

US-EPA

# EnviroAtlas Requirements for Data Submission

Data and metadata guidelines for submission

# Contents

<b>Introduction</b> .....	2
<b>Guidelines for Submitting Data for Inclusion in EnviroAtlas</b> .....	2
Process data into standardized EnviroAtlas format (Required): .....	3
Provide FGDC-compliant metadata using the EPA Metadata Editor (Required): .....	5
Provide EnviroAtlas Fact Sheets for each metric or group of related metrics (Required): .....	6
Inform EnviroAtlas that you have delivered data (Required): .....	6
Provide detailed documentation describing the methodology used to create your data so data can be regenerated in the future and include guidelines for data display (Highly Recommended): .....	6
Provide guidelines for data/metric display in the EnviroAtlas map viewer (Highly Recommended): .....	6
Provide ArcToolbox models and/or scripts to regenerate metrics in the future (EPA staff and EPA Contractors - Required*. Partners/Affiliates - Highly Recommended): .....	7
Archive all data associated with creation of metrics (EPA staff and EPA Contractors - Required. Partners/Affiliates - Highly Recommended): .....	7
Provide widgets for the EnviroAtlas interactive map viewer (Optional): .....	8
<b>Appendices: Creating EnviroAtlas Metadata with the EPA Metadata Editor</b> .....	9
Appendix A. Getting Started with the EME and the EnviroAtlas Custom EME Access Database and Metadata Template .....	10
Appendix B: Using the EME to Complete EnviroAtlas Metadata Records .....	11
Completing Fields in the EME Tab 1 – Basic Dataset Information .....	12
Completing Fields in the EME Tab 2 – Quality, Coordinate System, and Attribute Information .....	16
Completing Fields in the EME Tab 3 – Distribution and Metadata Information .....	20
Final Metadata Steps – Spell Check and Validation (Required) .....	21

## Introduction

This document was created to standardize data submission by EPA staff, contractors and cooperators to EnviroAtlas. Numerous steps must be completed before data can be published in EnviroAtlas. Once data are received following the procedures outlined below, the EnviroAtlas Data Submission Group will process data for publishing in EnviroAtlas. Communication regarding status and completeness of data submitted will be continued throughout the process of data publishing.

EnviroAtlas publishes data as mapping services (Esri and Open Geospatial Consortium formats) through the US EPA's GeoPlatform, and as such, metadata and documentation of data are required for data publishing. Associated metadata must be validated when data are submitted to EnviroAtlas. This step is important because many fields populated in the metadata will be used in EnviroAtlas map service descriptions and published to EPA's [Environmental Data Gateway \(EDG\)](#). The metadata must also satisfy the requirements of Data.gov and the EPA GeoPlatform. After validation, the next review tests for data completeness – verifying that attributes joined to original spatial data contained within EnviroAtlas produce appropriate results. The EnviroAtlas Data Submission Group will work with the data submitter to verify data and metadata. Once data and metadata are verified and accepted, maps will be published in a development environment for testing and placement in EnviroAtlas. At this point, the data submitter may be able to review mapped results depending on access to the EPA network. Input from data submitters will be sought, but final decisions about map symbology and the EnviroAtlas interactive map Table of Contents structure as well as map service naming conventions will be completed by the EnviroAtlas Data Submission Group. Once development map services have been finalized and approved, they will be deployed in the EPA GeoPlatform hosting environment for public access. As soon as services are deployed in the EPA GeoPlatform, they also become viewable through the EnviroAtlas interactive map. Metadata will be submitted to the EDG and will be searchable based on content in the metadata record.

## Guidelines for Submitting Data for Inclusion in EnviroAtlas

In an effort to complete timely updates to EnviroAtlas and to follow the guidelines required for publishing data to the EPA GeoPlatform, we request the following steps be used for submitting processed data:

- Summarize data to one of four standardized boundary datasets: Watersheds, Counties, Census Tracts, or Census Block Groups. In some instances, data may be submitted in a different boundary format than those listed above. These will be accepted on a case-by-case basis and will follow, when relevant, the remaining steps in the submittal process.
- Output calculations/data values into a single comma-separated values (CSV) file or an ArcGIS (version 10.x) file geodatabase (FGDB) with a single table (do not include polygon boundaries) with polygon identifier and one column for each metric (i.e., field, variable, value). Each CSV or FGDB table may contain multiple metrics.
- Complete Federal Geographic Data Committee (FGDC) metadata document for the CSV file or FGDB table – including a description for each metric field.
- Complete Fact Sheet document for each metric or group of related metrics.
- Complete additional documentation, scripts, or models to submit with the data to detail the process used to recreate and/or classify the data.

Optional steps:

- Create and provide scripts or Esri ArcGIS Toolbox models to recreate or update submitted data.
- Create additional tool or widget for the EnviroAtlas interactive map viewer.

To submit your data please contact: [EnviroAtlas Data Submission Group](mailto:EnviroAtlas@epa.gov) (EnviroAtlas@epa.gov)

After receiving your request, instructions for delivering data, metadata, and other ancillary documentation to EnviroAtlas will be provided. Scripts, ArcGIS Toolbox models, or widgets for the EnviroAtlas map viewer can also be included. If the data and metadata package can be compressed into a file of less than 5mb (or whatever data limit your email system has, if under 5mb), you can attach a compressed "zip" file with the file extension changed from ".zip" to ".zipfile".

Specific instructions for submitting data to EnviroAtlas continue below.

