Review of Climate and Forecast Metadata Conventions Implementation and Operational Suitability

NASA's Earth Science Data Systems Standards Process Group (SPG) is considering the Climate and Forecast (CF) Metadata Conventions, for adoption as a community standard. You are invited to review this Requests For Comment (RFC) in the context of your implementation experience with this specification and its suitability for operational use. You only need to answer questions that are applicable to you. Please send completed review to:

spg-rfc-021@lists.nasa.gov.

Implementation Experience questions:

1. (Your background) Describe in a sentence or two your overall implementation experience related to the proposed specification. (e.g., specification implementer, tools developer, data provider, scientific analyst, science user, etc.) Have you directly implemented the CF metadata conventions? Did you use pre-existing software, and if so, what did you use?

   We as a team implemented a handler to access HDF5 data (e.g. NASA HDF-EOS5 data) via DAP and the enhanced module of HDF4-OPeNDAP handler to access NASA HDF4 and HDF-EOS2 data via DAP. One person is the project manager and the software developer. Another one is the software developer. We directly implemented the handlers by following CF metadata conventions if possible when mapping the HDF-EOS5 and NASA HDF-EOS2/HDF4 data to DAP2.
   We didn't use any pre-existing software to implement CF conventions. But we used some CF-compliant OPeNDAP clients such as IDV, Panoply and GrADS to verify our output.

2. (Completeness) Does the specification (the online documents referenced) provide all the detail you need to implement it in software? (e.g., to read or write a data file; to implement or modify a profile or extension; or develop a tool such as a metadata translator) If not, describe what is missing in the specification.

The specification didn't provide all the details we need to implement in our software. However, we think the reason is that the specification doesn't aim to serve remote sensing satellite data. The following are our comments.

1) The example in section 5.2 of CF 1.4 document is short. It will be good if the document provides more descriptions and even an additional example.
2) In the chapter 4 of CF 1.4 - Coordinate Types, no clear definition for coordinates other than vertical height/depth/pressure/sigma/density/temperature or time can be found. For satellite remote sensing data, the coordinate can be band number etc. There is no description of such kind of coordinate variable in the current CF conventions. Although this doesn't affect us to support the DAP output to be accessed by OPeNDAP clients in our current implementation, it will be good that such information is provided in the specification.

3) Some HDF4 and HDF5 data provide 3-D array latitude and longitude such as

```
Float lat[yc = 180][xc = 360][zc =12]
```

The current CF conventions don't address this.

4) Appendix F: Grid Mapping
Sinusoidal projection and Space Oblique Mercator projection are not listed. These two projections are used by NASA HDF products.

3. (Accuracy) Do any parts of the specification contain inaccuracies, or internal inconsistencies? If so, please provide details.
Based on our current experience, so far we haven't found inaccuracies or internal inconsistencies in these documents.

4. (Clarity) Is any part of the specification ambiguous, or poorly explained? If so, please provide details.

Most parts we encounter are clear. Here are two cases we find more explanations will be helpful for developers to avoid confusion and to understand the exact convention to follow.
Case 1: The 2-D coordinate variable (section 5.2 of CF 1.4) should be explained in more details for the following scenario.
Supposed that we have the two-dimension latitude and longitude fields:

```
Float lat[yc=180][xc=1]
Float lon[yc=360][xc=1]
```

Although the number of the element for the second dimension is 1, they are still 2-D arrays. Some OPeNDAP clients (IDV,MATLAB) will treat them as 1-D arrays. We hope that these latitude and longitude fields are still regarded as 2-D arrays. If not, at least CF conventions should provide the explicit explanation for this case.

Case 2: The backward compatibility with COARD conventions may need to be clarified.
For example, OPeNDAP clients(Panoply and IDV) fail to plot a file that has the following latitude, longitude and level coordinate variables.

```
Float lat[lat=180][lon=360]
Float lon[lat=180][lon=360]
Int   level[level = 10]
```

The data variable is

```
Float temp[lat = 180][lon=360][level = 10]
```

The coordinate variable level is a 1-D array that follows COARD conventions but

coordinate variables lat and lon are 2-D arrays that follow CF conventions (section 5.2). In this case, IDV and Panoply still expect lat and lon to follow COARD conventions, which only allow 1-D lat and lon coordinate variables and require the dimension name to be the same as the coordinate variable name.

5. (Balance) Does the standard describe the right set of concepts and attributes and enable the appropriate operations for its intended users? In particular, have the guiding principles outlined in section 5.2 been followed in the development of standard names?

   We would say yes for the first question. I think our case is very special in term of following CF conventions. The intended users for the current CF conventions are modeling communities. So we believe, though we are not experts on modeling data, that the CF conventions indeed describe the right set of concepts and attributes for modeling communities.
   Section 5.2 is the most useful section (though it is a little short) for our implementation. We are glad that several important OPeNDAP clients closely follow this section and HDF swath and 2-D projection grid can be visualized properly by these clients to the best of our knowledge.

