EPA Environmental Dataset Gateway (EDG)

Standard Operating Procedure and Governance Document Version 2.3

DRAFT 9/23/2011

1. PURPOSE

This document identifies the oversight and authority organization of the EPA Environmental Dataset Gateway (EDG) and provides an overview of roles and responsibilities for implementing the EDG. Roles and responsibilities outlined within this document align upwardly with the U.S. Environmental Protection Agency's (EPA's) Enterprise Data Policy (EDP), Procedures, and related documents. This document complements the role structure outlined within the EDP Procedure for Metadata Management by clearly defining the relationship between EDP Procedure roles and EDG contribution and management responsibilities expected of the individuals fulfilling those roles.

2. SCOPE AND APPLICABILITY

Because the EDG is a central metadata management application that consolidates distributed information into a single location, a proper governance structure is important for outlining responsibility and authority for assets. Roles and responsibilities that are clearly outlined and aligned with EPA's EDP and consistently implemented across EPA programs are important for ensuring that information contributed to the EDG is effectively managed and can be traced to owners. Additionally, identifying this information will enable stakeholders to clearly understand requirements and expectations for participation in the EDG. Agreement to specific oversight responsibilities must be in place in order to make the EDG a useful resource. For additional information on the EDG, please visit the EDG Web site at https://edg.epa.gov/.

3. INTENDED AUDIENCE

The primary audience for this document includes EPA's Metadata Coordinator, Data Owners, Data Stewards, EDG Administrator, and EDG Stewards for identifying their roles and responsibilities for contributing and managing information within the EDG. This document may also be of interest to any additional EPA personnel or EPA contractor staff interested in using the EDG to meet their resource needs.

The secondary audience for this document includes EPA's owners and stewards of non-geospatial data. The EDG incorporates all datasets contributed to Data.gov. This document is a resource for anyone with an interest in EPA metadata.

4. BACKGROUND

The EDG is EPA's central metadata access point. It is an enterprise application that stores metadata about geospatial and non-geospatial assets maintained by Data Stewards across the Agency. The purpose of the EDG is to: 1) provide EPA staff with a single access point to EPA's data assets from various EPA Program and Regional Offices, and 2) meet legislative and regulatory mandates and Agency requirements for metadata compliance, sharing, and management. As the central location for distributed data resources, the EDG helps support external federal initiatives such as Data.gov.

The EDG is comprised of a set of Web-based interfaces and a metadata catalog. Information is contributed to the EDG catalog using a process called 'harvesting'. Harvesting allows distributed EPA offices to contribute metadata to the EDG central catalog using an automated, scheduled process. Providing an automated connection to the central catalog removes the burden of metadata maintenance from Data Stewards by ensuring that their EDG records are automatically refreshed along with their local records. EDG Stewards who choose to contribute to the EDG using harvesting only need to maintain records locally, and information stored in the EDG catalog is synchronized with the local files on a scheduled basis (e.g., weekly, bi-weekly, monthly).

5. AUTHORITY

General governance and oversight for the EDG is provided by EPA's Quality Information Counsel (QIC) in coordination with the Geospatial Information Officer (GIO). The QIC is responsible for overseeing the implementation of the EDP Policy, Procedure, and related documents among EPA programs. It will also act as the governing authority for questions related to EDG costs, implementation or structure across programs.

Senior Information Officials (SIO's) are responsible for decisions regarding costs and managing metadata for their individual programs. The SIO has responsibility for ensuring that the roles listed in this document are implemented appropriately within their program. The implementation of roles is expected to vary by organization; procedures for appointing individuals to fulfill roles (excluding the EDG Steward) are listed in the EDP Procedure for Geospatial Metadata Management. Information regarding the EDG Steward role (including appointing a steward) is described in Section 8.5.1. Additional information regarding the roles and responsibilities of the QIC and SIO are available at the following locations:

- http://intranet.epa.gov/oei/imitpolicy/qic/index.htm
- http://intranet.epa.gov/oei/imitpolicy/qic/ciopolicy/2101.pdf

If costs or questions arise that extend beyond the scope of an individual organization, the organization's SIO should consult the QIC. Costs for EDG enterprise resources, EDG training and application oversight are managed by EPA's Office of Environmental Information (OEI). Program offices may decide the degree to which they would like to be engaged in more specific oversight of the EDG.

