EPA IT/IM DIRECTIVE POLICY

Enterprise Data Management Policy

Directive No: 2135.2

Issued by the EPA Chief Information Officer,
Pursuant to Delegation 1-19, dated 07/07/2005

Enterprise Data Management Policy (EDMP)

1. PURPOSE

The availability of data, information, and evidence of appropriate quality is critical to EPA's ability to make sound policy, regulatory, and program management decisions to protect human health and the environment and to manage agency programs and operations.

This Policy updates EPA's standard approach to developing, managing, and using data and information to support the development of evidence and evidence-based decision-making. It strengthens EPA's data and information management principles, practices, and standards to ensure data are discoverable, understandable, accessible, of known quality, and useable and re-useable by internal and external users seeking to conduct analyses and develop evidence.

This Policy incorporates the data and information planning, governance, and OPEN (Open, Public, Electronic, and Necessary)[1] data management requirements of the Foundations for Evidence-Based Policymaking Act of 2018 (Evidence Act or the Act); codifies the Chief Data Officer's (CDO) role for leading and coordinating EPA's data management functions as defined by the Act; and provides a framework for promoting an EPA culture of open, active, collaborative, and evidence-based data management.

2. SCOPE

This Policy replaces EPA's Enterprise Information Management Policy (EIMP). It updates EPA policy on data and information[2] planning, management, and governance based on the requirements of the Evidence Act. This Policy applies to:

- Data and information created, collected, and stored in an EPA system or repository or system or repository of an authorized third-party provider regardless of the data's source or format. It also applies to data created or collected using EPA funds and stored in this same manner.

---

[1] Title II of the Foundations for Evidence-Based Policymaking Act of 2018, the OPEN Government Data Act, defines OPEN data as data that is maintained in an Open format, is Publicly and Electronically available, and is Necessary or essential to accomplishing agency goals, objectives, or operating priorities. The current version of this document is available at OPEN Government Data Act.

[2] The OPEN Government Data Act defines the term data to include all information. OMB Circular A-130 defines the term information to include data. This Policy applies to both data and information without distinction and does not seek to reconcile these definitional differences. This Policy does clarify the definition of the term data for purposes of improving EPA data management. It also characterizes EPA's view of the data-to-information relationship, which it defines as: information is the product or result of analyzing data. The current version of this document is available at OMB Circular A-130.

- Lifecycle data and information management activities, including but not limited to planning, collecting, acquiring, developing, producing, processing, cataloging, preserving, storing, securing, using, evaluating, publishing, disseminating, and disposing of data.

This Policy is effective once issued as final by the Chief Information Officer (CIO) and Deputy Assistant Administrator (DAA) for Environmental Information (EI) within the Office of Mission Support (OMS). EPA's CDO may issue guidance to support its implementation. This Policy will be updated, where appropriate, to reflect new OMB guidance.

3.   AUDIENCE

The audience for this Policy is all EPA organizations, officials, and employees who collect, develop, produce, evaluate, and manage EPA data and information and who use these data and information to develop research, analysis, and evidence to support agency decision-making. This includes organizations, officials, and employees who perform traditional data management functions and activities (herein after referred to as EPA data managers) as well as non-IT staff and managers who perform programmatic, policy, administrative, financial, scientific, or other types of research or analysis.

The audience also is non-EPA organizations performing work on behalf of EPA under a contract, interagency, extramural, or other agreement. EPA's state, local, tribal, and territorial governmental partners are obligated to accept the provisions of this Policy when operating under a trading partner agreement or other codified form of interagency or interorganizational partnership, including but not limited to Federal grant programs. Non-EPA organizations include but are not limited to, contractors; regulated entities; cooperative agreement holders and grantees; state, interstate; territorial; tribal, and local organizations; intergovernmental organizations; educational institutions, hospitals, and non-profits; other Federal government agencies; and volunteer organizations and other data providers.

4.   BACKGROUND

In 2015, the EIMP established a standard enterprise approach to developing and managing EPA data and information. The goal of this standard approach was to make it easier for internal and external users to discover, understand, access, use, and re-use EPA data in a secure manner in applications and analyses. The EIMP applied to all EPA data and information regardless of the source, type, or format. The impetus for EPA to develop the EIMP was two policies issued in 2013 that required Federal agencies to make their data more open, accessible, and useable. Those policies were:

- OMB Memorandum M-13-13, "Open Data Policy - Managing Information as an Asset," May 9, 2013

- Executive Order 13642, "Making Open and Machine Readable the New Default for Government Information," May 9, 2013

In 2019, Congress passed the Foundations for Evidence-Based Policymaking Act of 2018 to significantly increase the amount of Federal data that is maintained in an open and machine-readable format and that is available and accessible to the public. The Act requires agencies to make their data open and accessible to the public by default and establishes the Freedom of Information Act (FOIA) as the standard for an agency to determine whether a data asset can be made publicly available.