### Process data into standardized EnviroAtlas format (Required):

1. Most EnviroAtlas data will be summarized by one of the following spatial boundaries using datasets provided by the EnviroAtlas Data Submission Group. All datasets include FGDC-compliant metadata and are provided as Esri file geodatabases in Albers projection. Boundaries change frequently and it is absolutely critical that the data are calculated based on these standardized boundaries. **Metric data must be calculated using only the provided boundaries; other boundary datasets may produce invalid data. Standardized boundaries will be available for download soon.**
  - a. **Watersheds:** The National Hydrography Dataset (NHD) Plus V2 snapshot of the Watershed Boundary Dataset (WBD) will be used to define watersheds (12-digit Hydrologic Unit Codes (HUCs), n=97,428 polygons). The EnviroAtlas uses a version of the NHDPlus V2 WBD snapshot downloaded January 2015. Minor modifications to the international border HUC 12 features have been made and are documented in the associated metadata. Also, percent of land area based on the National Land Cover Dataset (NLCD) from 2011 is included as an attribute. The attribute "HUC\_12" is a character string (meaning that it should be contained in double quotes if possible) and will be the polygon identifier included in the submitted CSV file or FGDB table. Download the watersheds using [HTTPS](https://www.epa.gov/nhdplus) or [FTP](ftp://www.epa.gov/nhdplus).
  - b. **Counties:** The 2010 TIGER County dataset will be used to define the county boundaries (n= 3231 polygons). The attribute "STCO\_FIPS" is a character string (should be contained in double quotes); it will be the polygon identifier included in the submitted CSV file or FGDB table. To access data, please request from [EnviroAtlas Data Submission Group](mailto:EnviroAtlas@epa.gov).
  - c. **Census Tracts:** The 2010 TIGER Census Tract dataset will be used to define the census tract boundaries (n= 74,102 polygons). The attribute "STCOTR\_FIPS" is a character string (should be contained in double quotes); it will be the polygon identifier included in the submitted CSV file or FGDB table. To access data, please request from [EnviroAtlas Data Submission Group](mailto:EnviroAtlas@epa.gov).
  - d. **Census Block Groups:** The 2010 TIGER Census Block Group dataset will be used to define the census block group boundaries (n= 220,653 polygons). The attribute "STCOTRBG\_FIPS" is a character string (should be contained in double quotes); it will be the polygon identifier included

in the submitted CSV file or FGDB table. To access data, please request from [EnviroAtlas Data Submission Group](#).

2. For EnviroAtlas data not summarized by one of the standard reporting units, please adhere to these guidelines:
  - a. All raster/grid data must align with the [Multi-Resolution Land Characteristics Consortium \(MRLC\) 2011 National Land Cover Database \(NLCD\) 30m grid](#), Albers projection, NAD83, and must be provided in ArcGIS GRID format or as a TIFF file with TIFF world file.
  - b. For other non-standard data: Please contact the [EnviroAtlas Data Submission Group](#) for details on *data formats or spatial boundaries accepted for inclusion in EnviroAtlas*. Provide as much information as possible about the content, format, extent, and scale of data in your email.
3. All area calculations must be performed in the following Albers projection. This projection is defined in ESRI products as "USA\_Contiguous\_Albers\_Equal\_Area\_Conic\_USGS\_version" and is assigned an ESRI SRID value of 102008. The datum should be set to North American Datum 1983 (NAD1983). The Albers definition for the Continental United States is:

Projection: Albers  
False\_Easting: 0.0  
False\_Northing: 0.0  
Central\_Meridian: -96.0  
Standard\_Parallel\_1: 29.5  
Standard\_Parallel\_2: 45.5  
Latitude\_Of\_Origin: 23.0  
Linear Unit: Meter (1.0)

4. Data must be submitted as a single table in CSV format or an ArcGIS file geodatabase (FGDB) table. The CSV export format is available through ArcMap and ArcCatalog. Remove all fields except for the following:
  - a. Unique identifier field from the base geography used in processing. This will be either "HUC\_12" for the Watershed, "STCO\_FIPS" for the County, "STCOTR\_FIPS" for the Census Tract, or "STCOTRBG\_FIPS" for Census Block Group data.
  - b. One field or column for each metric you have calculated. The table may contain multiple metrics, but each metric must be represented by only one field. Field name length should be limited to no more than 20 characters and cannot include spaces or special characters.
  - c. For non-standard formats (e.g., unique data layers, reference data layers, special study areas), data will be submitted in a format agreed upon by the data submitter and the EnviroAtlas Data Submission Group.

HUC_12	RLA_30	RO_30	RNAT30	RFOR30	RFOR9030
010100010101	5259600	100	100	82.7002	82.7002
010100010102	5525100	100	100	70.3698	70.3698
010100010103	7668000	100	100	95.2465	95.2465
010100010104	20288700	100	100	90.7022	90.7022
010100010105	9958500	100	99.8102	90.4745	90.4745
010100010106	4599900	100	100	85.541	85.541
010100010107	4220100	100	98.7204	86.479	86.479
010100010108	5229000	100	99.8967	87.5043	87.5043
010100010109	13536900	100	99.9601	86.6698	86.6698
010100010201	8518500	100	99.8521	74.5906	74.5906
010100010202	7407000	100	99.9028	85.0668	85.0668
010100010203	2542500	68.63	99.469	80.4956	80.4956
010100010204	2899800	66.5	100	83.923	83.923
010100010205	1628100	53.25	99.5025	91.487	91.487
010100010206	4405500	100	99.8366	81.287	81.287
010100010207	5323500	100	99.9324	90.3804	90.3804
010100010208	12233700	97.13	99.7278	80.6665	80.6665
010100010301	0	0	0	0	0
010100010302	0	0	0	0	0

Example CSV file format

### Provide FGDC-compliant metadata using the EPA Metadata Editor (Required):

1. A single metadata file (in XML format) must be submitted for each dataset and each dataset submitted should contain data processed using the same general procedures. EnviroAtlas, the EPA GeoPlatform, and Data.gov cannot accommodate multiple metadata files for a single dataset. EnviroAtlas requires using the [EPA Metadata Editor \(EME\)](#); the current EME version for geospatial data is 3.2.1) to create metadata associated with EnviroAtlas dataset contributions. To assist EnviroAtlas data contributors with metadata creation, an Access database specific to EnviroAtlas and an EnviroAtlas metadata template, which includes several pre-populated fields, are available for use. An example metadata record is also available for download. Implementing the EME with the customized EnviroAtlas EME Access database and EnviroAtlas metadata template will significantly streamline the metadata creation process and will help ensure that EnviroAtlas metadata is consistent across multiple providers/analysts. Please contact the [EnviroAtlas Data Submission Group](#) with questions about this requirement.
2. Install and use [EME v3.2.1](#). The EME software can be populated through Esri's ArcCatalog or as a standalone application. The EME creates and saves metadata in XML format. **Detailed instructions for accessing and installing the EME 3.2.1 software and creating metadata with the EME are included at the end of this document ([Appendices: Creating EnviroAtlas Metadata with the EPA Metadata Editor](#)).**

