



Procedures for Storing Data at the EDG Data Download Locations

The EPA Environmental Dataset Gateway (EDG) provides Data Download Locations for EPA Regions, Program Offices, and Research Laboratories that want to make data available online for download but do not have their own server available to do so. Any EPA office that publishes metadata to the EDG may use the Data Download Locations to store and disseminate geospatial and/or non-geospatial data free of charge.

Three EDG Data Download Locations are available:

<https://edg.epa.gov/data/Public/> - Public-access location for files under 250 MB, available to all EDG users.

<https://edg.epa.gov/data/Restricted/> - Restricted-access location for files under 250 MB, available only to users with an EDG login. For larger restricted access files contact edg@epa.gov.

<ftp://newftp.epa.gov/EPADataCommons/> - High-performance file transfer server for very large public files.

Within each location, office-specific folders are created by the EDG Administrative Group. Data Stewards may choose to structure subdirectories within the folders if desired. Navigate to the URLs listed above to review the structure that is being used by offices currently utilizing EDG Data Download Locations.

Instructions for adding data:

1. Determine Data Accessibility Requirements

Determine whether the Public or Restricted EDG Download Location is appropriate for the data. This is generally driven by data sensitivity or access restrictions and should align with the Security Classification listed in your metadata record (public, restricted public, non-public). Files larger than 250MB that are intended for the public should be placed on the new high-performance file transfer server. For restricted access files that are larger than 250 MB, or for public files that exceed 20GB in total, please contact edg@epa.gov to discuss options.

2. Contact the EDG Administrative Group

Email the EDG Administrative Group (edg@epa.gov) and identify which storage location is desired.

3. Place Data in Data Download Location

You may add your data to the data download location by either providing your zipped dataset to the EDG Administrative Group by emailing edg@epa.gov or by uploading it to the EDG Data Download Locations using a FTP client tool (such as FileZilla).

To FTP data to the server, perform the following steps:

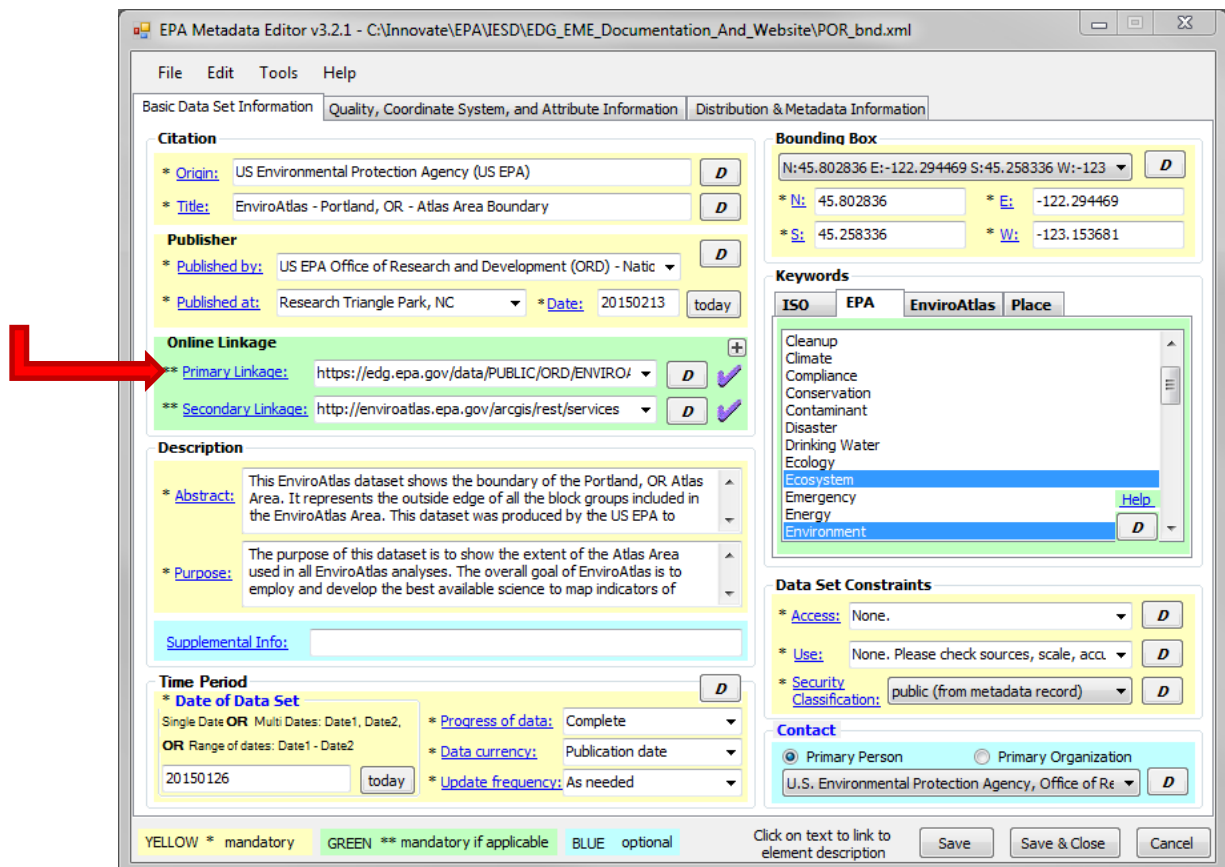
- 3a. Verify with the EDG Administrative Group that your AA account has been assigned to the internal EDG server.
- 3b. Open your FTP client tool.
- 3c. Use "edg-intranet.epa.gov" for the Host or URL, or "newftp.epa.gov" for the high-performance transfer server. For newftp.epa.gov, the protocol should be set to SFTP – insecure uploads are not permitted.
- 3d. Type in your user name and password (user name must be entered as "aa\{username}" for edg-intranet, and "username" for newftp.epa.gov).
- 3e. Connect to the site and navigate to your folder.
- 3f. Upload the zipped dataset to your folder.

