

ESDS–RFC-021 Technical Working Group Final Report

The ESDS-RFC-021 Technical Working Group (TWG) has completed a review of the Climate and Forecast (CF) metadata conventions RFC with the following conclusion that:

The Standards Process Group recommends ESDS-RFC-021 (CF Metadata Conventions) for endorsement as a NASA Recommended Standard.

History

The TWG conducted a review of the CF Metadata Conventions RFC dated April 2010 from the perspective of implementation and operational suitability. The CF Metadata Conventions are developed to promote interoperability among data providers, data users, and data services by providing a clear and unambiguous standard for representing geolocations and times of earth-science data, physical quantities that the data represent, and other ancillary information useful in interpreting the data or comparing it with data from other sources.

A set of review questions was adapted from the HDF5 and NetCDF 3 classic reviews. These can be found on the SPG web site (<http://www.esdswg.org/spg/rfc/esds-rfc-021>).

There were a total of 12 reviews received from the community that included data providers and managers, scientific analysts and programmers, and research scientists. The reviews were characterized with an overall positive response to the CF metadata conventions. Specifically, the review responses suggest a widespread usage of the CF metadata conventions across multiple disciplines that are crucial to promoting interoperability. Majority of the reviewers agreed that the RFC was clear and concise and there are no internal inconsistencies in the specification.

Recommendation

The TWG bases its recommendation on an analysis of the following factors.

Strengths:

The CF metadata conventions are used to describe attributes of various satellite datasets and earth system model outputs. Due to the needs of the different user groups, a wide collection of variable attributes is available in the CF conventions. As stated eloquently by reviewers:

“One of the major strengths of CF conventions is that the attributes are both machine and human readable form”

“CF is very practical, function(al), and to-the-point”

Because of this, interoperability is easily achieved when CF metadata conventions are used. Many open software and commercial data visualization clients are available to make use of the CF attributes from both netCDF and HDF data files. In addition, CF conventions are widely used by the Intergovernmental Panel on Climate Change Assessment Report numerical model outputs across various disciplines. In the words of one reviewer:

“The strength of the CF convention is its consensus approach bridging across several earth system modeling communities. Data sets which adhere to the CF standard and make wide use of its optional attributes contain rather well structured descriptions of the data set content which enable flexible distribution of data sets and also ease manual sharing of data, because the description is complete”

A few of the reviewers pointed out that CF allows reading of the axes, coordinate systems, data units, temporal attributes, maxima and minima and therefore more comprehensive than some of the other metadata conventions. Reviewers also commented on the fact that CF provides additional descriptive data that are useful to many users. According to one reviewer:

“Inclusion of these *local metadata* elements should make the product more self-descriptive, and thus enable users to more easily grasp the import of the product content”

While there were some complaints, almost all of the data providers agreed that the benefits outweigh the challenges leading them to use CF conventions.

Weaknesses:

There were two common weaknesses of the CF Metadata conventions mentioned by the reviewers.

- a. The CF conventions were originally developed to handle gridded model data; other types of data such as point and trajectory data needed modifications to the conventions.
- b. The CF conventions were originally tied to netCDF files and therefore implementing them with HDF files has been a challenge. In particular, there are not many examples and one group mentioned that it took them a fair amount of time to implement these conventions in HDF files.

A few reviewers also mentioned that while the standards names table is comprehensive, the current organization of standard names makes it time consuming to find the right name. Additionally, the units attribute in the CF conventions makes use of the udunits package, which is not described very well and lacks clarity according to two reviewers. Overall, most reviewers found the RFC to be complete, clear and concise for implementation of CF metadata conventions.

Applicability:

The CF metadata conventions are widely used across multiple disciplines and as suggested by the reviewers, it is **“crucial to promoting interoperability”**. Based on the reviews received, the CF conventions are used by many data centers with data volumes exceeding many petabytes of earth science data. In particular, one review from a NASA data center said the following:

“In general, this specification is useful for our needs, supplying excellent context for the conventions and the philosophy behind them. It is particularly useful for most Level 3 gridded and Level 2 swath data.”

It was also mentioned that the CF metadata provides valuable information about each array in the product and therefore CF compliant metadata may be more valuable to researchers who use the products directly. Additionally, software libraries and tools exist to support netCDF and HDF file creation and modification with the CF conventions. According to one reviewer:

“... many OPeNDAP visualization client tools, especially Java visualization client tools (IDV and Panoply) almost strictly follow CF conventions”

The CF conventions are particularly relevant to NASA Earth Science Data Sets. These conventions facilitate interoperability between NASA HDF data and available netCDF visualization tools. One reviewer summarizes this fact as follows:

“It greatly helped us to achieve the interoperability of NASA HDF data with existing netCDF tools”

Another mentioned:

“The CF Conventions are critical to ensure interoperability of data stored in the netCDF file format”

Limitations:

While the CF conventions are widely used with a great community support, the RFC reviewers identified important limitations. These are summarized as follows:

- a) These conventions work very well for two-dimensional data. In the third dimension, however, one is limited to vertical height/ depth/ pressure/ sigma/ density/ temperature or time. Representation of remote sensing data in the third dimension such as a band number is not currently an option, which limits the CF metadata conventions. In addition, one reviewer mentioned that vertical coordinate systems such as those used by remote sensing data sets from vertical sounders are problematic.

- b) The standards names table has names across various disciplines, but, it is not a complete set and it probably will never be. While this may be an obvious limitation, the user/ data provider groups can request that additional more appropriate names be added to the table to the CF conventions committee. Some reviewers pointed out that linking the CF standard names to existing ontologies could be beneficial.
- c) The HDF implementation of CF metadata conventions is time consuming and there is a lack of clarity in handling 2 dimensional latitude and longitude arrays.
- d) Geo-location of data is subject to the following limitations identified by the reviewers: 1. Since there is no way to store projection information using CF conventions, all the latitude and longitude values for all points will have to be stored, greatly increasing the size of the data files, 2. Datum information is not defined in the CF conventions, 3. Providing latitude/ longitude pairs does not indicate whether these define the center of the pixel or one of its corners, and 4. Does not allow for missing values inside the latitude and longitude arrays.
- e) The CF conventions provide descriptive metadata best suited for data in a collection, but not the collection itself, and therefore may not be sufficient for data discovery and cataloguing.
- f) The CF conventions document depends on udunits software package to provide canonical units. However, this is poorly explained and leads to confusion when dealing with some measurement units. In addition, some of the units are not available in the udunits package.
- g) As mentioned earlier, the standard names table could use better organization perhaps in a hierarchical structure with semantic relationships based on an ontological approach.
- h) Too many attributes are optional and too few data sets contain the attributes that are needed to fully understand the meaning and structure of the data. If major data providers and operational centers would get together and define an additional set of rules based on the CF convention (so that various optional arguments become mandatory in their context) one should have all that is needed.

Despite this long list of limitations, all of the reviewers overwhelmingly recommend that CF should be a NASA ESDSWG standard because the benefits of using these conventions far outweigh limitations. In addition, the CF metadata conventions have a large user base with very good support and tools both inside and outside of NASA.

RFC Overview

The RFC can be downloaded from: <http://www.esdswg.org/spg/rfc/esds-rfc-021/ESDS-RFC-021-v0.01.pdf>

Evidence of Implementation document: [http://www.esdswg.org/spg/rfc/esds-rfc-021/EOI CF.pdf](http://www.esdswg.org/spg/rfc/esds-rfc-021/EOI_CF.pdf)

The CF RFC review responses can also be found in this folder: <http://www.esdswg.org/spg/rfc/esds-rfc-021/implementation-review-comments>