

Metadata Template Documentation

This document provides guidance for completing each element within the metadata template directory. Within the top level of the directory you will find the following assets for recording your projects metadata:

- methods.docx
- abstract.docx
- metadata.xlsx

Additionally, within the metadata template directory is a sub directory with completed examples.

If you have any questions regarding metadata, please contact the acting CVPIA data managers at ecain@flowwest.com, erodriguez@flowwest.com, or sgill@flowwest.com.

Methods

In “method.docx” describe the methods followed in the creation of the dataset, including description of field, laboratory and processing steps, sampling methods and units, quality control procedures. What were the actual procedures that are used in the creation or the subsequent processing of the dataset? Also, describe processes that have been used to define or improve the quality of a data file, or to identify potential problems with the data file.

Please be specific, include instrument descriptions, or point to a protocol online. If this is a data compilation please specify datasets used, preferably their DOI or URL plus general citation information.

If the project includes more than one dataset, make a section in the document for each dataset and the method used.

Abstract

In “abstract.docx” give a brief overview of the data resource, dataset, and/or project. The abstract will be used for full-text searches, and it should be rich with descriptive text. In particular, descriptions should include information that does not fit into structured metadata, and focus on the “what”, “when”, and “where” information, general taxonomic information, as well as whether the dataset is ongoing or completed. Some general methods description is appropriate, and broad classes of measured parameters should also be included. For a large number of parameters, use categories instead of listing all parameters (e.g. use the term “nutrients” instead of nitrate, phosphate, calcium, etc.), in combination with the parameters that seem most relevant for searches.

The abstract can be at the project level or you can provide an abstract for each dataset associated with the project.

Metadata Workbook

All other information will be inputted into metadata.xlsx. Each section below corresponds to a sheet of the xlsx document. Columns that are gray are optional and columns that are blue are conditionally required but every other column must be filled out. Some columns contain drop down menus of acceptable values. In these columns you must input one of the predefined options listed in the dropdown menu.

If multiple data tables are associated with your project, fill out a different metadata.xlsx for each data table. If methodology or abstract differ between the data tables, you can add subsections with headers labeled for each data table to the abstract and methods documents.

Dataset

In the dataset sheet you will provide information on the name and type of dataset. If the type of your dataset is not tabular or vector contact the acting CVPIA data managers and we will help you prepare the metadata. Columns that are white are required and columns that are blue are conditionally required.

- **name** - Name of the dataset you are submitting
- **type** - Type of data (tabular, vector, raster)
- **geometry (conditionally required)** - **Required only if the dataset is type vector** EML schema accepts eight geometry values: Point, LineString, LinearRing, Polygon, Multipoint, MultiLineString, MultiPolygon, MultiGeometry.

Spatial Data

Vector For a dataset that contains vector data please fill out the metadata, abstract, and methods docs following the above instructions. Make sure to indicate type as vector in the dataset tab and to include a geometry in the geometry tab. EML schema accepts eight geometry values: Point, LineString, LinearRing, Polygon, Multipoint, MultiLineString, MultiPolygon, MultiGeometry.

Raster For a dataset that contains raster data please contact the acting CVPIA data managers at ecain@flowwest.com, erodriguez@flowwest.com, or sgill@flowwest.com. They will provide you with more information on how to format the metadata for raster data.

Personnel

In the personnel sheet you will provide information on the creator and the associated parties to the dataset. The creator is any person or organization who is responsible for the creation of the data, the creator will also be the contact for this dataset. You must have one person with the role “creator”. Columns that are gray are optional but every other column must be filled out.

- **first_name** - Person’s given name
- **last_name** - Person’s given last name
- **email** - Person’s email address
- **role** - Select the appropriate role from the dropdown, contact CVPIA data manager if you would like a role added to the template options
- **organization** - The organization that employs the person
- **orcid (optional)** - ORCID ID is a persistent digital identifier for researchers, register at <http://orcid.org/>

Title

The title should be fairly descriptive and between 7 and 20 words long. The short name must be less than the number of words present in the title and is your opportunity to give viewers a more accessible name to the dataset.

- **title** - A brief description of the dataset, providing enough detail to differentiate it from other similar datasets. For example: “Vernal pool amphibian density data, Isla Vista 1990-1996”.
- **short_name** - The short name provides a concise name that describes the dataset being documented. For example “vernal-data-1999”.

Keyword Set

The keyword set should include a list of keywords and the name of the controlled vocabulary registry (keyword thesaurus) they belong to. Keywords help users identify your dataset and the use of controlled vocabulary sets are helpful to keep keywords consistent across many different research efforts. If you choose not to use one of the controlled vocabularies linked to below, leave keywordThesaurus blank.