FOIA is often described as the law that keeps citizen in the know about their government. Pursuant to the Evidence Act, Federal agencies are required to disclose any information requested under the FOIA unless it falls under one of nine exemptions, which protect interests such as personal privacy, national security, and law enforcement. (See Definitions for list of FOIA exemptions). The FOIA Act also requires data assets be maintained in an open format to support internal analyses of program, operations, research, and policy questions, even if the data is not appropriate for release to the public.

The Enterprise Data Management Policy (EDMP) applies the OPEN data requirements of the Evidence Act to both data and information. This approach is intended to address any potential concerns resulting from differences in the definitions of the terms data and information. This Scope also is consistent with EPA's characterization of the data-to-information relationship, which defines information as the product or result of analyzing data.

The Evidence Act also reforms how agencies use data, information, and evidence to support policy, regulatory, and program decision-making. EPA's Evaluations and Evidence-Building Policy will implement the Act's evidence-building provisions, while the EDMP addresses the data management and data-related evidence provisions of the Act.

The EDMP builds on the foundation of the EIMP while incorporating the new OPEN data and evidence-related data management requirements of Titles I and II of the Evidence Act and Federal policy supporting the Act. The current version of each of these titles is briefly described below:

- Title I: Federal Evidence-Building Activities Act. This title provides evidence-related data management requirements for this Policy.

- Title II: OPEN Government Data Act. This title provides the OPEN data management and enterprise planning and governance requirements of this Policy. (OPEN means data that is maintained in an Open format; is Publicly and Electronically available; and is Necessary or essential to agency goals, objectives, and operating priorities.)

- OMB Phase 1 Memo: "Implementation of the Foundations for Evidence-Based Policymaking Act of 2018", M-19-23 (hereafter referred to as OMB Phase 1 Guidance) is a source of policy requirements, including: creating an agency Data Governance Body, developing an agency Learning Agenda, and refining the role of the CDO.

- Title III of the Evidence Act is the Confidential Information Protection and Statistical Efficiency Act (CIPSEA). This Policy does not include CIPSEA requirements but may in future updates.

The Evidence Act also emphasizes disclosing and disseminating data assets with the public and other data users as a policy priority; stresses that the value of data comes from its usefulness in developing high-quality evidence to support decision-making; requires agencies engage with the public and other data user to improve their use of agency data; and establishes new data management and governance requirements to ensure agencies can better integrate their data management and evidence-building functions and activities.

5. AUTHORITY

The Enterprise Data Management Policy is issued under the following Authorities. These citations are valid at the time of issuance of this Policy. Since these documents are subject to periodic review, users of this Policy should refer to the most recent version.

- The Foundations for Evidence-based Policymaking Act of 2018, passed 2019.
- OMB Phase 1 Memo: "Implementation of the Foundations for Evidence-Based Policymaking Act of 2018", M-19-23 (hereafter referred to as OMB Phase 1 Guidance) is a source of policy requirements, including: creating an agency Data Governance Body, developing an agency Learning Agenda, and refining the role of the CDO.
- Open Data Policy - "Managing Information as an Asset", OMB Memorandum M-13-13, May 2013
- Federal Data Strategy
- Circular No. A-130 – Managing Information as a Strategic Resource
- Executive Order 13642, "Making Open and Machine Readable the New Default for Government Information," May 9, 2013
- EPA Enterprise Information Management Policy, Procedure, and Metadata Standard, August 2019

6. POLICY

Data Management

It is EPA policy to develop and manage data and information as OPEN assets and to use these assets as critical inputs in developing research, analysis, and evidence to support agency decision-making. EPA organizations, officials, employees, and non-EPA organizations who develop, produce, manage, evaluate, and use EPA data and information will

1. Plan and manage data and information (hereafter referred to as data) according to a lifecycle process that reflects OPEN data principles, practices, and standards (PPS), including the standard of making EPA data assets OPEN by default.[3]

---

[3] Pursuant to the Evidence Act, the standard of making data OPEN by default, includes the practices of 1) Maintaining data in an *Open format*; 2 & 3) Making data available *Publicly* in an *Electronic*, machine-

<u>New Data Assets</u>: EPA data managers will conduct lifecycle planning that considers how to apply OPEN data PPS to new data assets across all phases of the lifecycle. This practice applies to collecting, creating or developing, or acquiring new data assets and to developing new data collection systems, mechanisms and technologies.

