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 within reasonable limits described below.

Two 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://gaftp.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.

Data Stewards may add data to the Data Download Locations for their Region, Program Office, or Research Laboratory free of charge by following these detailed instructions on how to prepare and submit datasets:

1. Contact the EDG Administrative Group
2. Upload Content to the EDG Download Location
3. Update the metadata record to provide the data download URL
4. Publish the metadata record in the EDG
5. Review the published EDG metadata record

1. Contact the EDG Administrative Group (edg@epa.gov)

Email the EDG Administrative Group and identify the following:

- What is the total size of the files you plan to host?
- How long do you need the data hosted?

The EDG Administrative Group will coordinate with you to assist with uploading the files and providing your online linkage information for your metadata records.

2. Upload Content to the EDG Download Locations

You may provide your zipped dataset to the EDG Administrative Group by either uploading it to the EDG Data Download Locations using, the share drive (\\edg-intranet.epa.gov\ftproot\), EPA's Go Anywhere application, FTP, or by emailing it to edg@epa.gov.

 To Connect to the share drive on the EDG intranet server, perform the following steps:

- Open Windows Explorer
- Type the UNC path, \\edg-intranet.epa.gov\ftproot into Windows Explorer's path.
- Once connected, navigate to your group's folder.

To Connect to EPA's Go Anywhere application (high-performance transfer server) using a web browser:

- Go to https://newftp.epa.gov in a web browser
- Login using your LAN ID and password
- Click on the "Navigate" button. Input "EPADataCommons" in the text box that opens.
- Navigate to your folder and upload your data.

To FTP data to the high-performance transfer server, perform the following steps:
- Open your FTP client tool.
- Use "newftp.epa.gov" for the Host or URL. The protocol should be set to SFTP – insecure uploads are not permitted.
- Type in your user name and password
- Connect to the site and navigate to your folder.
- 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:

Update the primary online linkage (<onlink>) element to point to the download location. The primary online linkage element is found in the Distribution section under Digital Transfer Options of EME 5. Use the full folder and file name of your storage location:
- EDG Intranet server (files under 250MB in size):
  - Ex: https://edg.epa.gov/data/Public/[office]/[file].zip
- HTTPS URL for High-Performance file transfer server:
  - Ex: https://gaftp.epa.gov/EPADataCommons/[office]/[file].zip
- FTPS URL for High-Performance file transfer server:
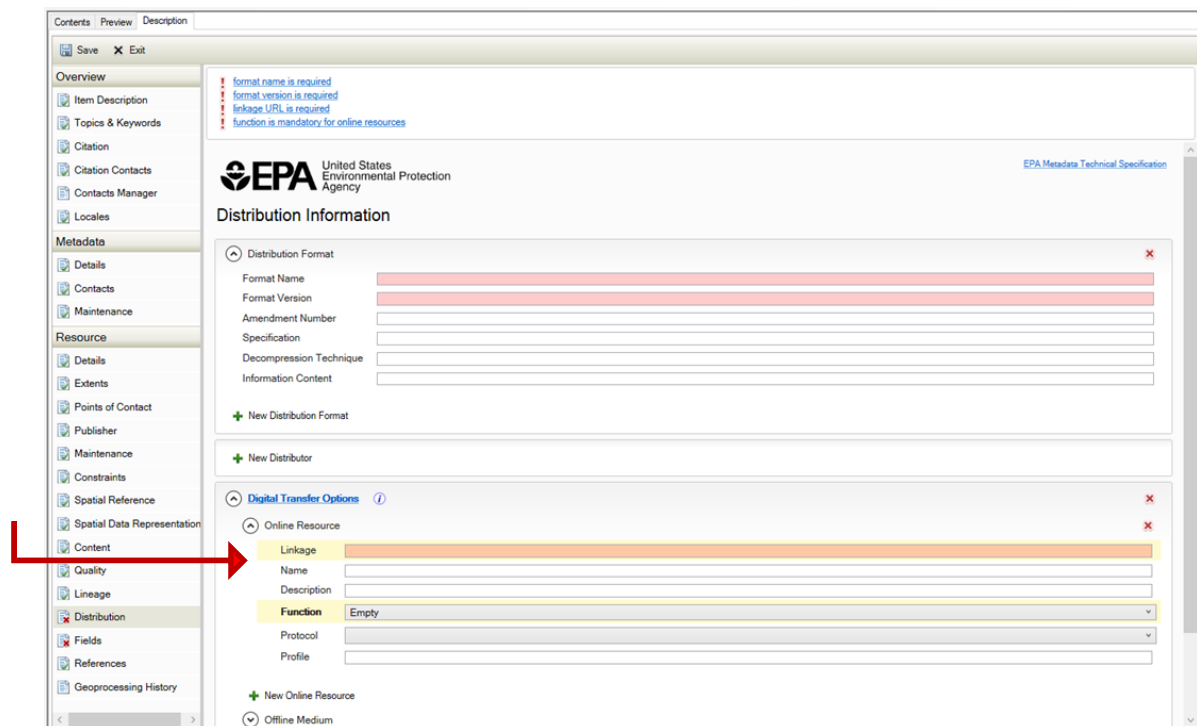 - Ex: ftps://newftp.epa.gov/EPADataCommons/[office]/[file].zip

Figure 1. Documenting Online Linkage in EME 5

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/[office]/[file].zip).

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 (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 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.*