## Provide EnviroAtlas Fact Sheets for each metric or group of related metrics (Required):

1. Begin with the [downloadable EnviroAtlas Fact Sheet template](#) in Microsoft Word format.
2. Complete one Fact Sheet for each metric field or group of related metrics. Name the Fact Sheet file as you would like the metric name(s) to appear on the [EnviroAtlas Fact Sheet web page](#) by replacing spaces with underscores. Please keep the name descriptive, but brief. EnviroAtlas reserves the right to modify the Fact Sheet file and metric name(s) based on EnviroAtlas requirements. The following links are for [national metric](#) or [community metric](#) data fact sheet examples.

## Inform EnviroAtlas that you have delivered data (Required):

1. Email the [EnviroAtlas Data Submission Group](#) and include your data and metadata submission. To assist the EnviroAtlas Data Submission Group with data quality control, please indicate the number of records included in the database. Also state, if there is a preference, where the data layer(s) could be placed within the EnviroAtlas map viewer Table of Contents. EnviroAtlas reserves the right to modify the placement of data layer(s) based on EnviroAtlas requirements. If your data submission files are too large (greater than 5mb) to email, please contact the [EnviroAtlas Data Submission Group](#) for additional instructions. EnviroAtlas is currently working to provide an alternative solution for transferring large files to the EnviroAtlas Data Submission Group. **Prior to submission, please review your data for completeness and quality. In the rare case where data must be resubmitted, please resubmit entire data package (i.e., data, metadata, other ancillary information). The EnviroAtlas Data Submission Group cannot accept partial data submissions (such as resubmitting only one data field).**

## Provide detailed documentation describing the methodology used to create your data so data can be regenerated in the future and include guidelines for data display (Highly Recommended):

1. For some datasets, when ArcToolbox models or other scripts are not available, detailed information (i.e., beyond what is described in your metadata) about input and output datasets, processes used, and ancillary information for deriving data will be stored with the original data in EnviroAtlas, but these files will not be made publicly available. The format for this ancillary information can be text, MS Word, MS Excel, or PDF (Portable Document Format).

## Provide guidelines for data/metric display in the EnviroAtlas map viewer (Highly Recommended):

1. Providing specialized classification instructions for mapping non-standard data or raster data is helpful to the EnviroAtlas Data Submission Group. Classification instructions should include guidelines to process null or no data values and suggested classification schemes to portray the data values

correctly. The format for this ancillary information can be a text, MS Word, MS Excel file or an ArcGIS layer file.

### Provide ArcToolbox models and/or scripts to regenerate metrics in the future (EPA staff and EPA Contractors - Required\*. Partners/Affiliates - Highly Recommended):

1. If ArcToolbox models and/or scripts were used to create the submitted data, please provide these tools to the [EnviroAtlas Data Submission Group](#). EnviroAtlas will only be relevant if data are periodically updated as new source datasets become available. When possible, our goal is to automatically update EnviroAtlas layers with the release of new source datasets. By providing the EnviroAtlas Data Submission Group with a script or model used to create your data and descriptions of inputs (datasets and values), this automation becomes more realistic. The model or script can include multiple tools, but please submit a single model or script that will process the entire workflow.
2. Give the script or model a name similar to the dataset submitted.
3. Include a text or MS Word document containing additional descriptions or information necessary for implementing the tool or script. Include default values used, look-up tables, and data input descriptions.
4. Document the script or model for others to understand and implement.

*\*Note for EPA staff and EPA contractors: providing ArcToolbox models and/or scripts used to create your metric dataset is required. The models and scripts should be archived in the same directory with the data you archive (Please see data submission Archive information below).*

### Archive all data associated with creation of metrics (EPA staff and EPA Contractors - Required. Partners/Affiliates - Highly Recommended):

1. **Partners/Affiliates:** The [EnviroAtlas Data Submission Group](#) recommends that data used to create metrics for EnviroAtlas, including specialized information or project-generated data, be retained, documented, and archived by contributors for future use and proper data stewardship. At times, there may be important raw or intermediate datasets that should also be submitted to EnviroAtlas for archival. In this case, EnviroAtlas partners are encouraged to email the [EnviroAtlas Data Submission Group](#) to discuss submission details.
2. **EPA staff and EPA contractors:** Some data associated with metric creation must be archived with EnviroAtlas using the following steps. Data that should be archived includes data that are not readily publicly available and that were compiled as an intermediate step in developing the metric. These should be limited to non-readily available data that would be necessary to reproduce the metric data. For some datasets, when ArcToolbox models or other scripts are not available, detailed information (i.e., beyond what is described in the metadata) about input and output datasets, processes used, and



ancillary information for deriving data should also be created and archived. The format for this ancillary information can be text, MS Word, MS Excel, MS PowerPoint, or PDF (Portable Document Format).

- a. Several directories have been created to store archival materials. The directories are organized by geographic extent (National, Communities, and Special\_Studies). For Communities and Special\_Studies, additional subdirectories have been created that reflect each community and/or special study area. Please locate the appropriate subdirectory on the ORD network shared drive, Aa\Ord\Las\Data\Priv\LEB\Projects\A-H\EnviroAtlas and create a new directory to store your archived data and scripts. Your new directory name should match your metric table name (e.g., iTree), and it must be unique relative to other pre-existing metric directories.
  - b. In your new directory, create a subdirectory to help identify team members associated with your data. The directory should be named using the following convention: \_LEAD LAST NAME\_NEXT TEAM MEMBER LAST NAME\_etc. (e.g., \_JACKSON\_SEARS). This subdirectory will remain empty and is only used to identify staff responsible for directory contents.
  - c. Create a "readme" file in your new directory. The file should be named "\_README.xxx" where ".xxx" is the file extension created by the software used to create the file (e.g, ".txt" for a text file, ".docx" for a Word file, etc.). The underscore at the beginning of the file name will ensure that the \_README file is shown at the beginning of your directory's file list. The \_README file should include a brief description of your new directory's contents.
  - d. Prior to copying your archival materials to your new directory on the shared drive, compress all data, models, and scripts into one ZIP file (using a software package such as WinZIP or using the compression tool in Windows).
  - e. Rename the \*.zip file to include the date the metric was submitted in the following format: metrictablename\_20120930.zip (YYYYMMDD, where Y = year, M = month, D = day). Adding the date to the ZIP file name will assist with version control.
  - f. Copy the ZIP file to your new directory on the shared drive. If you need assistance accessing this drive, please contact the [EnviroAtlas Data Submission Group](#).
3. For more information or if you have additional questions, please email the [EnviroAtlas Data Submission Group](#).