<u>Existing Data Assets</u>: Assistant, Associate (AAs), and Regional Administrators (RAs) will document their organization's Priority Data Assets (PDAs) and will develop a proposed method for applying OPEN data PPS to these assets in accordance with standard agency formats and processes provided by EPA's information and mission support organization. EPA data managers will manage their existing PDAs in accordance with the organization's methods for applying OPEN data PPS and in accordance with the standard agency formats and processes.

EPA's OPEN Data Principles, Practices, and Standards include:

- Identifying and managing the organization's PDAs as OPEN
  - o PDAs are essential to the agency's ability to evaluate, develop-evidence, make decisions, or communicate on questions of strategic importance (See Definitions)
- Applying OPEN data PPS as consistently as possible across the AA-ship or Region to foster an enterprise approach and view of its PDA
- Planning and managing PDAs as OPEN to the maximum extent practical, balancing the benefits with the costs and risks
- Considering the following factors in assessing risks, costs, and benefits:
  - o Risk assessment (RA) starts with applying FOIA's disclosure exemptions (See Definitions) to identify any legal limits on disclosing a data asset
  - o RA includes assessing the risk a data asset could be combined with other assets already in the public domain to create a re-identification, confidentiality, or security concerns
  - o RA must consider whether releasing a PDA is limited by intellectual property rights under Titles 17 or 35, or by contract or written agreement
  - o Cost considerations include the cost of converting a non-OPEN data asset into an open, electronic, and machine-readable format
  - o Benefit considerations include assessing the current or potential public value of the data. Data that no longer aligns with an agency goal or priority will be of lower value to the public than data that does align
  - o <u>FAIR Principles - GO FAIR (go-fair.org)</u> can inform benefit/cost analyses
- The value of PDAs increases with their use and re-use. Cataloging and labeling PDAs with metadata is vital to maintaining their value and usability
- EPA's cataloging and metadata standards are foundational to making PDAs discoverable, understandable, and useful (Refer to Section 6.2).

---

readable format; and 4) Identifying those priority data assets that are *Necessary* or essential to the agency's ability to analyze, evaluate, develop-evidence, make decisions, or communicate on questions of strategic importance to its mission or operations.

The SIO, or other management official from an AA or RA's information and mission support organization, will lead and coordinate the organization's OPEN data management initiative on the AA or RA's behalf. The SIO will work with the SRO to ensure adequate staff and resources are provided for this effort.

2. Catalog and label agency new and existing data assets in EPA's authorized data catalog in accordance with EPA's Cataloging Procedure and Minimum Metadata Standard (Refer to Section 8, Related Information), as applicable, and the metadata requirements of this Policy, including:
   - A description of the data asset, including all variable names and definitions
   - The original source(s) from which the data asset is derived
   - Whether the original data asset has been modified and the purpose and scope of the modifications
   - The results of the FOIA assessment to determine whether the data asset is
     - Disclosable as an open Government data asset
     - Subject to full or partial disclosure or exempt under FOIA
     - Not subject to open format or license requirements due to restrictions on the government's disclosure of the data
     - The date of EPA's FOIA disclosure assessment or planned disclosure assessment, or an indication of whether an assessment can be made.

EPA data managers also will update the metadata for a data asset, as appropriate, to maintain its currency, accuracy, and relevancy throughout the lifecycle. (Refer to Section 7: Roles and Responsibilities, EPA Data Managers role, for additional information on the responsibilities for cataloging new and existing data assets.)

3. Develop and maintain an authorized catalog of EPA's new and existing PDAs. The catalog will be a comprehensive inventory of the PDAs the agency creates, collects, controls, or maintains in EPA systems or repositories, or in systems or repositories maintained by authorized third parties. The CIO/DAA and CDO are responsible for creating and maintaining the catalog and for ensuring compliance this Policy's cataloging requirements. The CIO/DAA and CDO will:

   - Operate and maintain the authorized data catalog to ensure EPA data managers can appropriately catalog their PDAs and label them with metadata.
   - Ensure the catalog's data assets are comprehensive of the agency's PDA holdings. The AA and RA's PDA assessments (Refer to Section 6.1), as well as new data acquisitions (creation, collection, and procurement of new data) will provide the basis for determining agency's baseline for the inventory.
   - Ensure these data assets are labeled with current and appropriate metadata in accordance with this Policy (Refer to Section 6.2), including labeling geospatial data assets in accordance with the National Geographic Data Procedure. (Refer to Section 8, Related Information).
   - Review and revise on a biennial basis the scope and requirements for cataloging EPA data assets, as necessary, to ensure important EPA data assets are included in the catalog.