4. Update the Metadata Record to Provide the Data Download URL

After the data are loaded to the Data Download Location for your office, update the metadata record to indicate that the dataset is available as downloadable data.

FOR A GEOSPATIAL RESOURCE:

- 4a. Update the primary online linkage (<onlink>) element to point to the download location. The primary online linkage element is the first link located in Section 1 of the FGDC metadata record. Use the full folder and file name of your storage location (Ex: [https://edg.epa.gov/data/\[Public or Restricted\]/\[office\]/\[file\].zip](https://edg.epa.gov/data/[Public or Restricted]/[office]/[file].zip)). If you are using the EPA Metadata Editor (EME) Version 3.2.1 or earlier, the Primary Linkage is in the first tab (Basic Data Set Information) under Online Linkage (see Figure 1 below).



The screenshot shows the EPA Metadata Editor v3.2.1 interface. The 'Basic Data Set Information' tab is active. The 'Online Linkage' section is highlighted in green, and a red arrow points to the 'Primary Linkage' field. The 'Primary Linkage' field contains the URL <https://edg.epa.gov/data/PUBLIC/ORD/ENVIRO/>. The 'Secondary Linkage' field contains <http://enviroatlas.epa.gov/arcgis/rest/services>. Other sections include 'Citation', 'Bounding Box', 'Keywords', 'Description', 'Time Period', 'Data Set Constraints', and 'Contact'. A legend at the bottom indicates that yellow fields are mandatory, green fields are mandatory if applicable, and blue fields are optional.

Figure 1. Documenting Online Linkage in EME 3.2.1

- 4b. Update the record type (<recdesc>) element to indicate that the metadata record describes downloadable data. This is located in Section 6 of the FGDC metadata record. Specify that the metadata record describes Downloadable Data in this element. If you are using EME Version 3.2.1 or earlier, the Data Resource Type is in the third tab (Distribution and Metadata Information) under Distribution Information (see Figure 2 below).