### Provide widgets for the EnviroAtlas interactive map viewer (Optional):

1. As of January 2015, the EnviroAtlas map viewer uses the [ArcGIS Viewer for Flex](#) version 3.4. EnviroAtlas will begin to implement [ArcGIS for JavaScript API](#) in the near future.
2. For special cases, contributors may create specialized widgets using Flash Builder and Esri's supplied source code and documentation.
3. Include the source code when submitting the completed widget.

For specifics on interacting with existing EnviroAtlas layers, for information about the current EnviroAtlas development environment, and for discussion about future code work in JavaScript, please contact the [EnviroAtlas Data Submission Group](#).

## Appendices: Creating EnviroAtlas Metadata with the EPA Metadata Editor

The [EPA Metadata Editor \(EME\)](#) was designed to help users create and edit metadata records that meet EPA Geospatial Metadata Technical Specification and Federal Geographic Data Committee (FGDC) content standards. Metadata created for EnviroAtlas data must pass EME validation – this ensures EPA and FGDC requirements are met and allows for the metadata to be published for public discovery in the [EPA's Environmental Dataset Gateway \(EDG\)](#) and Data.gov. EnviroAtlas requires using the EPA Metadata Editor (EME) to create metadata associated with your EnviroAtlas dataset contributions. The EME includes two components: 1) the user interface used for creating and editing metadata, and 2) a customizable Microsoft Access database that supplies default information to the user interface. To assist EnviroAtlas data contributors with metadata creation, an Access database specific to the EnviroAtlas and an EnviroAtlas metadata template, which includes several pre-populated fields, are available for use.

*An important note before you dive in: creating meaningful metadata is critical to the credibility and success of EnviroAtlas and we want to support your metadata creation efforts. Please [contact](#) the EnviroAtlas Data Submission Group for assistance with any aspect of metadata creation from installing the EME to completing metadata fields and validation.*

## Appendix A. Getting Started with the EME and the EnviroAtlas Custom EME Access Database and Metadata Template

1. Download and install the EPA Metadata Editor (EME) v3.2.1.
  - a. Click [here](#) to download the EME. The latest version of the EME for geospatial records requires installation of Microsoft .NET Framework 3.5 and ArcGIS 10 or higher.
  - b. If using the EME v3.2.1 within Esri's ArcGIS 10 or higher, this software patch is required: [FGDC Metadata Style Patch](http://resources.arcgis.com/content/patches-and-service-packs?fa=viewPatch&PID=160&MetalD=1637) (<http://resources.arcgis.com/content/patches-and-service-packs?fa=viewPatch&PID=160&MetalD=1637>). Prior versions of the EPA EME for use with ArcGIS 9.2 or 9.3 are available (*if needed, please [contact](#) the EnviroAtlas Data Submission Group*).
  - c. The EME functions as a standalone application and also within Esri's ArcCatalog. To learn more about the EME and general metadata creation, review the EME training slides and EME Fact Sheet available under the <https://edg.epa.gov/EME/resources.html> on the EME website.
  - d. Once installed, you may view the EME documentation (click the "Help" button on EME interface).
  - e. Please be sure to follow our EnviroAtlas-specific metadata instructions when creating your metadata. Additional information about EnviroAtlas metadata creation is included in Appendix B, "Using the EME to Complete EnviroAtlas Metadata Records".
  
2. Download and install the customized EnviroAtlas Access database for use with the EME interface.
  - a. Click [here](#) to download the customized Access database.
  - b. Once you have successfully installed the EME (Step 1 above), you will need to replace the default EME Access database with the EnviroAtlas Access EME database.
    - i. Ensure the EME application is closed before replacing the default Access database. The default EME Access database in your user directory can be replaced with the EnviroAtlas Access EME database without administrator privileges on your computer. If you have difficulties replacing the default EME database, please contact the [EnviroAtlas Data Submission Group](#).
    - ii. When installing the EME, the default EME Access database (metadata.mdb) is placed in your user directory (e.g., "C:\Users\\AppData\Roaming\Innovate! Inc\EPA Metadata Editor v3.2.1\" for Windows 7 or "C:\Documents and Settings\\Application Data\Innovate! Inc\EPA Metadata Editor v3.2.1\" directory for Windows XP).
    - iii. If you cannot view the Application Data folder in your user directory, go to Tools>Folder Options>View tab and choose "Show hidden files, folders, or drives" under the "Hidden files and folders" header.).
    - iv. Replace the default EME Access database with the EnviroAtlas Access EME database. **Do not change the Access database name – the database must be named metadata.mdb.** This will make EnviroAtlas-specific metadata field options available to you through the EME interface.
  
3. Download the EnviroAtlas metadata template.

- a. Click [here](#) to download the EnviroAtlas metadata template.
  - b. Once you have installed the EME and replaced the default EME Access database with the EnviroAtlas EME database, open the EnviroAtlas metadata template in the EME (within the EME, select File> Open> and browse to EnviroAtlas\_metadata\_template\_EME321.xml).
  - c. The EnviroAtlas metadata template includes several pre-populated fields, some of which will be updated based on dataset contents; other fields will not be altered. Please note: the metadata template includes completed coordinate system and projection metadata elements. It is important to use the metadata template with these pre-populated elements when creating your metadata. Additional information about updating appropriate metadata fields is provided in Appendix B, “Using the EME to Complete EnviroAtlas Metadata Records”.
4. To assist with your understanding of EnviroAtlas metadata requirements, an example metadata record is available for download from any EnviroAtlas data layer or by searching for “EnviroAtlas” at <http://edg.epa.gov>.