6. RELATED DOCUMENTS

- Federal Initiatives
  - *E-Government Act of 2002 (H.R. 2458/S. 803)*, December 17, 2002.
  - *Data.gov Metadata Template, version 2.0*
  - *Executive Order 12906, National Spatial Data Infrastructure*, April 13, 1994, Federal Register, Volume 59, Number 71, pp. 17671 – 17674.
  - *Federal Geographic Data Committee Standard (FGDC-STD) 001-1998, Content Standard for Digital Geospatial Metadata, Federal Geographic Data Committee*, June 1998.
  - *Office of Management and Budget (OMB) Circular A-16, Coordination of Geographic Information and Related Spatial Data Activities*, August 19, 2002.
- EPA Policies, Procedures, Standards, and Guidances

- *EPA Geospatial Metadata Technical Specification, Version 1.0*, January, 2, 2007
- *EPA Guidance, Enterprise Data Policy Implementation Guidance, Office of Environmental Information* (*In Progress*)
- *National Geospatial Data Policy, EPA Order 2121*, August 7, 2005

7. PROCEDURES FOR CONTRIBUTING TO THE EDG

This section provides detail on metadata development and sharing processes for operations specifically tied to the EDG. It outlines steps that may be taken to implement the primary EDG contribution responsibilities described in Section 7.5. Users wishing to obtain more detail on metadata management outside of EDG operations should consult the EDP Procedure for Geospatial Metadata Management.

Establishing a framework within each organization for contributing to the EDG usually includes a standard set of process steps (Figure 1). These steps are intended as guidance for EDG Stewards and may change based on a particular organization's configuration. In general, EDG contribution process steps typically include:

- The Data Owner recommends an EDG Steward(s) for their organization. The EPA Metadata Coordinator and/or EPA EDG Administrator are notified of the selection as appropriate.
- The EDG Administrator (or the EDG Administrator's support team) provides an overview of EDG contribution expectations and requirements to the EDG Steward, including an overview of available tools.
- The EDG Steward works with Data Stewards assigned for their organization to create an EDG Inventory that identifies which data and/or other assets will have metadata in the EDG. Stewards should outline all assets that their organization produces that should be documented and shared. The Inventory should be prioritized if possible and indicate where metadata already exists. The Inventory will serve as a master asset list that will be used to guide metadata development.
- The EDG Steward determines which metadata standard to use. The type of data being documented (geospatial data, non-geospatial data, services, etc.) will determine whether the steward uses the FGDC CSDGM, the Data.gov template, the ISO standard, or other standards as applicable.
- The EDG Steward reviews the appropriate metadata template (with the EDG Administrator as appropriate) and has the option to create a metadata template for his/her organization. The EDG Steward is encouraged to use the EPA Metadata Editor for documenting assets and creating a template.
- The EDG Steward works with the EPA Data Steward to begin creating and/or updating metadata for his/her organization. This step usually involves coordination among the EDG Steward and multiple Data Stewards.
- The EDG Steward identifies the method(s) of access for information being documented (e.g., live data and services, downloadable data, etc).
- The EDG Steward contributes to the EDG via upload, publish and/or harvest.
- The EDG Steward verifies harvesting reports and validation results. For geospatial metadata the EME and EDG can be used for validation. For non-geospatial metadata for data sets, the EDG can be utilized for validation.
- The EDG Administrator verifies quality of the newly-submitted metadata records. The EDG Administrator approves or denies records as appropriate.
- The EDG Steward identifies which records should be made available to the public by flagging records as unrestricted.
- The EDG Steward oversees local file synchronization with the EDG metadata catalog by removing records from the EDG catalog if they are deleted locally.
- The EDG Steward manages and maintains assets in the EDG by regularly reviewing records.