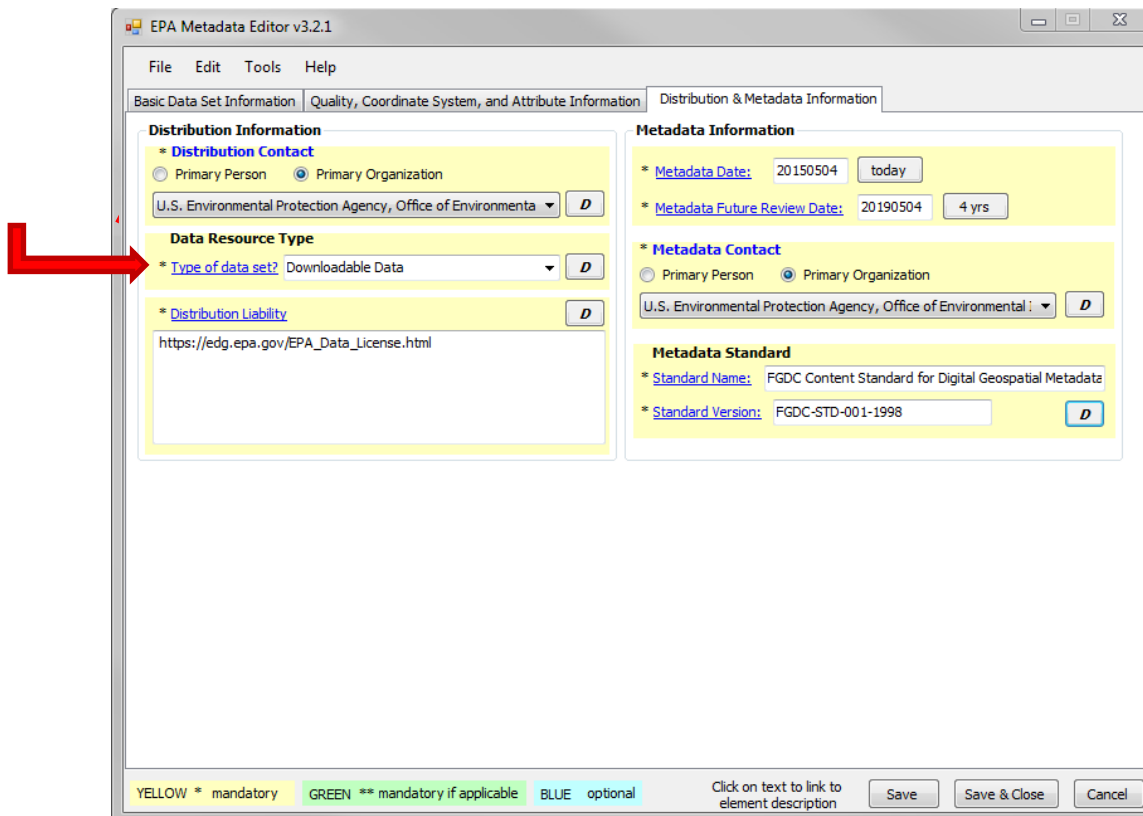


Figure 2. Documenting “Downloadable Data” in EME 3.2.1.

FOR A NON-GEOSPATIAL RESOURCE:

Update the Linkage element to point to the download location. The Linkage element is found in the Distribution Info section of the ISO metadata record. Use the full folder and file name of your storage location (Ex: [https://edg.epa.gov/data/\[Public or Restricted\]/\[office\]/\[file\].zip](https://edg.epa.gov/data/[Public or Restricted]/[office]/[file].zip)). If you are using EME Version 4.0, the Linkage is in the second tab (Distribution & Metadata Information) under Distribution → Digital Transfer Options. Also update the Online Function Code to indicate that the data are available as a download.

4.1. Create download linkage file (This step is only for files uploaded to the high-performance EPADDataCommons site)

Files uploaded to the high-performance transfer server have a FTP URL that should be easy to determine based on your upload (Ex: [ftp://newftp.epa.gov/EPADDataCommons/\[office\]/\[file\]](ftp://newftp.epa.gov/EPADDataCommons/[office]/[file])). We are working to also enable HTTP access to these files, but at present, FTP is the only option.