## Appendix B: Using the EME to Complete EnviroAtlas Metadata Records

The EME interface includes three tabs: **Tab 1** – Basic Dataset Information, **Tab 2** – Quality, Coordinate System, and Attribute Information, and **Tab 3** – Distribution and Metadata Information. The following sections describe individual tabs and associated fields that require input or modification based on EnviroAtlas metadata requirements.

The EME includes “Default” buttons throughout the interface. This button sets the selected field to the default value stored in the EnviroAtlas custom EME Access database. Note: the default EnviroAtlas values are also pre-populated in the EnviroAtlas metadata template.

If using acronyms or abbreviations in the submitted metadata record, please include their definitions. For example, MRLC would be denoted as Multi-Resolution Land Characteristics Consortium (MRLC).

After you have completed all metadata fields for your dataset, you will validate the metadata to ensure compliance with metadata standards. This process is explained in greater detail in the following sections.

## Completing Fields in the EME Tab 1 – Basic Dataset Information

The screenshot shows the EPA Metadata Editor v3.2.1 interface. The 'Basic Data Set Information' tab is active. The interface is divided into several sections:

- Citation:** \* Origin: US Environmental Protection Agency (US EPA) (Step 1a), \* Title: EnviroAtlas - (Step 1b)
- Publisher:** \* Published by: US EPA Office of Research and Development (ORD) - Natic (D), \* Published at: Research Triangle Park, NC, \* Date: 2012 (today)
- Online Linkage:** \*\* Primary Linkage: https://edg.epa.gov/data/PUBLIC/ORD/ENVIRO/ (D, ✓), \*\* Secondary Linkage: http://enviroatlas.epa.gov/arcgis/rest/services (D, ✓)
- Description:** \* Abstract: (Step 2a), \* Purpose: (Step 2b), Supplemental Info: (Step 2c)
- Time Period:** \* Date of Data Set (Step 3a), Single Date OR Multi Dates: Date1, Date2, OR Range of dates: Date1 - Date2, \* Progress of data: Complete, \* Data currency: Publication date, \* Update frequency: As needed
- Bounding Box:** N: 51.6773602 E: -65.20336449 S: 22.7348894 W: - (D), \* N: 51.6773602 (Step 4a), \* S: 22.7348894, \* W: -128.02839229
- Keywords:** ISO, EPA, EnviroAtlas, Place. List includes: Drivers of Change - Ecosystem Services, Ecology, Economy, Ecosystem, Ecosystem Services (Step 5a/b), Elevation, Emergency, Endangered, Energy, EnviroAtlas, Environment, Environmental Atlas (Help, D)
- Data Set Constraints:** \* Access: None (D), \* Use: None. Please check sources, scale, acc. (D), \* Security Classification: public (from metadata record) (D)
- Contact:** Primary Person (selected), Primary Organization, U.S. Environmental Protection Agency, Office of R&D (D)

Legend: YELLOW \* mandatory, GREEN \*\* mandatory if applicable, BLUE optional. Buttons: Save, Save & Close, Cancel.

Figure 1. EPA Metadata Editor Tab 1 – Basic Data Set Information. Green outline boxes denote metadata fields to complete. Green text corresponds to the number assigned to each Tab 1 requirement (see information below) to be completed by EnviroAtlas data provider. Blurred areas are default settings completed by using the EnviroAtlas metadata template for EME version 3.2.1.

### 1. Citation

- Complete Origin (Required).** Add the name of the organization that developed the data. If US EPA is not the originator, remove the provided default text, “US Environmental Protection Agency (US EPA)”, and type the name of the organization that developed the data.

Example Origin: NatureServe, US Geological Survey (USGS), US Forest Service (USFS), New Mexico State University.

- b. Complete Title (Required).** Add a title that represents the data content. Begin by typing descriptive text after the words, "EnviroAtlas". Include the geographic location (e.g., Durham, NC) or name of special study area (e.g., LIPS - Midwest) at the end of the title. The text, "EnviroAtlas" must remain at the beginning of the dataset title.

Example Title: "EnviroAtlas – Domestic Water Demand by 12-Digit HUC for the United States"

## 2. Description

- a. Complete Abstract (Required).** Add a narrative summary about the dataset and include general information about EnviroAtlas (For general EnviroAtlas information, please see the green highlighted text below in the Example Abstract. Please copy this general information into your Abstract, but be sure to replace "US EPA" with your organization.). Begin by typing narrative text directly into the Abstract text input field. If desired, double-click in the box to open a larger version of the text input box. Descriptive text can be added by typing directly into the input field or by copying and pasting from another source. TIP – to populate the Abstract field, you may want to use information included in your metric Fact Sheets under "Overview".

Example Abstract:

"This dataset includes national domestic water demand attributes that provide insight into the amount of water currently used for indoor and outdoor residential purposes in the contiguous United States. The values are based on 2005 water demand and 2010 population distribution and have been summarized by watershed or 12-digit hydrologic unit code (HUC). For the purposes of this metric, domestic water use includes residential uses, such as for drinking, bathing, cleaning, landscaping, and pools. Depending on the location, domestic water can be self-supplied, such as by private wells, or publicly-supplied, such as by municipalities. Sources include either surface water or groundwater. Estimates are for primary residences only (i.e., excluding second homes and tourism rental). **This dataset was produced by the US EPA (in place of "US EPA", insert your organization) to support research and online mapping activities related to EnviroAtlas. EnviroAtlas (<http://enviroatlas.epa.gov>) allows the user to interact with a web-based, easy-to-use, mapping application to view and analyze multiple ecosystem services for the contiguous United States. The dataset is available as downloadable data (<https://edg.epa.gov/data/Public/ORD/EnviroAtlas>) or as an EnviroAtlas map service. Additional descriptive information about each attribute in this dataset can be found in its associated EnviroAtlas Fact Sheet (<http://enviroatlas.epa.gov/EnviroAtlas/DataFactSheets>).**"

- b. Complete Purpose (Required).** Add descriptive text describing the purpose of the dataset and general information about EnviroAtlas (for general EnviroAtlas information, please see the green highlighted text below in the Example Purpose. Please copy this general information into your Purpose.). Begin by typing narrative text directly into the Purpose text input field. If desired, double-click in the box to open a larger version of the text input box. Descriptive text can be added by typing directly into the input field

or by copying and pasting from another source. TIP – to populate the Purpose field, you may want to use information included in your metric Fact Sheets under "How could I use this information?"