Metadata should be updated according to the maintenance schedule outlined for the data set. Outdated metadata should be archived/retired along with data.
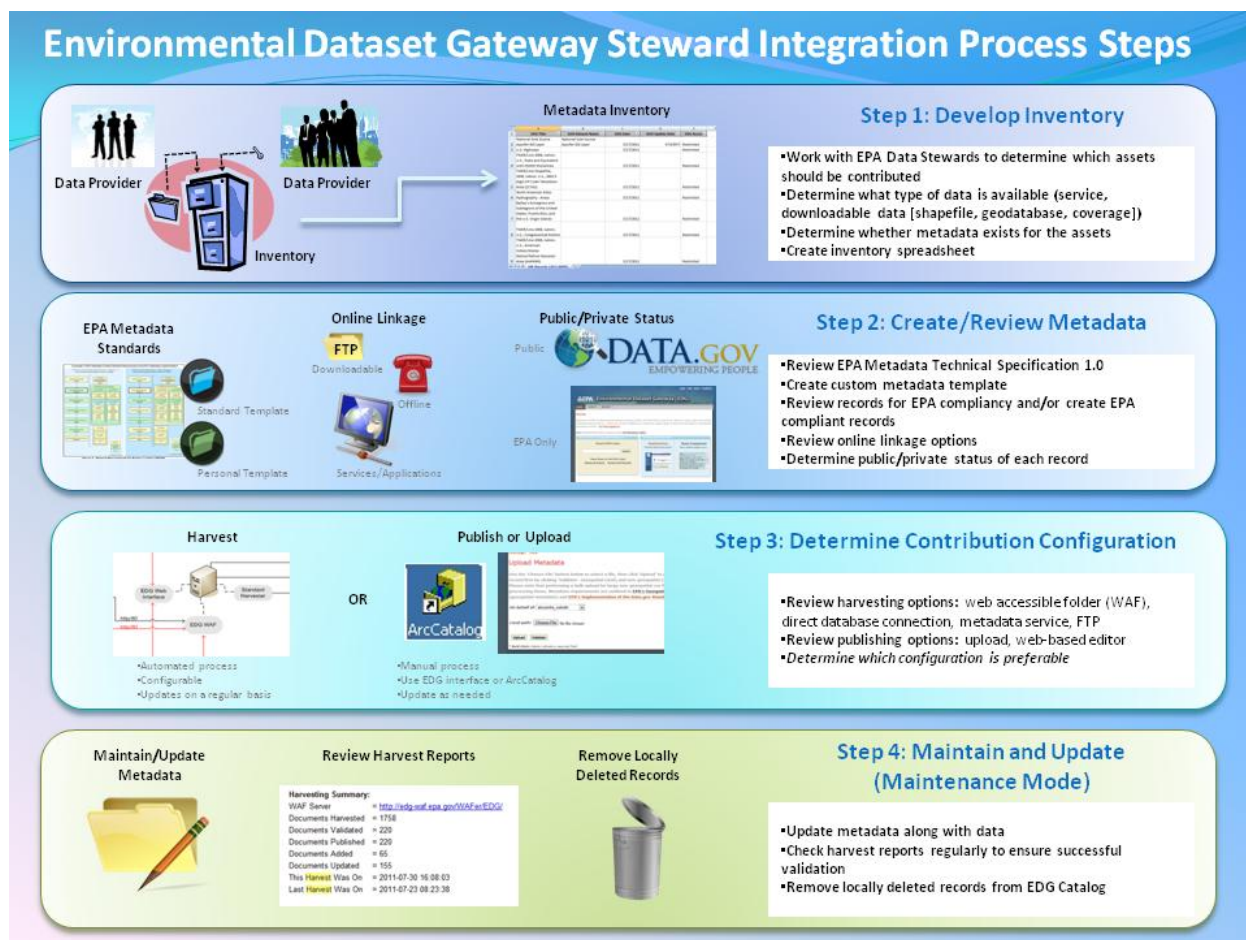

Figure 1. EDG Steward Integration Process Steps

Data stored and used by EPA programs may be derived from a number of sources. Some data is produced and maintained by EPA organizations, but a large majority of data is obtained from outside sources and used locally for performing regular business operations. Information obtained from external agencies may be supplied with metadata of varying quality, or it may be supplied without any metadata. As such, a number of key questions arise when addressing metadata sharing and compliancy, including:

- What is EPA-Compliancy?