- **keyword** - The keyword
- **keywordThesaurus (conditionally required)** - **Required only when keyword is from a controlled vocabulary.** A string identifying the controlled vocabulary the keywords originate from, do not include if keywords are not from a controlled vocabulary

In order to promote consistency, please search the following resources for keywords:

- LTER - Long Term Ecological Research.
- AGROVOC - A controlled vocabulary covering all areas of interest of the Food and Agriculture Organization (FAO) of the United Nations, including food, nutrition, agriculture, fisheries, forestry, environment.
- U.S. Board on Geographic Names - USGS place names dictionary.

License

The intellectual rights information associated with the dataset. For projects funded under CVPIA authority, we have preselected two licenses, CC0 and CC BY. You can select either of these by adding “CC0” or “CCBY” in the default_license column and leave the rest of the columns blank.

CC0 - The most permissive license, appropriate for data in the public domain.

CC BY - Attribution required

If neither of these two licenses fit the intellectual right associated with the dataset you must provide information for all of the conditionally required columns:

- **default_license** - Select the default license if applicable to your project.
- **license_name (conditionally required)** - The official name of the license that applies to the data and metadata
- **license_url (conditionally required)** - The persistent URL for the license
- **license_identifier (conditionally required)** - The official identifier for the license, which should be drawn from the SPDX license list, or a similar well-known license registry
- **intellectual_rights_description (conditionally required)** - Description of rights under the license

Example:

- **license_name:** Creative Commons Attribution Non Commercial Share Alike 4.0 International
- **license_url:** <https://spdx.org/licenses/CC-BY-NC-SA-4.0.html>
- **license_identifier:** CC-BY-NC-SA-4.0 ***intellectual_rights_description:** This license lets others remix, adapt, and build upon your work even for commercial purposes, as long as they credit you and license their new creations under the identical terms. This license is often compared to “copyleft” free and open source software licenses. All new works based on yours will carry the same license, so any derivatives will also allow commercial use. This is the license used by Wikipedia, and is recommended for materials that would benefit from incorporating content from Wikipedia and similarly licensed projects.

Project

Use this sheet to input a project personnel if you wish to specify a specific project personnel. If project personnel is left blank we will automatically assign project personnel to the prior defined personnel with the role of creator. To give a different project personnel please fill out all sections of this sheet. Columns that are blue are conditionally required and columns that are gray are optional.

- **first_name (conditionally required)** - Required if you want to add a specific project personnel Person's given name
- **last_name (conditionally required)** - Required if you want to add a specific project personnel Person's given last name
- **email (optional)**- Required if you want to add a specific project personnel Person's email address
- **role (conditionally required)** - Required if you want to add a specific project personnel Select the appropriate role from the dropdown, contact CVPIA data manager if you would like a role added to the template options
- **organization (conditionally required)** - Required if you want to add a specific project personnel The name of the organization that the person is employed by
- **orcid (optional)** - ORCID iD is a persistent digital identifier for researchers, register at <http://orcid.org/>

Funding

Funding information of the dataset. Use this sheet to describe the funding awarded to the project. We have provided a set of CVPIA default funders: "USBR", "CDFW", "CDWR". If you input one of the previous listed in the **funder_name** column you must still provide an **award_title** but you may leave the remaining columns blank.

If the default funding options are not applicable to your project you must fill out all required and conditionally required columns. Columns that are blue are conditionally required and columns that are gray are optional but every other column must be filled out.

- **funder_name** - Organization or individual providing the funding. If applicable select a default funder from the dropdown list.
- **funder_identifier (optional)** - CVPIA specific finding identifiers are in development.
- **award_number (conditionally required)** - The identifier assigned by the funding agency to identify this funding award. Required if default funder is not selected.
- **award_title** - Title of the dataset or project which received funding.
- **award_url (optional)** - to include a link to information about the funding award on the funding organization's webpage.

- **funding_description (optional)** - To provide a short description of the funding received.

Example:

- **funder_name** - National Science Foundation
- **funder_identifier** - <http://dx.doi.org/10.13039/1000000001>
- **award_number** - 1656026
- **award_title** - LTER: Beaufort Sea Lagoons: An Arctic Coastal Ecosystem in Transition *award_url* - https://www.nsf.gov/awardsearch/showAward?AWD/_ID=1656026 **funding_description**- BLE LTER is supported by the National Science Funder award #1656026 (2017-08-01 to 2022-07-31).

Maintenance

Maintenance information of the dataset. Use this sheet to describe the status of the data collection as well as the frequency at which you plan to update the data. All columns required.