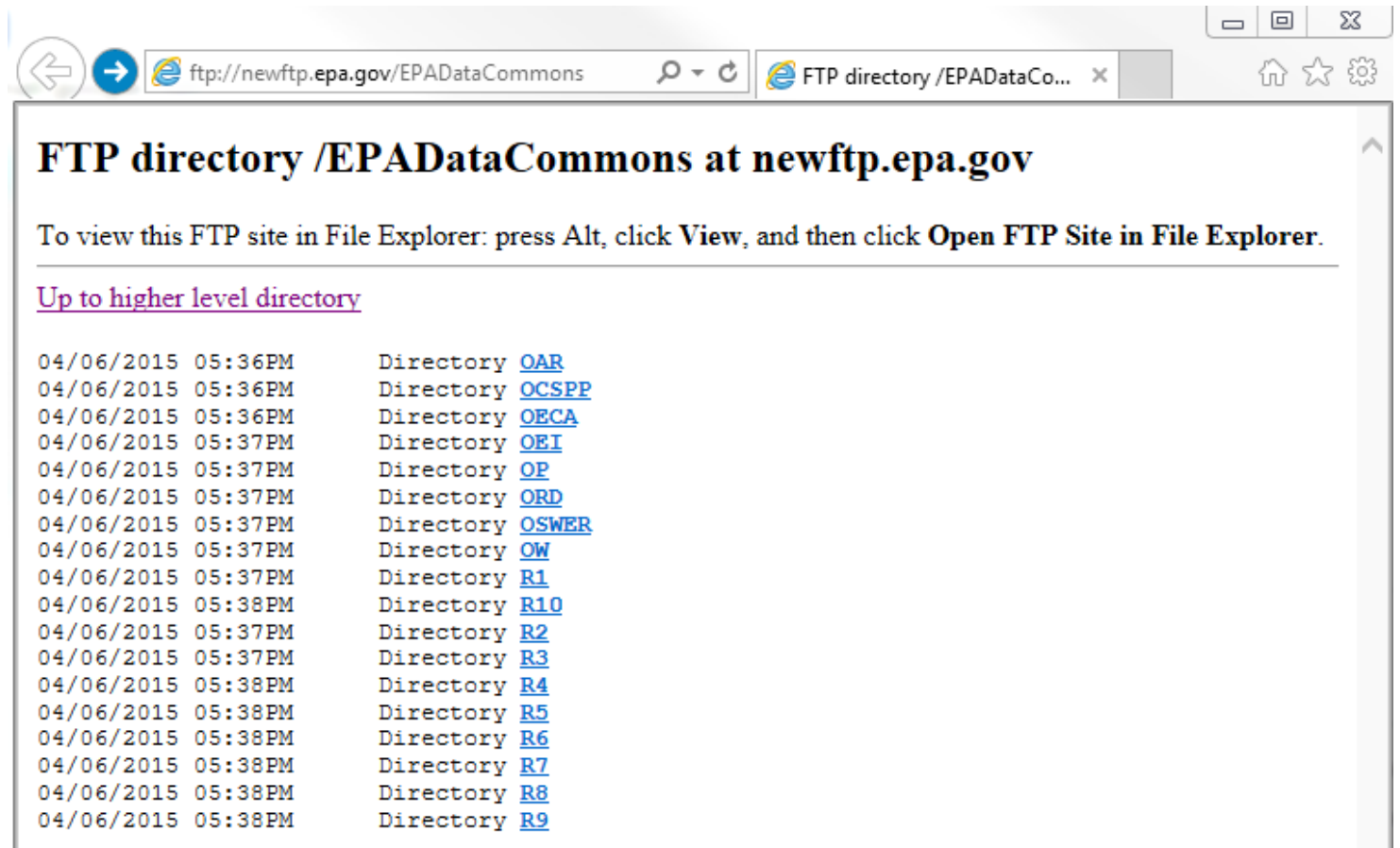


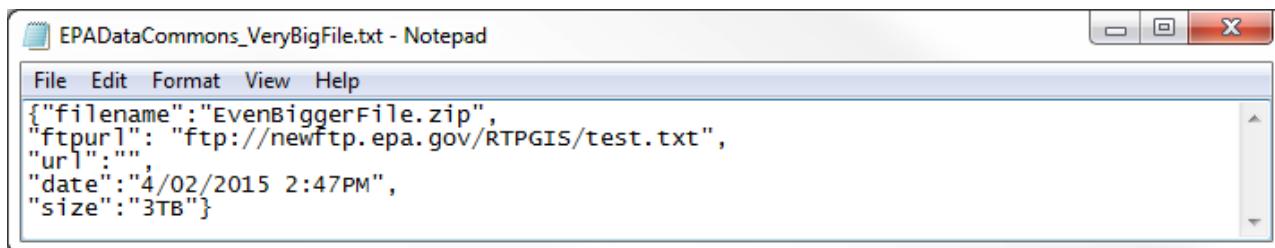
Figure 3 - EPADDataCommons FTP site for anonymous public in Internet Explorer 11

- 4.1a. Follow steps 4a and 4b above to add this URL ([ftp://newftp.epa.gov/EPADDataCommons/\[office\]/\[file\]](ftp://newftp.epa.gov/EPADDataCommons/[office]/[file])) with the FTP link to your metadata record.
- 4.1b. So that a link to the uploaded EPADDataCommons file will also appear in the <https://edg.epa.gov/data> table of contents, place a small text file in your folder on that system with the following text:


```

{"filename": "VeryBigFile.zip", "ftpurl":
"ftp://newftp.epa.gov/epadatacommons/[office]/[VeryBigFile.zip] ", "url": "",
"date": "1/13/2015", "size": "3GB"}

```



The name of this placeholder text file should begin with the prefix EPADDataCommons_ and have a file extension of either .txt or .json. An intuitive example might be “EPADDataCommons_VeryBigFile.txt”. The empty “url” parameter is a placeholder for when HTTP access to these files is enabled – we will update this document and notify all data stewards at that time. See Figure 4 below for an example result of creating placeholder file.



Figure 4. Example of links to FTP download files from EDG Data Download folder (note underlined FTP links).

5. Publish the Metadata Record in the EDG

Once the metadata record has been updated, it is ready to be published in the EDG. Use your standard process for publishing your new metadata to the EDG (manual upload or WAF synchronization).

6. Review the Published EDG Metadata Record

After updated metadata has been published to the EDG, it is recommended that you search the EDG for your metadata record to verify that it displays correctly, both with the appropriate content type identifier (Downloadable Data) and the correct online linkage.

General Recommendations for Uploading Data to the EDG Data Download Locations

- Data stored in the EDG Data Download Locations should be maintained along with the regular maintenance cycle used for managing your data locally. At a minimum, please revisit the information uploaded on these sites semi-annually to ensure that the data are current.
- When shapefiles are being uploaded, it is recommended that you zip all datasets into a single file.
- Data stored the EDG Data Download Locations (Public or Restricted) may not exceed 4GB within each folder, or 20GB per folder on the high-performance transfer server. *Exceptions may be made if needed. Contact the EDG Administrative Group at edg@epa.gov for additional assistance.*