- Which metadata are records must be EPA-compliant?

- Which metadata records should be contributed to the EDG?

- Which metadata records should be shared with the public?

Basic guidelines have been developed in order to help clarify these aspects of metadata prioritization and sharing. These are summarized in Table 1 below. These guidelines should be used by organizations as a basis for prioritizing metadata development, updates and sharing with internal and external parties.

Table 1. EDG Contribution and Distribution Criteria

EDG Contribution, Prioritization, and Distribution Criteria

Asset Type	EPA Compliancy Required?	Required to Contribute to EDG?	Make Available to the Public?
Assets (data, applications, and services) that EPA produces	Yes, EPA compliancy is required for all geospatial assets that originate at EPA.	Yes, required to contribute to the EDG	Yes, unless restricted due to data sensitivity, license agreement, or other EPA restriction on public access.
Assets acquired from other agencies but with value-added modifications made by EPA personnel	Yes, EPA compliancy is required. Personnel should ensure that all EPA-required fields are present in the metadata, but errors are allowable.	Yes, required to contribute to the EDG	Yes, unless restricted due to data sensitivity, license agreement, or other EPA restriction on public access.
Assets acquired from other agencies but are of critical importance or may be hard to find by other EPA personnel	EPA compliancy is encouraged, but not required. Personnel should ensure that all EPA-required fields are present in the metadata, but errors are allowable.	No, but highly encouraged to contribute to the EDG	No

7.2 EPA-Compliancy Requirements for Geospatial Metadata

EPA-compliant metadata is required for all resources that <u>originate</u> at EPA. If your organization develops geospatial resources for EPA, metadata should be developed for those resources that meets the *EPA Geospatial Metadata Technical Specification v1.0.* EPA-compliancy is *encouraged* for all geospatial resources <u>used</u> by EPA organizations. EPA-compliancy means that all EPA-required fields are included in your metadata record and that your record passes minimum FGDC compliancy. If you have questions about this determination, please contact the EDG Administrator or the EPA Geospatial Metadata Coordinator.

It is recognized that high value geospatial resources used at EPA may be derived from external groups but that the corresponding metadata records may contain some pre-existing FGDC compliancy errors. In those cases, it is *permissible* to allow records that have all minimum EPA requirements to be contributed to the EDG (but still retain FGDC errors). Personnel should review the record to ensure that it contains high-quality information (e.g., it has key information regarding the data set) and should use the EPA Metadata Editor (EME) to apply basic EPA fields if they do not exist. However, it is not required that external metadata records meet full EPA-compliancy requirements.

It should be recognized that EPA requirements have been developed so that metadata records are easily shared and consistent across the Agency. For example, the EPA-required field "Resource Description" should have consistent values so that the information is easily discoverable in the EDG and provides access to the resource as intended. As such, personnel are encouraged to apply EPA required fields to pre-existing metadata records even if the information is not fully EPA-compliant.

7.2.1 EPA Compliancy Requirements for Non-Geospatial Metadata

Non-geospatial metadata contributed to the EDG should comply with the Data.gov Standard. The EDG can be used to validate both geospatial and non-geospatial data to check for EPA compliancy.

7.3 Determining What to Contribute to the EDG

EPA requires that metadata for geospatial assets produced or 'improved' by EPA personnel and/or EPA contractor staff be contributed to the EDG. Metadata records that describe resources that would benefit other EPA organizations are encouraged but not required for contribution. The term "improved" refers to any type of value-add process that alters the data set or resource from its original form.

