Metadata Standards

Cal-Adapt
Published: 2024-01-01
Last modified: 2026-07-10

For attribution, please cite this work as:

Cal-Adapt. (2026). *Metadata Standards*. Cal-Adapt.
https://analytics.cal-adapt.org/data-tools/data-documentation/metadata-standards.html

Why Use Standard Conventions?

Readability Modern conventions are designed in a way that metadata is readable by humans and concurrently organized in a way that can be read by computers.

Interoperability Data developed to support state energy-related climate research and planning should be easily accessed by research partners on California's climate assessments, national climate assessments which utilize similar conventions, and global climate modeling efforts (i.e. CMIP). The end goal is to allow a user to incorporate data from multiple data sources directly on the Analytics Engine, without the need for customization.

Public Availability In order to work within the wide range of data serving options and analytics tools on the Analytics Engine (interactive programming environments, direct downloads), input data must be in formats which can be incorporated into the Analytics Engine, and directly read in by end users who may work with the data in the cloud computing environment or download the data directly.

Documentation Datasets hosted on the Analytics Engine should be self-describing; each dataset should contain information about when the product was produced, which version of the dataset is being used, what data went into the data product, relevant references, and points of contact for questions about the data.

Best Practices for GIS Data Formats

All geospatial datasets should include relevant metadata adhering to the Content Standard for Digital Geospatial Metadata by the Federal Geographic Data Committee (FGDC). While the specification may appear complex to newcomers, the goal is to communicate basic information, empowering downstream consumers of the dataset.

At a bare minimum, the title, originator contact, and citation details must be provided. Additionally, the abstract and purpose sections should be completed with a concise narrative and intention behind the data. The entity and attribute information needs to provide descriptions of the variables and values present, such as coded classifications, entity types, and domains. Desktop GIS applications like ArcGIS and QGIS assist in authoring and automating metadata generation, avoiding the need to edit the XML source. A validation service is available from USGS.

Best Practices for NetCDF

NetCDF files should be generated so as to be easily transferred between producer and end user, and self-describing so that extensive documentation files are not needed. Metadata included with a NetCDF file should be easy to read by users but also formatted in a way that present day software can read.

Data or products hosted on the platform as a NetCDF file should follow standard best practices as established by Unidata, including using a standard convention for organizing and describing data products. CF Conventions are preferred because they include categories of nearly all relevant topics (i.e. atmospheric, hydrology, ocean/lake, surface conditions), but for datasets outside these topics alternative conventions will be considered. Adoption of CF conventions assures end users can use NetCDF products developed by CEC-supported research using only information contained within the NetCDF.

Minimum specifications

Each NetCDF should contain the following components at a minimum:

Convention List the standard convention used to organize the data. CF Convention is preferred.

Coordinate System Dimensions of time and space should be included as standalone variables, each self-described using a standard convention system, including a description of the units and a naming attribute (i.e. for latitude, units of “degrees_north” and a long_name of “grid latitude, positive northward”). Coordinate variables should not contain missing values.

Variable Naming Whenever possible, units should be named using industry standard names for interoperability. Naming variables (and their attributes) in conformation to the Udunits package will allow users of common open access software and packages to load NetCDF files directly, without extensive modification. This is critically important for data being uploaded to the Analytics Engine.

Variable Attributes At a minimum, variable attributes should contain units and a long name descriptor (e.g. descriptive enough to label plots).

Missing Data A fill value or `valid_range` attribute should be applied to any variable with missing data.

Time Coordinates The time variable should represent time elapsed since a reference date. Either Gregorian or Julian calendar dates could be used, but Gregorian is preferred as it is supported by more software platforms. Storing the data as an ISO 8601 object allows for easy reading and interpretation by users, as well as a robust interface for interaction with modern software. The Udunits package provides easy to use tools for encoding of dates.

Suggested practices

NetCDF Version NetCDF-4 is the preferred and currently supported version.

Packing of Large Datasets Large datasets may be condensed in size via packing, or a reduction in precision through use of a scale factor and offset value. Most modern NetCDF tools and software automatically unpack data during load commands.

Global Attributes Global attributes describe the NetCDF file and, more broadly, the research project and dataset itself, rather than a component of the NetCDF file. Following CF Convention standard best practices, we encourage the inclusion of the following fields as Global Attributes:

Table 1: Recommended Global Attribute fields for NetCDF files, following CF Convention standard best practices.

Field	Description
Title	What's in the file
Institution	Where it was produced
Source	How it was produced, e.g. model version, instrument type
References	Audit trail of processing operations
History	Pointers to publications or web documentation
Comment	Miscellaneous

Test for Compliance Before submission of the NetCDF product, consider using an online test for your data product.