



GeoData Gateway

Connecting EPA's Geospatial Resources

Procedures for Storing Data at the GeoData Gateway (GDG) Download Locations

Some GDG contributors would like to provide online access to their data but may not have the capability or do not want to provide direct access to the information on their servers. For those users, the GDG provides 2 storage locations that may be used to store and provide online access to data. The locations are referred to as the '**GDG Download Locations**'. Any EPA office that publishes metadata to the GDG can use the GDG Download Locations to store their data as long as the total size of the content stored does not exceed 4 GB.

One of the GDG Download Locations is an **internal only URL** that is available only to those users who are able to access the EPA's intranet; the other is an **internal/external location** that is made available to both internal EPA and external public users. Procedures for using and maintaining information posted to these locations are outlined below.

The base URLs for the download locations are:

- ◆ **Internal:** <http://geogateway-intranet.epa.gov/data>
- ◆ **External:** <http://geogateway.epa.gov/data>

Offices posting information to GDG will be provided with office-specific folders to which they can post information. The naming convention used for the download locations is as follows: [http://\[base_URL\]/data/\[office_name\]](http://[base_URL]/data/[office_name]), with [office_name] replaced by R1, R2, OAR, OEI, etc. (For example, OEI has a folder at the internal storage location called <http://geogateway-intranet.epa.gov/data/OEI/>). Directories are set up for each office as needed.

The step-by-step process is as follows:

1. Determine if you would like to store data at the **Internal** or **External** GDG Download Location (or both). This is generally determined by your data sensitivity or access restrictions and should align with your metadata access policy (restricted/unrestricted).
2. Email Jessica Zichichi jzichichi@innovateteam.com and/or Michelle Torreano torreano.michelle@epa.gov and identify which storage location is desired. Jessica and Michelle will coordinate with you to assist with posting files and providing your online linkage information for your metadata records.
3. Post your data to the GDG Storage Locations. You can choose to either **email** your data directly to Jessica and Michelle or **post** your data to the GDG server using ftp. To ftp data to the server, take the following steps:
 - Verify with Jessica and/or Michelle that your AA account has been assigned to the internal GDG Server
 - Open your ftp client tool.
 - Type 'geogateway-intranet.epa.gov' for the Host or URL
 - Type in your user name and password
 - Connect to the site
 - Navigate to your folder
 - Upload information to your folder

4. If you are planning to make your data publicly available, the data will be moved to the GDG External Storage Location.
5. After data are loaded, it is important to make sure to update your metadata record to point to the download location and to make sure that the record type is specified appropriately.
 - a. Pointing the metadata record to the data storage location is done using the online linkage (<onlink>) element. The online linkage element is located in Section 1 of the metadata record and may be repeated. The first link in the record is known generally as the primary linkage. The primary linkage should be filled out so that it points to your data storage location, with the full folder and file name included:
 - **<onlink>http://geogateway-intranet.epa.gov/data/[office]/[file]</onlink>**
 - If you are using the EPA Metadata Editor (EME) in ArcCatalog, then you would document online linkage in Tab 1 (See Figure 1 below).
 - b. The record type is specified using the resource description (<resdesc>) element (located in Section 6 of the metadata record). Users should specify that the metadata record describes **Downloadable Data** in this element. If you are using the EPA Metadata Editor (EME), then you would document the resource description in Tab 3 (See Figure 2 below).

The screenshot displays the EPA Metadata Editor (EME) window, specifically the 'Basic Data Set Information' tab. The interface is divided into several sections for data entry:

- Citation:** Fields for Origin (USEPA Office of Water), Title (National Sole Source Aquifer GIS Layer), Publisher (U.S. EPA Office of Water), and Published at (Washington, D.C.).
- Online Linkage:** A section with a red arrow pointing to the 'Primary Linkage' field, which contains the URL 'http://geogateway-intranet.epa.gov/data/ow/S/'. A secondary linkage is also present.
- Description:** Fields for Abstract and Purpose, describing the data as a vector polygon GIS layer showing available materials.
- Time Period:** Fields for Date of Data Set (20091216), Progress of data (In work), Data currency (Publication date), and Update frequency (As needed).
- Bounding Box:** Fields for North, East, South, and West coordinates.
- Keywords:** A list of keywords including 'Inland Waters', 'imageryBaseMapsEarthCover', etc.
- Data Set Constraints:** Fields for Access, User, and Security Classification.
- Contact:** Fields for Primary Person or Organization (USEPA Office of Water | A. R. Anzolin).

A legend at the bottom indicates field requirements: YELLOW * mandatory, GREEN ** mandatory if applicable, and BLUE optional. A note at the bottom right says 'Click on text to link to element description'.

Figure 1. Documenting Online Linkage in the EME

Figure 2. Documenting 'Downloadable Data' using the EME

6. Once the metadata has been updated, you are ready to contribute it to the GDG. Use your standard process for contributing your new metadata to the GDG (harvest or upload).
7. Once your updated information has been harvested or published to the GDG, it is recommended that you perform a search for your data at the GDG and make sure that it shows up correctly, both with the appropriate content type identifier (Downloadable Data) and correct linkage.

General Recommendations for Posting Data to the GDG Download Locations:

- Information stored in the GDG Download Locations should be maintained along with the regular maintenance cycle used for managing your data locally. At a minimum please revisit the information posted on these sites semi-annually to ensure that the data are current.
- It is encouraged that, if you use shape file format for storing data, you zip your information into a single file for each layer.
- Information stored at either of the **GDG Download Locations** may not exceed 4GB within each folder. *Exceptions may be made if needed.*

If you have any questions, please contact Michelle Torreano at torreano.michelle@epa.gov