7.4 Differentiating Between Public and Private Metadata Contributions

EPA organizations have the ability to share metadata records with the general public by flagging them as 'unrestricted' in the EDG. These records are then accessible to the public at EPA's public metadata service, Web Accessible Folder (WAF), GOS, and Data.gov. Select records from GOS are flagged for inclusion in Data.gov. Some groups may be unsure which records should be shared with the public through these channels. In general, EPA should make metadata records that describe information that *reflects EPA's mission areas and business* available to the public. That includes all metadata records that describe geospatial resources that have been developed by or have had substantial value-add by EPA personnel. This may exclude those records that are considered sensitive, have not been published, are restricted by license agreement, or have draft status. Other EPA restrictions on public access may also be considered.

When making metadata records available to the public, it is highly encouraged to make the resource described in the record also available to the public. This will ensure that external parties that discover your metadata records can access the resources directly. It is recommended to review the online linkage field to verify that the URL entered is available to external parties. Additional information about this topic is provided in Section 7.5.

7.5 Providing Access to Resources (Online Linkage)

Information contributed to the EDG should be documented such that it may be readily accessed by others via the Internet. Providing this type of access makes the EDG more useful to those searching for information, and it also allows data providers to avoid handling multiple queries from EDG users hoping to access the information.

Within a metadata record, access is provided to data through the use of the online linkage element. A number of methods exist for providing access using online linkage, but the most commonly used methods include live data services, data download, and direct connection to a database. In cases where data owners do not wish to provide direct online access to their data, they may choose to contribute records to the EDG that are classified as 'offline data' and may provide contact information for users who wish to gain access to the data.

In general, where possible, it is always recommended to provide more than one method of access to data. Providing access through multiple venues ensures that the largest numbers of users will be able to reliably access the data once they have discovered the metadata. It also provides a source of back up for data owners during events where infrastructure bottlenecks make online distribution difficult.

The guidance provided below is intended to inform users about considerations for choosing between different types of data distribution options. This does not include details regarding how to document the online linkage element properly within a metadata record. The options identified below are listed in order of ease of access from the EDG (most easily accessible from the EDG to least easily accessible from the EDG).

7.5.1 Live Data and Services

Benefits (from the EDG perspective):

- Data provided in a service may be fed directly from a production data store and as such will be up-to-date.
- Data services are easily accessed using a variety of Web-based mapping or client tools, which allows a number of different types of users to access the information (e.g., users do not need to install special software).
- Data services are integrated directly within EDG map viewers, which typically provide a user with a better experience when accessing the information from the EDG.

Drawbacks

- Data services may be more difficult to administer than other simpler data access formats.
- Data services may be more unreliable than other data formats during periods of high network traffic.
- Data services typically require more expensive hardware, software, and system administration than other formats.

7.5.2 Downloadable Data

Benefits (from the EDG perspective):

- Downloadable data is typically the simplest and most reliable format for different users trying to access data.
- Downloadable data does not require as much software, hardware, or other infrastructure robustness as other formats.
- Downloadable data is relatively easy to administer.

Drawbacks

- Data made available for download typically requires additional maintenance because it is usually separated from the production data store and must be replicated to a Web-accessible folder.
- Downloadable data may be limiting for some users who do not have access to the software required to load or view it.
- Once users have downloaded data, the author may never be sure if users will return to obtain newer, more up to date versions. If the source data changes frequently, repeating the download process very often may be onerous for a user.

7.5.3 Direct Database Connection

Benefits (from the EDG perspective):

- Direct database connection may be made to the production data store and as such the data accessed by the user will be up-to-date.
- Direct connection to a database may require less maintenance as it does not require that owners maintain live services or generate multiple copies.

Drawbacks

- Read-only passwords cannot be provided through metadata (as per EPA security restrictions); as such direct database connection will typically cause a user to go through extra steps to access the information.
- Some production databases may not be accessible through Anytime Anyplace Access (AAA). As such, connection to a Spatial Database Engine (SDE) database may be limiting for some EPA contractors or offsite EPA employees.
- Providing a number of different users with direct access to a production database may not be desirable for data owners if usage is extremely high or if they are not willing to share the entire database.
- Opening direct database connections for SDE databases that are on public servers may be risky in terms of high-volume traffic and usage.