- **status** - status of the project or dataset this can be either “ongoing” or “complete”
- **update_frequency (conditionally required)** - **Required if status is “ongoing”** provide the frequency of which the project or dataset is updated. Available update frequencies: daily, weekly, monthly, annual

Example (ongoing maintenance):

- **status**: ongoing
- **update_frequency**: monthly

Coverage

Geographical and temporal coverage associated with the dataset, all columns are required. Geographic and temporal will allow users to quickly search for your dataset based on geographic location and time.

- **geographic_description** - A description of the locations of research sites and areas related to the data
- **west_bounding_coordinate** - The west cardinality limit
- **east_bounding_coordinate** - The east cardinality limit
- **north_bounding_coordinate** - The north cardinality limit
- **south_bounding_coordinate** - The south cardinality limit
- **begin_date** - The starting date for the dataset or project. Dates must be provided in ISO 8601 format, YYYY-MM-DD
- **end_date** - The projected or actual end date for the dataset or project. Dates must be provided in ISO 8601 format, YYYY-MM-DD.

Example:

- **geographic_description**: North Slope drainage basin: Bounding box encompasses 42 drainage basins totaling the North Slope drainage basin, Alaska, USA
- **west_bound_coordinate**: -160.594000
- **east_bound_coordinate**: -134.1048
- **north_bound_coordinate**: 71.2383
- **south_bound_coordinate**: 67.865
- **begin_date**: 1980-01-01
- **end_date**: 2010-12-31

Taxonomic Coverage

Taxonomic coverage for the dataset, this sheet is optional. We have provided a set of CVPIA common taxa: "chinook", "delta_smelt", "white_sturgeon", "green_sturgeon", or "steelhead". If you select one of the previous listed in the **CVPIA_common_species** column you can leave the remaining columns blank. If you are adding a new taxon you must fill out all of the other columns. When adding a new taxon please use <https://www.itis.gov/> for full taxonomic coverage. Contact the CVPIA data managers if you wish to add a set of taxa to the CVPIA common taxa list. All blue columns are conditionally required and only need to be filled out if you do not select a **CVPIA_common_species**.

Example:

- **kingdom_value** - Animalia
- **phylum_value** - Chordata
- **class_value** - Mammalia
- **order_value** - Carnivora
- **family_value** - Felidae
- **genus_value** - Panthera
- **species_value** - Panthera Leo
- **common_name** - Lion
- **taxon_id** - 183803

Attribute

For each column in your dataset, you need to describe in detail the type of information encoded in that column. In EML, columns are “attributes”. The type of information needed for an attribute varies based on its measurement scale. This document explains the different types of measurement scales, how to select the appropriate one given the data within a column, and which columns within the ‘attribute’ tab of the metadata excel workbook are required given the measurement scale.

The instructions included below detail how to complete all the columns. Columns that are gray are optional and columns that are blue are conditionally required depending on the type of measurement scale:

- **attribute_name** - Name of the attribute as it appears in the dataset
- **attribute_definition** - A precise and complete definition of the attribute being documented
- **measurement_scale** - The type of scale from which values are drawn for the attribute (allowed values and additional details see [Measurement Scales](#) section)
- **attribute_label (optional)** - A descriptive label that can be used to display the name of an attribute.
- **domain** - Please select a domain from the dropdown. Domain options vary depending on measurement_scale. See the measurement_scale sections below to see which domains are appropriate for each measurement_scale.
- **type (conditionally required)** - Required for measurement_scale: *Ratio*, *Interval*. See the measurement_scale sections on *Ratio* and *Interval* for additional information.
- **units (conditionally required)** - Required for measurement_scale: *Ratio*, *Interval*. See the measurement_scale sections on *Ratio* and *Interval* for additional information.
- **number_type (conditionally required)** - Required for measurement_scale: *Ratio*, *Interval*. See the measurement_scale sections on *Ratio* and *Interval* for additional information.
- **unit_precision (optional)** - This is a decimal value indicating how precise the measured value is.
- **date_time_format (conditionally required)** - Required for measurement_scales: *Date Time* See the measurement_scale section on *Date Time* for additional information.
- **date_time_precision (optional)** - Optional for measurement_scales: *Date Time* See the measurement_scale section on *Date Time* for additional information.
- **minimum (conditionally required)** - Required for measurement_scale: *Ratio*, *Interval*, *Date Time*. Appropriate minimum values vary depending on measurement_scale. See the measurement_scale sections below to see what values are appropriate for each measurement_scale.

- **maximum (conditionally required)** - Required for measurement_scale: *Ratio, Interval, Date Time*. Appropriate maximum values vary depending on measurement_scale. See the measurement_scale sections below to see what values are appropriate for each measurement_scale.

The ‘attributes’ tab is the most in the metadata standards weeds, we can provide technical assistance with completing this tab.