Example Purpose:

"This data could be used in conjunction with other data to identify potential imbalances and trends between supply and demand. In areas with significant imbalances or detrimental trends, measures to further understand and alleviate pressure on the water supply could be implemented. In subsequent updates to EnviroAtlas, future demand scenarios will be incorporated based on projected population increases. **The overall goal of EnviroAtlas is to employ and develop the best available science to map indicators of ecosystem services production, demand, and drivers for the nation.**"

- c. **Complete Supplemental Info (Not Required – may be left blank).** Add descriptive text describing the purpose of the dataset. This is a good location to add descriptive information about data caveats or limitations. Begin by typing text directly into the Supplemental Info text input field. If desired, double-click in the box to open a larger version of the text input box. Descriptive text can be added by typing directly into the input field or by copying and pasting from another source. TIP – to populate the Supplemental Info field, you may want to use information included in your metric Fact Sheets under "What are the limitations of the data layer?"

Example Supplemental Info:

"Reporting complexities exist within these data. For example, coefficients are derived based on available data, such as from public suppliers, and they provide estimates for areas in which there might be less data available, such as self-supplied well withdrawals. People routinely cross reporting boundaries for activities, such as for work or vacation. In addition to these data complexities, another limitation in estimating water use is the inherent nature of water withdrawals and delivery. Self-supplied withdrawals (e.g., private wells) are calculated in the county in which the withdrawals occur. Publicly-supplied deliveries (e.g., from a municipality) may cross county, even multiple county or municipal, borders. For example, surrounding or upstream counties may supply urban counties, such as New York, NY, or Washington, DC, that have limited access to local natural resources. Considerable effort and diligence have been taken by the US Geological Survey (USGS) in standardizing, acquiring, estimating, and reporting the most accurate available data. These data represent the best available data."

### 3. Time Period

- a. **Complete Date of Data Set (Required).** Add the date that reflects the time period of the dataset. The date information can be recorded as a single date (201207026), range of dates (20110726 - 20120726), or individual dates (20100726, 20110726, 20120726). Date information should be documented using the following date format (YYYY, YYYYMM, or YYYYMMDD, where Y = year, M = month, D = day). No slashes are permitted. If date is unknown, then enter the text "Unknown". Please note that the example metadata record demonstrates how to denote multiple dates within the Date element. This is not a requirement, but just part of the example - multiple dates can be denoted if desired.

#### 4. Bounding Box

- a. **Select Bounding Box (Required).** Select the bounding box text from the “N: S: E: W:” pull down menu that best describes the spatial extent of your dataset. For community data, please select the state where the community resides – standardized values will be updated later. The N, S, E, and W fields are automatically populated once a selection has been made from the pull down menu. Please do not modify the values in the N, S, E, and W fields.

#### 5. Keywords

- a. **Select EnviroAtlas Tab Keywords (Required).** Choose one or more keywords from the EnviroAtlas keyword tab (labeled "EnviroAtlas") that generally describe the content of your dataset. The EnviroAtlas keyword list was created specifically for EnviroAtlas and is included in the customized EnviroAtlas EME database. Several default keywords have been pre-selected in the metadata template – please do not remove these selections. However, add to the default EnviroAtlas keywords by holding the Ctrl key and using your mouse to select additional keywords of interest. To ensure proper validation of your metadata record, EnviroAtlas staff has also added default keywords to the metadata template for the ISO and EPA keyword tabs. Please do not remove or add to the ISO or EPA default selections.
- b. **Select Place Tab Keywords (Required).** Choose one or more keywords from the Place keyword tab (labeled "Place") that generally describe the geographic location characterized by your dataset. To choose multiple keywords, hold the shift key and use your mouse to select additional keywords. For example, if your dataset is for Durham, NC, then select “Durham, NC” and while holding the shift key, also choose “North Carolina”. Making both selections will aid in the data discovery process. If you are documenting a dataset for a special study area, please choose all states that intersect the special study area. For example, for the special study area, “LIPS-Midwest”, select Illinois, Indiana, Iowa, Kansas, Kentucky, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, and Wisconsin. If you are documenting a dataset that encompasses the entire US, please select "United States" and then select each state from the Place keyword list (TIP – to select all Place keywords, select the first Place keyword (i.e., Alabama), scroll to the end of the Place keyword list, hold down the shift key, and also select the last Place keyword (i.e., Wyoming). Then, hold down the control key and deselect any Place keywords that do not apply.). Adding each state Place keyword (along with the Place keyword, “United States”) for US datasets will assist with data discovery, especially if a potential data user searches by state name.



## Completing Fields in the EME Tab 2 – Quality, Coordinate System, and Attribute Information

The screenshot shows the EPA Metadata Editor v3.2.1 interface. The main window is titled "EPA Metadata Editor v3.2.1 - D:\work\EnviroAtlas\FY2015\_docs\Metadata\_files\EnviroAtlas\_metadata\_template\_EMEnv321\_v2.xml". The menu bar includes File, Edit, Tools, and Help. The tabs are "Basic Data Set Information", "Quality, Coordinate System, and Attribute Information", and "Distribution & Metadata Information".

