized rules and processes for data management throughout the data lifecycle.
- Create standards for data domains and metadata to improve governance and help draw connections between EPA managed data and Agency missions and statutory mandates.
- Identify data assets that would be particularly useful for specific AI use cases and ensure that those data assets are aligned for ease of use by AI.

5. Open Data Goal Requirements

To meet the requirements of open data provisioning, EPA must ensure that our staff are skilled in managing data throughout the data lifecycle. Additionally, our technical infrastructure must provide the necessary tools to meet our open data requirements. EPA strives to provide training opportunities and convenes communities of staff to provide knowledge sharing and best practices discussions across various data-related disciplines.

Current Activities

- Continue to support the EPA Data Governance Community, championed by the EPA CDO, through regular meetings and activities to discuss and work on a broad variety of data-related topics, in order to strengthen the data manager, analyst and user connections through information exchange.
- Continue to populate the EPA Enterprise Data Catalog and maintain the process to submit our metadata entries into Data.gov.
- Continue to expand the EPA data skills training hub with a variety of data skills training resources that are available to all staff at any skill level.

Additional Planned Actions to Support Open Data Goal Requirements

- Develop an enterprise data model to enhance data governance and management across all agency components.
- Identify data skills training opportunities for staff at all levels, from beginning data users to expert data scientists.
- Work closely with the Chief Information Officer and Chief IT Architect throughout our IT governance process to ensure that IT asset procurements or developments are aligned with enterprise data management efforts.

6. Prioritizing Public Data Asset Review

Understanding what data assets are most valuable for dissemination to the public is an essential step in meeting open data requirements. At the time of creation of this plan, EPA submits 6,565 entries to Data.gov. For nearly all entries, EPA provisions those data to the public in open formats. EPA must continue to identify those data assets that we do not disseminate in open formats and that would be of highest value for public dissemination. Public input on which data assets to prioritize is a valuable aid in helping EPA make informed decisions about which data assets to prioritize for public access.

Current Activities

- Continue to populate the EPA Enterprise Data Catalog and maintain the process to submit our metadata entries into Data.gov.
- Identify those data assets that are not yet represented in our data inventory across all component offices, with an emphasis on high-value data assets, for inclusion in the catalog and dissemination where not already disseminated.

- As our open data concierge contact matures, build mechanisms to leverage that service to identify additional data assets requested by stakeholders to make accessible.
- Build stronger criteria for identifying additional data assets to prioritize for open data access.
- Develop processes to leverage Freedom of Information Act requests to identify additional data assets to prioritize for open data dissemination.
- Identify data assets that may be particularly useful for AI tools and strive to make those data open and AI ready.

7. Compliance Status

EPA will comply with all requirements of [OMB M-25-05](#) and the requirements of 44 U.S.C. § 3511 when disseminating public data assets pursuant to the Evidence Act.[4] Comprehensive data inventories and catalogs that aid in discovery are valuable tools enabling the ability to find and understand data holdings. In support of this, EPA shares its data inventory with the Federal Data Catalog, Data.gov, and will continue to do so moving forward. We also recognize the value of creating high-quality metadata in support of providing findable and accessible data assets in the context of a cross-federal data catalog. Without complete and accurate metadata, data becomes challenging to find, reference, or use for analysis and decision-making. To support these and all other data governance efforts EPA created a data advisory committee, comprised of data officers representing each EPA program and regional office, led by the EPA Chief Data Officer.

Current Activities
- Manage the EPA Enterprise Data Catalog and maintain the process to submit our metadata entries into Data.gov.
- Coordinate across all EPA components to ensure that our metadata entries are complete and provide sufficient information for end users to understand the asset, its context, and the appropriate uses for each data asset.

Additional Planned Actions to Support Compliance
- Coordinate with EPA's Chief Freedom of Information Act Officer, Senior Agency Official for Privacy, Chief Information Officer, General Counsel, Statistical Official, members of our Data Governance Advisory Committee, and other relevant Agency officials to ensure that any public release of data assets is lawful, appropriate, and adheres to data privacy requirements.
- Establish processes to ensure that entries in the EPA Enterprise Data Catalog and subsequent submissions to the Federal Data Catalog at Data.gov conform to the most recent U.S. Office of Management and Budget authorized metadata schema.

---

[4] [44 U.S.C. § 3511: Data inventory and Federal data catalogue](#).