6. (Usefulness) How well does this specification meet your information sharing needs? (e.g., Does it properly represent and describe your datasets? What are the pros and cons of these metadata convention attributes?)

   As we discussed in the previous questions, the specification largely meets our need, especially section 5.2. It will be good if CF conventions provide conventions for other third dimension variable (band number) used by the satellite remote sensing community. To the best of our knowledge, the standard attribute names are meaningful and complete in terms of modeling communities. Although we are not certain whether these attributes address all types of satellite remote sensing data, we feel that it covers most usage of NASA HDF data such as Fillvalue, units, long_name etc.

7. (Implementation) What implementation challenges does the proposed standard present? (e.g., does it conflict with other metadata requirements for your data? Is it compatible with the data formats you use?)

8. We have to translate some attribute values to follow CF conventions. For example, the attribute “units” of latitude field for some NASA HDF data is “degrees” rather than “degrees_east” required by CF conventions. We have to change the “degrees” to “degrees_east”. In terms of data formats, we believe that whether following CF conventions or not is irrelevant for most data formats, at least for HDF4, HDF-EOS5 and HDF5.

9. (Flexibility) In what software environment(s) have you used the CF metadata conventions (e.g., Solaris, Linux, Windows, Mac OS X)?
   Linux and Mac OS X.

10. (Standard Names) In your opinion, does the standard name table provide an adequately comprehensive set of names for the metadata representation?

    Again, the standard name table may be good enough to serve the need for modeling communities but may not be complete enough for satellite remote sensing communities. For example, the current standard name sheet may not address the metadata information stored in ArchiveMetadata and CoreMetaData attributes or datasets in HDF-EOS2/5 and HDF4. Another small issue is the standard name table listed at the NASA SPG website is in XML and is difficult for us to read through.

Operational Suitability questions:

11. Do you currently use or plan to use CF conventions in a production setting? What types of applications do you use with CF Conventions? Does the metadata model work well with the data types and data manipulations in your application?

    We as a team are just responsible for implementing these OPeNDAP HDF handlers. To our best knowledge, NASA GES DISC is using these handlers now and they are satisfied with the capability that CF-compliant OPeNDAP client tools can access most of their data.
12. Why do you choose to use the CF metadata conventions for your applications?

    Because many OPeNDAP visualization client tools, especially Java visualization client tools(IDV and Panoply) almost strictly follow CF conventions. To help those end-users that like to use these Java tools, the only solution is for us to implement the handlers by following CF conventions.
13. Have you or your users encountered any difficulty when using some of the data access or visualization tools (e.g., IDL, GrADS, etc.) on files with CF metadata? If you have, please provide a brief description of your experience.
14. So far it seems that if files following CF conventions, there are not much difficulty. The only problem is what we addressed in question 5. The clarification about backward compatibility with COARD in the CF document is necessary for individual tools to follow.

15. Does the CF metadata conventions meet your requirements for discovering, accessing, providing interoperability of data and metadata? (e.g., Can it handle the data types in your applications? Do you provide catalog services that utilize CF conventions?)

    It greatly helped us to achieve the interoperability of NASA HDF data with existing netCDF tools.
16. What operational challenges or limitations do the CF metadata conventions present? (e.g., Does it take a long time to learn how to use it? Does it require advanced processing power, large amounts of memory, complex configuration, etc.)
17. It took one of us fairly amount of time to learn and apply to our software packages. We think that it is because the developer didn't have much background on Earth Sciences.

18. What benefits do CF conventions present? Do the benefits of CF conventions outweigh the challenges? (e.g., Do the conventions offer the flexibility you want to package the data types in your applications? Do they facilitate interdisciplinary studies?)

19. Yes, based on our previous answers, it is obvious that we think CF conventions help us to help our users.

20. How much data do/ will you provide using these CF metadata specifications? (number of distinct data products or data sets, total data volume, number of files.)

21. Again, we are neither data providers nor data distributors, so we cannot answer this question. But we recommend that if possible, it is better to follow CF conventions(providing CF attributes) in new products regardless of data storage formats.

22. How many users and user-groups do you have or expect to have for data using CF metadata conventions, and what is your expected user community?

Again, cannot answer this question.

23. (User comments) Any additional comments, observations or criticisms of CF metadata conventions and the RFC can be provided here.
    In addition to what we described in the previous questions, one thing we feel CF conventions should address is the missing values inside latitude and longitude. These are so common for Level 1 and some Level 2 swath data. It will be great that CF conventions allow the missing values inside lat and lon. The tools then can follow the CF conventions to support the visualization of the file containing lat and lon having missing values.

    We are not certain how well HDF-EOS point data is supported by CF.