Measurement Scales

There are five types of measurement scales to choose from. Based on the selected measurement scale, additional columns must be filled out to produce a valid EML document. Below each measurement scale is defined and guidance on the additional required columns are provided.

Nominal Used to define categorical scale attributes. Nominal is used when numbers have only been assigned to a variable for the purpose of categorizing the variable. An example of a nominal scale is assigning the number 1 for male and 2 for female.

- **domain** - Please list either "text" or "enumerated". A nominal attribute with domain “enumerated” is one in which the variable types can be encoded with integers. For example 1 = “cloudy”, 2 = “sunny”. A “text” type usually denotes a name or an id for example a site name or number.
 - if “enumerated” you need to provide a definition for each code in the **code_definitions** (see Code Definitions section) sheet.

Ordinal Used to define ordered scale attributes. Ordinal is used when the categories have a logical or ordered relationship to each other. These types of scale allow one to distinguish the order of values, but not the magnitude of the difference between values. An example of an ordinal scale is a categorical survey where you rank a variable 1=good, 2=fair, 3=poor.

- **domain** - Please list either "text" or "enumerated". A nominal attribute with domain “enumerated” is one in which the variable types can be encoded with integers. For example 1 = “cloudy”, 2 = “sunny”. A “text” type usually denotes a name or an id for example a site name or number.
 - if “enumerated” you need to provide a definition for each code in the **code_definitions** (see Code Definitions section) sheet.

Interval Used to define interval scale attributes. Intervals define data which consist of equidistant points on a scale. For example temperature data, mark grading, IQ scale, etc. Intervals can be negative while ratios cannot

- **domain** - The domain for an interval value should be numeric. Please select numeric from the dropdown menu.
- **type** - The type for an interval value should be interval. Please select interval from the dropdown menu.
- **units** - The units assigned to this attribute’s values. These must be in standard units, other common non-SI units are also allowed. A full list of allowable units can be viewed [here](#).
- **unit_precision (optional)**- This is a decimal value indicating how precise the measured value is.
- **number_type** - one of natural (1, 2, 3, ...), whole (0, 1, 2, 3,...), integer (-2, -1, 0, 1, 2) or real.
- **minimum** - Theoretical or allowable minimum value. Values can be larger than or equal to this number.
- **maximum** - Theoretical or allowable maximum value. Values can be less than or equal to this number.

Ratio Used to define ratio scale attributes. Ratios define data which consists not only of equidistant points but also has a meaningful zero point, which allows the ratio to have meaning. For example measurement heights, flow rates, weight, length, etc.

- **domain** - The domain for a ratio value should be numeric. Please select numeric from the dropdown menu.
- **type** - The type for a ratio value should be ratio. Please select ratio from the dropdown menu.
- **units** - The units assigned to this attribute's values. These must be in standard units, other common non-SI units are also allowed. A full list of allowable units can be viewed [here](#).
- **unit_precision (optional)**- This is a decimal value indicating how precise the measured value is.
- **number_type** - one of natural (1, 2, 3, ...), whole (0, 1, 2, 3, ...), integer (-2, -1, 0, 1, 2) or real.
- **minimum** - Theoretical or allowable minimum value. Values can be larger than or equal to this number.
- **maximum** - Theoretical or allowable maximum value. Values can be less than or equal to this number.

Datetime Used to define date and time attributes. DateTime is used when the values fall on the Gregorian calendar system. DateTime values are special because they have properties of interval values (most of the time it is legitimate to treat them as interval values by converting them to a duration from a fixed point) but they sometimes only behave as ordinals (because the calendar is not predetermined, for some dateTime values one can only find out the order of the points and not the magnitude of the duration between those points). The most encompassing format is: YYYY-MM-DDThh:mm:ss.

- **domain** - The domain for a datetime value should be dateTime. Please select dateTime from the dropdown menu.
- **date_time_format** - The format your date/time attribute is recorded in. ISO 8601 standard should be used (YYYY-MM-DD) for date, (YYYY-MM-DDThh:mm:ss) for datetime and (hh:mm:ss) for time.
- **date_time_precision (optional)** - To what level time is being measured
- **minimum** - minimum allowable date/time
- **maximum** - maximum allowable date/time or “ongoing”

Code Definitions

The code definitions sheet is where you will define the “enumerated” variable types specified in the attributes sheet. This is simply where you will list out all the types for an enumerated variable and each of their definitions.

All columns of this section are blue because they are conditionally required and only needs to be filled out if there are attributes that are “enumerated”.

Example:

code	definition	attribute_name
1	Clear weather	weather
2	Cloudy weather	weather
3	Rain	weather