- Maintain the catalog to enable Data.gov to collect and publish EPA public data assets as open Government data assets.
- Modernize the catalog to leverage opportunities to improve data discovery, accessibility, understandability, and useability for internal and external users; address new data assets and metadata requirements; and address new data collections methods, systems, and technologies.

4. Strive to improve the management of EPA data assets by implementing best data management practices to the extent practical; implementing existing data standards where applicable: and developing new data standards, where appropriate, to support data sharing, integration, use, and exchange. EPA data and systems managers will

- Implement best data management practices including evaluating and benchmarking existing data management practices and implementing best data lifecycle management practices where applicable.
- Implement existing EPA data standards to support data integration and interoperability. If a suitable EPA data standard does not exist, then collaborate with EPA's partners to adopt, adapt, develop, and implement an existing consensus-based international, national, federal, or industry, or consortium data standard in accordance with EPA's Data Standards Policy and Procedures for new applications, systems, or data collections mechanisms.

Data Governance and Planning

It is EPA Policy to leverage enterprise data planning and governance to ensure its data and information needs are met and that these assets support agency goals and priorities. EPA will implement data governance and planning through leadership, coordination, and collaboration with internal and external stakeholders. Pursuant to the Evidence Act:

5. EPA's CDO will lead and coordinate the agency's data and information governance function with the support of EPA's Data Governance Council (DGC) and CIO Senior Advisory Committee (CIO-SAC).

6. EPA will develop and maintain a Strategic Information Resources Management (IRM) Plan that describes how information resources support the mission and how data collection and management activities will support the development and implementation of EPA's Learning Agenda and Evaluation Plan. The CIO/DAA will lead EPA's strategic IRM planning process and function in coordination with the CDO, SIOs, Chief Technology Officer, Evaluations Officer, Statistical Officer, all AA-ships and Regions, and key external stakeholders.

7. EPA will develop and maintain an Open Data Plan as a part of EPA's Strategic IRM Plan. The CDO will lead the development and implementation of EPA's Open Data Plan in coordination with the DGC, CIO-SAC, and EPA stakeholders. Pursuant to the Act, the Plan will address EPA's plans to develop processes and procedures that:

- Require new data collection mechanisms be available in an open format

- Actively engage the public, businesses, and other external stakeholders and collaborators to improve their use of EPA data. To inform this public engagement process, EPA's Open Data Plan also will
  - o Describe its plans to collect, analyze, and publish information on the usage of EPA data by both external and internal data users
  - o Provide the public and all external collaborators with an agency point of contact to assist and respond to questions and concerns about the quality, accessibility, usability, and openness of EPA data
  - o Develop and implement a process to evaluate and improve the timeliness, completeness, consistency, accuracy, usefulness, and availability of EPA's open Government data assets
  - o Identify the agency's PDAs and develop a plan to evaluate these assets for disclosure in the EPA's Data Catalog and on Data.gov.

  The Plan also will address the requirements of new OMB Open Data Plan guidance.

8. This Policy codifies the CDO's leadership and coordination responsibilities in supporting the implementation of the Evidence Act. Pursuant to Title 44 USC Section 3520 and OMB's Phase 1 Guidance, EPA's CDO is responsible for:

- EPA Lifecycle Data Management, including

  - o Implementation of best data management practices where practical
  - o Developing and maintaining EPA's Open Data Plan and supporting agency strategic IRM planning activities
  - o Carrying-out the requirements for data collection and paperwork control
  - o Leading agency information dissemination activities and requirements
  - o Managing EPA's data assets, including leading on data sharing and publishing, and standardizing data formats to support these activities
  - o Carrying-out records management.

- EPA Data Governance and Support Function around Finding and Using Data:

  - o Leading and coordinating EPA's data governance process and activities
  - o Supporting EPA's Learning Agenda by coordinating data access and management activities that support evidence building activities
  - o Coordinating with EPA officials responsible for generating, using, protecting, and disseminating data to ensure that EPA's data needs are met
  - o Engaging EPA employees, the public, businesses, contractors, and other external stakeholders on how to effectively use EPA public data assets and support collaborative approaches to improving the use of these assets
  - o Supporting the Performance Improvement Officer in identifying and using data to implement performance planning, measurement, and assessments
  - o Supporting the Evaluation Officer in obtaining data to carry-out evaluation functions and activities, including development of EPA's Learning Agenda and Evaluation Plan
  - o Ensuring EPA maximizes its data to support the production of evidence, including research, regulatory analysis, cybersecurity, and EPA operations
  - o Serve as EPA's liaison to OMB on data matters and EPA's member on the Federal CDO Council.

