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?

group experience:

- contributed to discussion on CF standard names for chemistry as co-coordinator of TFHTAP multi-model experiments
- production of numerical model results for scientific analysis of chemistry climate interactions; contributions to IPCC assessment
- development of WCS server for TFHTAP multi-model data and linkage to other model products and observational data sets
- programmed several analysis tools for numerical model results and model evaluation
- member of MACC management board where operational composition forecasts and analyses are made available

CF implementation experience:

- implementation of a CF checker for uploading data onto Jülich TFHTAP server (python)
- contributed to definition of tables for CMOR program
- used bits and pieces of CF convention in NCL or IDL analysis tools

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.

It takes considerable time to digest the information in the CF documentation and there are many cross-dependencies so that one must develop full understanding before being able to implement the standard. If one reads often and careful enough, the documentation appears complete. Specific section which would need updating is the guidelines for

standard_name definition. Here, the tables showing the possible entries for each "field" are outdated and standard_names exist for more chemical species and more processes than listed in the guidelines document.

3. *(Accuracy)* Do any parts of the specification contain inaccuracies, or internal inconsistencies? If so, please provide details.

We have found no inconsistencies, but there are a few places where the standard is rather loose. Specifically, the description of the cell_method attribute leaves room for interpretation. The time formats are not always ISO-conformant (no mandatory use of 4 digits for year).

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

Generally, a problem with CF is the optional character of many attributes. This makes it hard to develop tools which should rely on certain properties in the data sets. A specific example are climatological time variables. There are various indicators which could help the tool to decide whether a time variable is meant to be a "climatological average", but no specific attribute value that can be parsed reliably.

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?

Yes

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?)*

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. The main problem is that in reality too few data sets adhere to the CF convention or – even if they formally do – they contain only a small subset of the attributes which are needed to fully understand the meaning and structure of the data.
A specific example are the optional "axis" attributes: these are extremely helpful when data sets shall be delivered through a WCS server – however they are optional, so the server cannot automatically rely on them.
One fundamental problem we encountered in the context of hosting and sharing the TFHTAP multi-model data is that the definition of new standard names takes a lot of time (due to extensive discussions on the CF mailing list prior to acceptance) and not all names that are needed can be defined as "standard" names. For example, we have specific tracers with fixed lifetimes defined in some model experiments which are of

interest only to our community. It would contradict the concept of "standard" names to adopt these, but on the other hand a standard_name attribute would be very valuable in order to facilitate delivery of such data through our WCS server. The CF convention states that "either long_name or standard_name attribute should be present", but the long_name attribute is a poor replacement for a non-standard standard_name which at least follows the guidelines for construction of standard names.

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?)*

As we adopted the CF convention for all our netcdf data files, we don't see conflicts with other conventions. The model raw output is generally not standard-conformant and has to be converted e.g. using the CMOR tool.
The major challenge for the implementation is the handling of optional attributes. If you have control over the format of the data sets, this can be overcome. If your tools shall also handle external data, then this is very hard to accomplish.

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

Linux

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

In general: yes. As stated above there are occasions where one would wish an extension. Getting a proposal for new standard_names through the discussion list takes time and there is a danger that individual groups or communities seek their own private extensions, which of course defeats the purpose. It is recognized that there must be some compromise here, but the rules for this have not been spelled out very clearly.

Operational Suitability questions:

10. 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?

CF-conformant files are made available through a WCS server. We are currently cleaning up attributes in existing files and re-design the data access interface. We also plan to distribute daily data files from a European project using these conventions.

11. Why do you choose to use the CF metadata conventions for your applications?

In general the CF conventions provide a very good description of the data file content and are structured well enough to be interpreted by software tools and to enable interoperability. They have become a widely used standard in the climate modeling community.

12. 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.

Normally, adherence to the CF convention does not cause problems when data shall be used in existing software tools (with the possible exception of the _bnds variables). However, we have not come across any major tools yet which try to make full use of the CF attributes. We believe that this is a consequence of the fact that too many of these attributes are optional and too few existing data sets make use of them. We also often found that data sets carry the Conventions="CF-1.0" label but don't pass a CF checker test.

13. 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?)*

If a reasonably large number of CF attributes is used consistently and the problem of undefined standard_names can be solved, then yes. As said above, we provide WCS catalogue services.

14. 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.)*

"Using" CF as data user is relatively simple and the convention can be of great help for interpreting the data files which the user receives. Using CF to provide data is a lot more challenging. What would be helpful is a set of fully CF conformant data files that can serve as examples. The information how to create a "good" CF conformant file is rather scattered across the documentation and it requires a certain dedication to the cause to dive deep into the documentation and understand what is meant. Using CF as a tool developer is even more challenging, because the standard leaves too much room for optional arguments and you will have to define additional rules in order to make sure that you can interpret the data files you receive. We believe that a more stringent standard would make it a lot easier to develop flexible data analysis and interoperability tools.

15. 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?*)

The benefits still outweigh the challenges. The convention is rather versatile and broad and it has a large group of "believers", even if not all of them live strictly by the rules. 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. Perhaps this can be achieved as CF-2.0 ?

16. 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.)*

We currently serve about 8 different types of data files from the TFHTAP multi-model intercomparison experiments. Total number of files ~20,000, total data volume ~2.5 TByte.

17. 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?

The TFHTAP group consists of about 80 scientists worldwide who make use of the model data from the intercomparison exercise. We expect to attacrt additional users with the further development of our WCS server capabilities.

18. *(User comments)* Any additional comments, observations or criticisms of CF metadata conventions and the RFC can be provided here.