7.5.4 Offline Data

Benefits (from the EDG perspective):

- Protects data from use by unknown customers.
- Ensures that draft, sensitive, or restricted data are able to be discovered through the EDG but are not directly accessed by parties without communication with data owners.

Drawbacks

- May be onerous for data providers in responding to multiple inquiries for data
- Least useful access method for EDG users and most likely to be discarded when returned from a search for information.

7.6 Metadata Review and Validation

Metadata that is contributed to the EDG should be reviewed for consistency with the *EPA Geospatial Metadata Technical Specification v1.0 (geospatial records)* or *EPA's Implementation of the Data.gov Standard (non-geospatial).* EDG Stewards and EPA Geospatial Data Stewards may choose to validate records locally using the EPA Metadata Validation Service. This service is available as a part of the *EPA Metadata Editor Version 3.1.1.* Review and validation will also be performed by the EDG Administrator upon contribution of information to the EDG. The EDG Administrator will review non-compliant records with EDG Stewards to ensure that they can be updated and made compliant. Records that are not compliant with EPA Requirements (as per the *EPA Geospatial Metadata Technical Specification, v1.0*) will not be approved for contribution to the EDG. For more information on using the EPA Metadata Validation Service as part of the *EPA Metadata Editor Version 3.1.1,* please visit the EDG Web site (http://edg.epa.gov) and download the *EPA Metadata Editor Version 3.1.1* (http://edg.epa.gov/EME).

7.7 (Key EDG Contacts as of August, 2011)

Contacts listed below identify key EDG operational personnel. A list of key EDG Stewards will be made available at the EDG Web site http://edg.epa.gov.

- EDG Project Manager: Dave Parrish (OEI-OIC-IESD-ISSB)

  parrish.david@epa.gov; 214.665-8352

- EDG Administrator: Jessica Zichichi (Contractor, Innovate!, Inc)

  zichichi.jessica@epa.gov; 774-487-0445

8.   ROLES AND RESPONSIBILITIES

Roles and responsibilities outlined in this section reflect metadata-specific duties identified for contributing to and maintaining information posted to the EDG. The roles and responsibilities are derived from the EDP Procedure for Metadata Management, with the exception of the EDG Steward, which has been added to the framework in order to uniquely identify responsibilities of the EDG Steward. The specific responsibilities outlined in this section are a subset of the entire suite of geospatial metadata management responsibilities, as identified in the EDP Procedure. Responsibilities listed in this document only include those that relate specifically to operations involving the EDG. For a full set of geospatial metadata management responsibilities listed for each of these roles under the EDP Procedure and for direction on how to appoint these roles within an organization, please reference the EDP Procedure for Geospatial Metadata Management, located at http://www.epa.gov/geospatial/docs/2131.pdf.

8.1 EPA Metadata Coordinator

- **Primary Role:** Oversight for Inter-Agency metadata coordination and reporting.
- **Primary Responsibilities:**
  – Provides oversight for and coordination of metadata contributions to Geospatial One Stop (GOS), Data.gov, and Data.gov
  – Prepares and delivers metadata reports to OMB.
  – Provides support for cross-Agency metadata coordination efforts, including Data.gov activities.

8.2 EPA Data Owner

- **Primary Role:** Management of assets for an entire organization.
- **Primary Responsibilities:**
  - Provides oversight for their organization's data.
  - Ensures development of metadata for information created within their organization.
  - Recommends EDG Steward(s) for their organization as appropriate, and notifies EDG Administrator when there are personnel changes that impact EDG Stewardship.