7. ROLES AND RESPONSIBILITIES

Only the roles and responsibilities directly related to this Policy are outlined in this section.

Chief Information Officer (CIO):
- Issues this IT/IM policy and its supporting procedures, standards, and guidance
- Empowers the CDO to lead the implementation of this Policy
- Coordinates with the CDO on implementation and compliance matters
- Maintains EPA's enterprise data catalog in coordination with the CDO
- Leads EPA's Strategic IRM Planning Process

Chief Data Officer (CDO):
- Leads and coordinates the agency's data and information governance function with the support of EPA's Data Governance Council (DGC) and CIO Senior Advisory Committee (CIO-SAC) -- including considering the DGC's advice on data planning, governance and management policies, plans, strategies, and initiatives, and coordinating with the CIO/DAA and CIO Senior Advisory Committee (CIO-SAC) on adopting and/or approving these and other data and information policies, plans, strategies, and initiatives
- Chairs DGC and coordinates with EPA officials responsible for using, protecting, disseminating, and generating data to ensure EPA's data needs are met.
- Coordinates with the CIO/DAA on developing and updating this Policy and its implementing procedures, standards, and guidance Leads EPA's implementation of this Policy, including
  - o Promoting and monitoring agency compliance with the policy
  - o Coordinating with EPA organizations to incorporate the EDMP in their internal policies, procedures, standards, and guidance
  - o Providing tools, templates, guidance, and technical advice and assistance to support its implementation
- Implements CDO responsibilities pursuant to Title 44 USC Section 3520 and OMB Phase I Guidance (Refer to Section 6.8) to strengthen EPA data planning, management, governance, and evidence-building activities

CIO Senior Advisory Council (CIO-SAC):
- Advises CIO/DAA and CDO on IT/IM policy, procedures, and standards
- Advocates for timely implementation of these policies and procedures
- Reviews and advises on progress and challenges with policy implementation

Assistant and Associate Administrators (AAs) and Regional Administrators (RAs):

- Promote the strategic value of quality data and information in developing evidence to support agency and program decision-making
- Empower the SIO, or other senior official in their information and mission organization, to lead the organization's senior leadership team in developing and implementing an EDMP implementation plan
- Provide the resources necessary to effectively implement and monitor compliance with the EDMP Policy
- Monitor the organization's progress implementing the EDMP and make strategic course-corrections as necessary

Senior Resource Officials (SROs):

- Advise AA/RA on the resources necessary to effectively implement this Policy
- Coordinate with the SIO and organization's senior leadership team to assess the resource requirements necessary to fully implement this Policy
- Ensure the organization's acquisition plans align and support EPA's Learning Agenda and Evaluation Plan

Senior Information Officials (SIOs):

- Promote the strategic value of quality data and evidence in decision-making Lead and coordinate the organization's OPEN data management initiative on the AA or RA's behalf, including developing measures and controls to oversee progress
- Coordinate with the SRO to ensure adequate staff and resources are provided for this effort
- Report to the AA or RA, CDO, Evaluations Officer, and other appropriate officials on the progress and challenges with implementing this Policy
- Coordinate across the organization to ensure data collection and management plans and activities align with priority questions in the EPA Learning Agenda and Evaluation Plan
- Coordinate with the organization's DGC representatives on the development and implementation of data management and governance plans and initiatives
- Advise and concur on requests for waivers to this Policy or its procedures

Data Governance Council (DGC):

- Advise and assists the CDO on enterprise data governance and strategic data management matters in accordance with its charter and OMB Phase 1 Guidance. The DGC will advise the CDO on:
  - facilitating coordinated and collaborative approaches to
    - data collections and acquisitions
    - lifecycle data management processes and procedures
    - cataloging and metadata management
    - data sharing and dissemination
  - improving data interoperability
  - implementing the open/transparent release of EPA data assets

- advancing the use of EPA data both internally and externally
- supporting and promoting an open data culture
- improving the use of data to support evidence-based decision-making, and
- identifying data and information to support evidence development and implementation of EPA's Learning Agenda and Evaluation Plan.
- Members will establish a dialogue with their organization's SIO and IMO on DGC plans and activities and will serve as a leader in their organization on data management and governance matters.