The "Quality" section includes:
 

- Integrity Tests:** Tests for integrity have not been performed. (D)
- Completeness of Data:** Features represented have not been tested for. (D)
- Horizontal Positional Accuracy:** Data were collected using methods that have unknown accuracy. (D)
- Vertical Positional Accuracy:** (D)

The "Coordinate System Information" section includes:
 

- Horizontal:**
  - Projection: Map Projection (D)
  - Projection Name: (D)
  - Units: meters (D)
  - Datum: North American Datum of 1983 (D)
  - Resolution: X: 0.000001, Y: 0.000001 (D)

The "Entity and Attribute Information" section includes:
 

- Overview Description:** (D)
- Citation:** <http://enviroatlas.epa.gov/EnviroAtlas> (D)

The "Lineage" section includes:
 

- Source Information:** No sources recorded. (D)
- Processing Steps:** No processing steps recorded. (\* only 1 processing step required) (D)

The "Spatial Data Organization Information" section includes:
 

- Direct Spatial Reference:** (D)
- Indirect Spatial Reference:** (D)

Green outlines and text indicate fields to be completed by the user:
 

- Step 1a:** Lineage section.
- Step 2a/b:** Entity and Attribute Information section.

Legend: YELLOW \* mandatory, GREEN \*\* mandatory if applicable, BLUE optional. Buttons: Save, Save & Close, Cancel.

Figure 2. EPA Metadata Editor Tab 2 – Quality, Coordinate System and Attribute Information. Green outline boxes denote metadata fields to complete. Green text corresponds to the number assigned to each Tab 2 requirement (see information below) to be completed by EnviroAtlas data provider. Blurred areas are default settings completed by using the EnviroAtlas metadata template for EME version 3.2.1.

### 1. Processing Steps

- a. **Complete Processing Steps – Date and Description (Required).** \* Processing steps detail information about data processing, conversions, and/or transformations. Each processing step includes a date and description. If you cannot recall a specific date, a general date is permitted (i.e., 20120101), but the following date format must be maintained. Format the date as YYYYMMDD, where Y = year, M = month, D = day (no slashes are permitted). No matter which input method is selected (Please read \*note below), you should include detailed processing steps that thoroughly describe data sources and how your dataset attributes were derived. This information allows the process to be repeated.

### Example Processing Steps (Date and Description):

20111128 Acquired 2005 water use data from USGS.

20111201 Calculated domestic water use per capita by county using USGS water use spreadsheet data from 2005, resulting in gallons per person per day per county.

20111209 Joined domestic water use per capita results to Esri's county-level data layer.

20111216 Calculated median domestic use by state using Zonal Statistics.

20111223 Multiplied state median results by dasymetric data (Tim Wade, Landscape Ecology Branch, US EPA) using the Raster Calculator.

20111230 Summarized dasymetrically distributed values by 12-digit HUC (Watershed Boundary Dataset, March 2011) using Zonal Statistics.

20120127 Adjusted Null and 0 value HUCs to account for water bodies in National Land Cover Database (NLCD).

*\*Note: The EME includes two metadata elements that allow for the addition of information describing how your data was processed to generate metrics. The first element is the Processing Steps, where step-by-step information and associated dates are provided on separate lines. The second element is Entity and Attribute Overview Description text input box. The Overview Description text input box allows for all metric-creation steps to be included in one block of text that can be typed directly in the text box or copied and pasted from other documents.*

*EnviroAtlas allows you to choose between one of two methods for describing the steps used to create your metrics: 1) Add comprehensive steps using the Processing Steps metadata element and include a note in the Entity and Attribute Overview text input box that indicates that a comprehensive list of metric creation steps is included in the Processing Steps or 2) Complete a short, generalized list of processing steps in the Processing Steps metadata element and also complete a detailed description of your processing steps in the Entity and Attribute Overview Description metadata element. Option 2 is shown in the Processing Steps and Entity and Attribute Overview Description examples given in this document. Processing steps do not have to be described in both locations but must be described in at least one location.*

## 2. Entity and Attribute Information

This section of the metadata record is used for describing data creation methods (Overview tab, Overview Description) and documenting each attribute in the dataset (Detailed tab). The Overview Description allows for the input of text that thoroughly describes the methodology used to create your dataset and associated attributes. The tab marked Detailed provides a method to individually describe each attribute using a standard format that includes the attribute name, description, and domain.

- a. **Complete Overview Description (Required)\*\*.** The Overview tab includes the Overview Description text input field. Begin by typing text directly into the Overview Description input field. If desired, double-click in the box to open a larger version of the text input box. Descriptive text can be added by typing directly

into the input field or by copying and pasting from another source. Include detailed information that thoroughly describes data sources and how your dataset attributes were derived. This information allows the process to be repeated.

Example Overview Description:

"In general terms, domestic use per capita per day per county was estimated using United States Geological Survey (USGS) 2005 Water Use data. A median per capita use by state was calculated to smooth outlying county values. These median values were weighted according to the estimated population per 30 meter square or pixel on the ground. Water use was summarized for each 12-digit HUC or watershed. More specifically, the initial data used to provide this metric is from the USGS. The USGS estimates a variety of water use data at 5-year intervals through the National Water Use Information Program (NWUIP). Data are collected at the state level by USGS water-use study chiefs and estimates are provided at the county level. Compilation methods vary slightly by state, depending on available data. Coefficients are derived and applied as needed. Overall guidelines used by the water-use study chiefs and the original data are available through the NWUIP website. To calculate the domestic use per capita by county, the total domestic use was divided by the total population served per county and adjusted as applicable to represent the gallons used per person per day.  $(DO\_Total * 1,000,000) / (TP\_TotPop * 1000)$  where a) DO\_Total is the Domestic Total Use (Withdrawals + Deliveries), b) 1,000,000 converts the domestic use from millions of gallons of water per day to gallons of water per day, c) TP\_TotPop is the total population of the county, and d) 1,000 converts the total population from thousands to individuals. Due to reporting complexities, such as the use of coefficients and the fact that people routinely cross reporting boundaries, estimates sometimes lead to outlying county results of gallons of water per day per person. To reduce the impact of outliers in the county data, a median domestic use was calculated by state. The state median results were then applied to distributed population data. The distributed population data are known as dasymetric population data. The dasymetric data were derived by Tim Wade (Landscape Ecology Branch, U.S. EPA, 2011), using 2006 National Land Cover Database (NLCD) and 2010 US Census data. To represent these results in the EnviroAtlas, the water use values were then summarized by 12-digit HUC (Watershed Boundary Dataset, March 2011). To account for water bodies, Null and 0 sum HUCs were evaluated in correlation with 2006 NLCD water (11). Null and 0 sum HUCs containing at least 75% water were given sums of -1 to represent water in the EnviroAtlas. The remaining Null HUCs were reviewed, verified, and given sums of 0 in the EnviroAtlas. The remaining 0 sum HUCs remain unchanged in EnviroAtlas."