8.3 EPA Data Steward

- **Primary Role:** Management of individual data sets, databases, and/or applications that comprise part or all of a programs data holdings.
- **Primary Responsibilities:**
  - Participates in metadata management for their organization, including creating metadata as appropriate for existing resources and planned acquisitions (marketplace records).
  - Oversees metadata upkeep along with data upkeep and processing.
  - Ensures geospatial metadata compliancy with the *EPA Geospatial Metadata Technical Specification v1.0* or the *EPA Implementation of the Data.gov Standard.*

8.4 Environmental Dataset Gateway Administrator

- **Primary Role:** Management of EDG application and metadata catalog.
- **Primary Responsibilities:**
  - Manages and approves information contributed to the EDG catalog, including validating records contributed to the EDG for EPA compliancy and approving or disapproving records based on validation results.
  - Oversees Web interfaces and related content provided to users at the EDG Web site.
  - Provides technical support for implementing and maintaining the EDG connection to GOS, Data.gov, and NSDI.
  - Provides training, outreach, and instruction on contributing to the EDG.
  - Contacts EDG Stewards on a yearly basis to ensure that the EDG maintains appropriate stewardship status and that any changes in personnel are properly addressed.

8.5   Environmental Dataset Gateway Steward

- **Primary Role:** Management and coordination of an organization's contributions to the EDG.
- **Primary Responsibilities:**
  - Works with Data Steward to organize metadata for contribution to the EDG, including creating metadata as appropriate.
  - Works with EDG Administrator to contribute metadata to the EDG catalog for their organization.

8.5.1   EDG Stewardship

The EDG Steward role is highlighted in this section because of its importance in the EDG governance framework. The EDG Steward acts as the central 'Authority' for assets contributed from each organization and will serve as the contact point for EDG implementers working through processes identified for EDG contribution.

Within EPA programs, the identification of the EDG Steward may vary depending on the nature of the

particular organization's structure. Typically, the Data Owner will recommend an EDG Steward for their organization. The EDG Steward identified for an organization may also fulfill one of the other roles listed in this document (e.g., EPA Data Steward or EPA Data Owner). The specific EDG responsibilities listed for the EDG Steward complement other responsibilities identified for the EPA Data Steward and EPA Data Owner roles. If an EDG Steward's employment is terminated, the Steward will be required to complete an exit interview with his/her supervisor in accordance with EPA standard termination processes.

In general, the EDG Steward will work with EPA Data Stewards in their organization to:

- Identify and prioritize local assets to contribute to the EDG in order to create an initial metadata inventory.
- Identify available metadata for assets identified.
- Create or oversee creation of metadata for assets as appropriate.
- Identify data set constraints and characteristics.
- Determine public/private nature of information.
- Determine data availability and format (online linkage).

Additionally, the EDG Steward works with the EDG Administrator to:

- Understand how to use the EDG.
- Understand non-geospatial metadata requirements according to the Data.gov Standard.
- Understand EPA geospatial metadata requirements according to the *EPA Geospatial Metadata Technical Specification Version 1.0*.
- Contribute metadata (upload/publish/harvest) to their organization's 'EDG Metadata Collection'.