Information Management Officers (IMOs):

- Promote the strategic value of quality data and evidence in decision-making
- Review, comment, and concur on EPA data and information management policy, procedures, standards, and guidance
- Ensure staff and managers across the organization understand this Policy
- Support the SIO in overseeing the implementation of the EDMP
- Share and report on progress and lessons learned implementing this Policy
- Collaborate with the organization's DGC representatives on developing and implementing data management and governance plans and initiatives
- Coordinate with the staff and managers cross the organization to ensure the data requirements of the EPA's Learning Agenda and Evaluation Plan are understood and appropriately supported
- Advise and concur on requests for waivers to this Policy or its procedures

EPA Data Managers (includes all EPA and non-EPA organizations, employees or individuals with data management responsibilities):

- Plan and manage data and information according to a lifecycle process that reflects OPEN data principles, practices, and standards
- Catalog and label agency data assets in EPA's authorized data catalog in accordance with EPA's Cataloging Procedure, Minimum Metadata Standard, and this Policy. New data assets will be cataloged when created, collected, or acquired. Existing PDAs will be cataloged according to the organization's method for making these data assets OPEN
- EPA data managers will update the metadata for a data asset, as appropriate, to maintain its currency, accuracy, and relevancy throughout the lifecycle
- Strive to improve the management of EPA data assets by implementing best data management practices to the extent practical; implement existing data standards where applicable: and develop new data standards, where appropriate, to support data sharing, integration, use, and exchange

Director of EPA's Acquisition Organization:

- Ensures that applicable EPA Requests for Proposals and contracts incorporate a provision requiring compliance with applicable information management policies, procedures, and standards, including the EDMP.

Director of EPA's Grants and Disbarment Organization:

- Ensures that applicable EPA Requests for Proposals and conditions of grants, interagency agreements, and cooperative agreements incorporate a provision requiring compliance with applicable information management policies, procedures, and standards, including the EDMP.

EPA Program Managers:

- Promote the strategic value of quality data and evidence in decision-making
- Ensure the program's data managers are aware and implement this Policy
- Develop and implement a plan for carrying-out the requirements of this Policy
- Implement controls to measure the program's implementation progress
- Ensure program and data management staff and managers collaborate in the planning, identification, access, creation, and management of high-quality data to support evidence-building projects and activities

- See EPA Data Manager role

8. RELATED INFORMATION

These citations are valid at the time of issuance of this Policy. Since these documents are subject to periodic review, users of this Policy should refer to the current version.

- CIO Procedure 2135-P-01-0, Enterprise Information Management Policy (EIMP) Cataloguing Information Resources Procedure
- CIO Standard 2135-S-01.0, Enterprise Information Management Policy (EIMP) Minimum Metadata Standards
- CIO 2131-P-01-0, National Geographic Data Procedure
- Freedom of Information Act (FOIA), EPA Freedom of Information Act Guidance
- CIO 2105.1, Environmental Information Quality Policy
- EPA Evaluations and Evidence-Building Policy

The CIO/DAA and CDO may issue procedures, standards, guidance, tools, and templates to support the implementation of this Policy.

9. DEFINITIONS

Accessible: Open data are made available in convenient, modifiable, and open formats that can be retrieved, downloaded, indexed, and searched. Formats should be machine-readable (i.e., data are reasonably structured to allow automated processing). Open data structures do not discriminate against any person or group of persons and should be made available to the widest range of users for the widest range of purposes, often by providing the data in multiple formats for consumption. To the extent permitted by law,

these formats should be non-proprietary, publicly available, and no restrictions should be placed upon their use. Open Data Policy

Data: The term data means recorded information, regardless of form or the media on which the data is recorded. Foundations for Evidence-based Policymaking Act 2018. EPA further defines data as a quantitative or qualitative representation of values, facts, observations, or ideas in a formalized manner capable of being transmitted, processed, stored, analyzed, interpreted, and/or communicated by some process, whether on paper or in electronic form.

- Qualitative data—is descriptive.
- Quantitative data—is numerical.
- Primary data—are data observed, collected, stored, or generated directly for a specific purpose.
- Existing data—are data that have been collected, derived, stored, or reported in the past or by other parties (for a different purpose and/or using different methods and quality criteria). Sometimes referred to as data from other sources.
- Metadata— Metadata is structured information that describes, explains, locates, otherwise makes it easier to retrieve, use, or manage an information resource.