- b. Complete Detailed information (Required).** The tab marked Detailed includes multiple input fields that allow the entry of very specific information about individual dataset attributes. Include the dataset name (Entity Information, Label), a definition of the dataset (Entity Information, Definition), and the source of the dataset definition (Entity Information, Definition Source). This same information must be included for each dataset attribute as well as the domain for each attribute. Additional information that describes each input field can be located in the EME Help document. To access the overall EME Help documentation, click the word "Help" on the EME interface. To open the EME Help documentation for a

specific topic, click the text associated with the field of interest (for example, click on “Domain Information” to open the EME Help documentation relevant to Domain Information).

Example Detailed Information:

Entity Information:

Label - DWDbyHUC.csv

Definition - Domestic Water Demand by 12-Digit HUC

Definition Source - US EPA

Attribute Information:

Label - DWDbyHUC

Definition - Sum of the domestic water demand within the HUC12 (gallons of water per day). -1 indicates >75% and no applicable or zero dasymetric data.

Definition Source - US EPA

Domain Information:

Range - Min:-1; Max: 287,472,000

*\*\*Note: The EME includes two metadata elements that allow for the addition of information describing how your data was processed to generate metrics. The first element is the Processing Steps, where step-by-step information and associated dates are provided on separate lines. The second element is Entity and Attribute Overview Description text input box. The Overview Description text input box allows for all metric-creation steps to be included in one block of text that can be typed directly in the text box or copied and pasted from another document (such as your Fact Sheet).*

*EnviroAtlas allows you to choose between one of two methods for describing the steps used to create your metrics: 1) Add comprehensive steps using the Processing Steps metadata element and include a note in the Entity and Attribute Overview text input box that indicates that a comprehensive list of metric creation steps is included in the Processing Steps or 2) Complete a general list of metric creation steps in the Processing Steps metadata element and also complete a detailed metric creation description in the Entity and Attribute Overview Description metadata element. Option 2 is shown in the Processing Steps and Entity and Attribute Overview Description examples given in this document.*

### **3. Coordinate System Information (Required, but does not need any data provider action).**

Coordinate system information is required for proper metadata validation. However, the current version of the EME offers limited projection types and does not include the Albers projection required for data submission to EnviroAtlas. The EME coordinate system information fields do not accept text input, therefore, Albers projection information cannot be input using this method. However, the correct Albers coordinate information has been included with the EnviroAtlas metadata template; therefore, you do not need to complete this information when using the template.

## Completing Fields in the EME Tab 3 – Distribution and Metadata Information

EPA Metadata Editor v3.2.1 - D:\work\EnviroAtlas\FY2015\_docs\Metadata\_files\EnviroAtlas\_metadata\_template\_EMEnv321\_v2.xml

File Edit Tools Help

Basic Data Set Information Quality, Coordinate System, and Attribute Information **Distribution & Metadata Information**

**Distribution Information**

\* **Distribution Contact**

Primary Person  Primary Organization

U.S. Environmental Protection Agency, Office of Research and **D**

**Data Resource Type**

\* **Type of data set?** Downloadable Data **D**

\* **Distribution Liability** **D**

Although these data have been processed successfully on a computer system at the Environmental Protection Agency, no warranty expressed or implied is made regarding the accuracy or utility of the data on any other system or for general or scientific purposes, nor shall the act of distribution constitute any such warranty. It is also strongly recommended that careful attention be paid to the contents of the

**Metadata Information**

\* **Metadata Date:**  **Step 1a**

\* **Metadata Future Review Date:**  **Step 1b**

\* **Metadata Contact**

Primary Person  Primary Organization

U.S. Environmental Protection Agency, Office of Research and **D**

**Metadata Standard**

\* **Standard Name:** FGDC Content Standard for Digital Geospatial Metadata

\* **Standard Version:**  **D**

YELLOW \* mandatory GREEN \*\* mandatory if applicable BLUE optional Click on text to link to element description Save Save & Close Cancel

Figure 3. EPA Metadata Editor Tab 3 – Distribution & Metadata Information. Green outline boxes denote metadata fields to complete. Green text corresponds to the number assigned to each Tab 3 requirement (see information below) to be completed by EnviroAtlas data provider. Blurred areas are default settings completed by using the EnviroAtlas metadata template for EME version 3.2.1.

## 1. Metadata Information

- a. **Complete Metadata Date (Required).** To complete this field, please click the "today" button to the right of the Metadata Date input field.
- b. **Complete Metadata Future Review Date (Required).** To complete this field, please click the "4 yrs" button to the right of the Metadata Future Review Date input field.

### Final Metadata Steps – Spell Check and Validation (Required)

Now that the metadata record is completed for all metadata fields in the dataset, you must spell check and validate the metadata record.

To run the EME spell checker, click on the "Edit" menu and select "Spell Check". Please repeat the spell check function **for each tab** and complete updates as necessary.

Properly validated metadata ensures that your metadata record can be published to the EDG. To validate the metadata, click on the "Tools" menu and select "Validate Metadata". Each metadata field that does not validate properly is highlighted in the user interface with a red balloon symbol. When you use your mouse to hover over each red balloon symbol, specific information about the error is displayed. Use this information to update and re-validate the metadata field until validation is successful. Typical validation errors include incorrect date formatting or blank metadata fields. For more information about validating metadata using the EME, open the EME help documentation to the section, "Getting Started, Using the EPA Validator".

Once your metadata validates successfully, save your final record (File > Save as) and provide a copy to the [EnviroAtlas Data Submission Group](#) with your dataset.