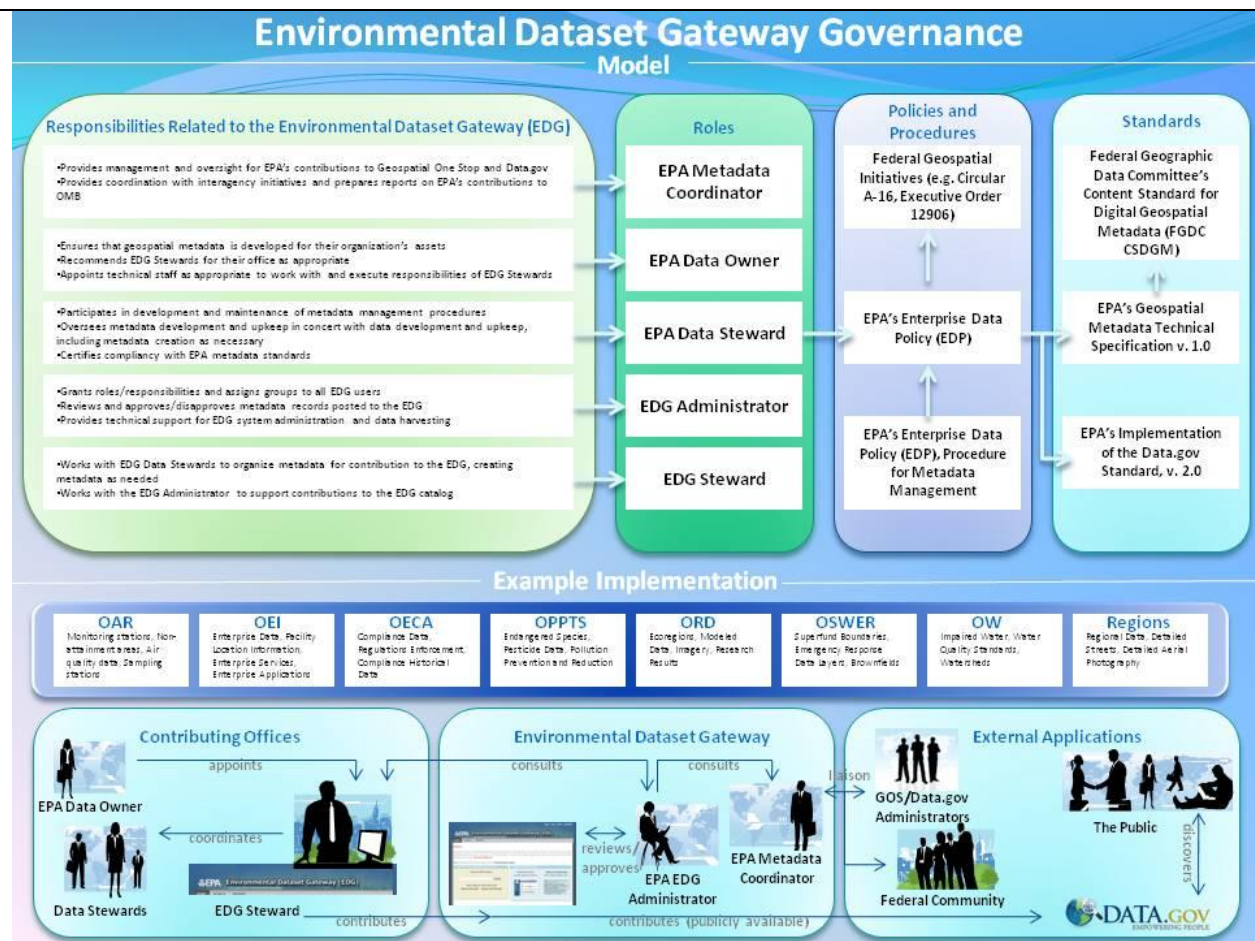   Maintain resources on a regular basis, as outlined in the EDG Maintenance Plan.

Figure 2. U.S. EPA Roles and Responsibilities Related to the EDG

9. ACRONYMS AND ABBREVIATIONS

Acronym	Definition / Clarification
AAA	Anytime Anyplace Access
EPA	U.S. Environmental Protection Agency
FGDC	Federal Geographic Data Committee
EDG	GeoData Gateway
GIO	Geospatial Information Officer
GOS	Geospatial One Stop
EDP	National Geospatial Data Policy
NSDI	National Spatial Data Infrastructure
OEI	Office of Environmental Information
OMB	Office of Management and Budget
QIC	Quality and Information Counsel
ROI	Return on Investment

| SDE | Spatial Database Engine |
| SIO | Senior Information Officer |

10. WAIVERS

    a.   **Waiver Process.** The Agency's CIO may grant waivers to selected provisions of this Procedure for sufficient cause. The CIO may re-delegate this authority.

    b.   **Applications.** Applications for waivers to specific provisions shall contain (1) identification of the Procedure provision; (2) a listing of reasons why the Procedure can not be applied or maintained; (3) an assessment of the impacts resulting from non-compliance; and (4) the signature of AA, RA, or Laboratory Director responsible for the spatial data collection in question.

Notification. The CIO will notify the requesting office in writing of the disposition of the waiver within 60 days of receipt.

11. RELATED PROCEDURES AND GUIDELINES

Required implementation guidelines for this Procedure are presented in the Enterprise Data Policy implementation Guidelines.

12. ADDITIONAL INFORMATION

For further information about this Procedure, please contact the EPA Office of Environmental Information, Office of Information Collection.