Data-to-information relationship: This Policy reflects EPA's characterization of the data-to-information relationship, which defines information as the product or result of analyzing data. Source: EPA's Chief Data Officer

Data Asset: A collection of data elements or datasets that make sense to group together. Each community of practice identifies the data assets specific to supporting the needs of their respective mission or business functions. Notably, a data asset is a deliberately abstract concept. A given data asset may represent an entire database consisting of multiple distinct entity classes or may represent a single entity class. A data asset may also be a data or information product. Source: Glossary for resources.data.gov

Dataset: 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 Extensible Mark-up Language (XML) file, a geospatial data file, or an audio, video, image, or text-based files. Source: Glossary for resources.data.gov

EPA Data Manager: The requirements of this Policy apply to a wide range of EPA staff and managers with lifecycle data management responsibilities. This Policy uses the generic title of "EPA data manager" to simplify references to these staff and managers.

Enterprise Data Management (EDM): Enterprise Information Management (EIM) is defined as an integrative discipline for structuring, describing, and governing information assets across organizational and technological boundaries to improve efficiency, promote transparency, and enable business insight. This policy adopts this definition while substituting the term "data" for "information". Gartner Glossary

Evaluation Plan: Consistent with OMB Phase 1 Guidance, an Evaluation Plan describes the evaluation activities the agency intends to conduct based on its Learning Agenda and includes the key questions to be addressed by each significant evaluation study or phase of a study the agency plans to begin in the next fiscal year. OMB Phase 1 Guidance

Evaluation and Program Evaluation: OMB defines evaluations and program evaluations as in 5 U.S.C. 311(3) to mean "an assessment using systematic data collection and analysis of one or more programs, policies, and organizations intended to assess their effectiveness and efficiency". Phase 4 Implementation of the Foundations for Evidence-based Policymaking Act of 2018

Evidence: OMB broadly defines evidence as "the available body of facts or information indicating whether a belief or proposition is true or valid. As such, evidence can be quantitative or qualitative and may come from a variety of sources, including foundational fact finding (e.g., aggregate indicators, exploratory studies, descriptive statistics, and other research), performance measurement, policy analysis, and program evaluation. OMB Phase 1 Guidance

Evidence-building activities: OMB defines evidence-building activities to include program evaluation; policy analysis; foundational fact finding; performance measurement; data collection, acquisition, and management; research plans, research roadmaps, enterprise learning agendas, or evaluation strategic plans; and planning and managing or conducting specific evaluations and other evidence-building activities and bringing evidence to bear in decision-making. OMB Phase 1 Guidance

Freedom of Information Act (FOIA): FOIA provides the public the right to request access to records from any Federal agency. It is often described as the law that keeps citizen in the know about their government. Federal agencies are required to disclose any information requested under the FOIA unless it falls under one of nine exemptions, which protect interests such as personal privacy, national security, and law enforcement. The nine exemptions are:

1. Classified national defense and foreign relations information
2. Internal agency rules & practices
3. Information prohibited from disclosure by federal law
4. Trade secrets and other confidential business information
5. Inter-agency or intra-agency communications protected by legal privileges
6. Information involving matters of personal privacy
7. Information compiled for law enforcement purposes
8. Information relating to the supervision of financial institutions
9. Geological and geophysical data, including maps, concerning wells.

Source: EPA Learn about FOIA

Information: Any communication or representation of knowledge such as facts, data, or opinions in any medium or form, including textual, numerical, graphic, cartographic, narrative, or audiovisual forms. OMB Circular A-130

Information or Data Lifecycle: The stages through which information or data passes, including creation or collection, processing, dissemination, use, storage, and disposition. (This Policy adopts the same definition while using the term data and information interchangeably.) Source: Glossary for resources.data.gov

Learning Agenda: A Learning Agenda identifies the priority policy questions an agency will seek to answer to support effective policy, regulatory, and program development and decision-making. A Learning Agenda is a multi-year document that serves as a driver for an agency to identify and plan the evaluations and data collection activities necessary to answer priority policy questions. OMB Phase 1 Guidance

Machine-Readable: Refers to data that is in a format that can be easily processed by a computer without human intervention while ensuring no semantic meaning is lost. Source: Glossary for resources.data.gov

Metadata: Structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource.
Foundations for Evidence-based Policymaking Act 2018

OPEN data: The letters of this acronym mean: data that is in an Open format; is Publicly available in Electronic format; and is Necessary or essential to an organization's mission.
Foundations for Evidence-based Policymaking Act 2018

Open Government Data Asset: A public data asset that is
A. machine-readable
B. available (or could be made available) in an open format
C. not encumbered by restrictions, other than intellectual property rights, including under titles 17 and 35, that would impede the use or reuse of such asset; and
D. based on an underlying open standard that is maintained by a standards organization.
Foundations for Evidence-based Policymaking Act 2018

Open Format: Open format means a non-proprietary file format that has a published specification that is maintained by a standards body or a community driven process and has no restrictions placed upon it use. Wikipedia

Open Standard: Use of a standard developed or adopted by voluntary consensus standards bodies, both domestic and international. These standards include provisions requiring that owners of relevant intellectual property have agreed to make that property available on a non-discriminatory, royalty free or reasonable royalty basis to all interested parties. Source: Glossary for resources.data.gov

Open License: The term open license means a legal guarantee that a data asset is made available at no cost to the public and with no restrictions on copying, publishing, distributing, transmitting, citing, or adapting such asset.
Foundations for Evidence-based Policymaking Act 2018

Priority Data Assets: Data and information that have significant value to the public, or which are essential to EPA's efforts to research, analyze, evaluate, develop-evidence,

make decisions, or communicate on questions of strategic importance to the agency (e.g., data that is relevant to enterprise or regional goals, objectives, priorities, or emerging environmental issues or risks such as Covid-19 and PFAS), and therefore should be managed as OPEN. Data assets that are or have been determined to be ancillary to addressing questions of strategic importance to the agency are not considered Priority Data Assets. (Source: EPA's Data Governance Council, Inventory Subcommittee)

Public Data Asset: The term public data asset means a data asset, or part thereof, maintained by the Federal Government that has been, or may be, released to the public, including any data asset, or part thereof, subject to disclosure under FOIA. Foundations for Evidence-based Policymaking Act 2018

Unstructured data: Unstructured data (or information) is information that either does not have a pre-defined model or is not organized in a pre-defined manner. Unstructured information is typically text-heavy, but may contain data such as dates, numbers, and facts as well. This results in irregularities and ambiguities that make it difficult to understand using traditional programs as compared to data stored in fielded form in databases or annotated (semantically tagged) in documents. Wikipedia

Semi-structured data: Semi-structured data is a form of structured data that does not obey the tabular structure of data models associated with relational databases or other forms of data tables, but nonetheless contains tags or other markers to separate semantic elements and enforce hierarchies of records and fields within the data. Therefore, it is also known as self-describing structure. In semi-structured data, the entities belonging to the same class may have different attributes even though they are grouped together, and the attributes' order is not important. Examples of semi-structured include emails organized by inbox, sent, draft, or other category, or folders that are organized by subject title or topic. Wikipedia

10. WAIVERS

- Waiver Authority: EPA's CIO/DAA may grant waivers (directly or via delegated authority) to selected provisions of this policy for sufficient cause
- Application Requirements: Applications for waivers to specific provisions must be submitted to the CIO/DAA and shall contain:
  o Identification of the policy provision
  o Detailed justification why the policy cannot be applied or maintained
  o An assessment of the impact from non-compliance
  o The signature of an Assistant or Associate Administrator (AA) or Regional Administrator (RA) responsible for the information in question
- Notification:
  o The CIO/DAA will notify the requesting office in writing of the disposition of the waiver within 60 days of receipt
  o The CIO/DAA also will make the rationale for the waiver public unless this disclosure is prohibited by law or Federal policy.
- Exception: There is no waiver to develop a method for making existing PDAs publicly available.

11. MATERIAL SUPERSEDED

This policy supersedes the CIO Policy 2135.1, Enterprise Information Management Policy.

12. CONTACTS

For additional information about this policy, please contact the EPA Office of Mission Support, Office of Digital Services and Technical Architecture.

Vaughn Noga
Deputy Assistant Administrator for Environmental Information
and Chief Information Officer
U.S. Environmental Protection Agency

APPENDIX A:
ACRONYMS & ABBREVIATIONS

AA	**Assistant or Associate Administrator**
CFR	**Code of Federal Regulations**
CDO	**Chief Data Officer**
CIO	**Chief Information Officer**
CIO-SAC	**CIO Senior Advisory Committee**
CTO	**Chief Technology Officer**
DAA	**Deputy Assistant Administrator**
DGC	**Data Governance Council**
EDMP	**Enterprise Data Management Policy**
EI	**Environmental Information**
EIMP	**Enterprise Information Management Policy**
EPA	**Environmental Protection Agency**
EV	**Foundations for Evidence-Based Policymaking Act or Evidence Act**
FOIA	**Freedom of Information Act**
IM	**Information Management**
IMO	**Information Management Officer**
IT	**Information Technology**
OMB	**Office of Management and Budget**
OPEN	**Open, Public, Electronic, and Necessary**
RA	**Regional Administrator**
SIO	**Senior Information Officer**