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?
   Our Program is a provider of atmospheric observational data in NetCDF format and its operations predate the CF conventions. We have not implemented the conventions. We are considering whether and to what extent we should move towards CF compliance for some of our data.

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.
   We have not fully determined whether all needed detail is provided. The specification is quite detailed, and probably contains most or all needed detail.

3. *(Accuracy)* Do any parts of the specification contain inaccuracies, or internal inconsistencies? If so, please provide details. None has been identified.

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

   Most is well explained. One part was unclear: the specification document, p. 10, says: "Unless it is dimensionless, a variable with a standard_name attribute must have units which are physically equivalent (not necessarily identical) to the canonical units..." A specific example in our case related to whether our temperature variables in degC would need to be converted to the CF canonical unit K. After several Web searches we think we determined the answer is no (see http://www.mail-archive.com/cf-metadata@cgd.ucar.edu/msg00454.html). Follow-ups in this thread suggested that the specification wording would benefit from clarification of this concept of "physically equivalent" units; we support providing such clarification. In addition, as suggested by another thread, the Udunits table should be Web viewable without having to download and install the

full package. Our conclusion about not needing units conversion is based in part on the earlier version of the Udunits material, which is available on the Web if you search for it. Easy accessibility of the standard and its components, to both the data provider and the data user communities, is an important precursor to acceptance and success.

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, it appears the principles have been followed well.

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?)*
   We will need to look harder at some aspects of the specification (flags and specification of missing values as fill values, for example). It seems likely that the specification can work for us, as far as it goes. We feel that additional metadata beyond what the standard appears to cover is also important in locating and understanding our data (such as instrumentation specification and details). See further comments under item (10) below.

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?)*
   The main challenge for us is the massive and diverse collection of our data, and the concomitant effort and resources needed to convert some of them into the standard while maintaining compatibility with our users' existing tools. A specific challenge is the difference in the way we represent time in netCDF files, compared with the CF specification.

8. *(Flexibility)* In what software environment(s) have you used the CF metadata conventions (e.g., Solaris, Linux, Windows, Mac OS X)?
   Our infrastructure would be using these primarily in Solaris and Linux environments. However, a large portion of the user community works with our data in Windows, so Windows support is also needed.

9. *(Standard Names)* In your opinion, does the standard name table provide an adequately comprehensive set of names for the metadata representation?
   No, it is clear that some, perhaps many additional names would need to be requested/approved for us to provide all our data. Many of the names are already defined, however.

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?

We do not currently use the conventions. Applying it to our millions of files with diverse content would require considerable resources. Applying the conventions to a subset of the data would be more feasible. The metadata model permits additional attributes we could use to provide additional metadata, which we consider vital to adequately document observational data. We feel the CF conventions should be extended to assist in providing structured metadata. For example, GCMD and FGDC use category decisions leading to useful hierarchical representations of metadata. Instrument categories leading to instruments, and method categories leading to methods, are examples of such structured metadata.

11. Why do you choose to use the CF metadata conventions for your applications?
    Our data are intended for use by (and are used by) climate modelers. CF conventions are currently used for much climate model output. Converting some data to become CF compliant would facilitate further use of the data by 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.

    Our Archive has no such experience. Others in our group might have such experience.

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?)*
    The most promising approach to retrofitting our data for CF compliance appears to involve wrappers. We would need to explore the question of how such an approach might affect our discovery and access process.

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.)* Learning how to retrofit the conventions to the extensive data (or to parts of it) will require time, as will implementing changes in processing environment. Perhaps a proof-of-concept exercise would let us better assess the challenges, and whether there are substantial limitations. For example, we hope the wrapper approach to CF compliance (e.g., NcML) could handle, or could be extended to handle, the different way in which we represent time.

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 obvious benefit of CF conventions is easier interoperability with climate model output and other parts of the climate modeling community. Also, we would like significant parts of its data to be available through the Earth System Grid infrastructure, which strongly encourages CF compliance.

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.)* This is still to be determined. Data volume could range from tens of GB to tens of terabytes (perhaps more) for a few to dozens of data sets. Thousands to millions of files could be involved. It all depends on the scale of effort devoted to compliance. The most important data products could be provided based on values towards the low end of these ranges.

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? Dozens to hundreds of users in several or more user groups could be expected. Many atmospheric scientists are already using data in their current form. We expect that the climate modeling user community would make more use of our data if it were CF compliant.

18. *(User comments)* Any additional comments, observations or criticisms of CF metadata conventions and the RFC can be provided here. The document ESDS-RFC-021v0.02 (Russ Rew, April 2010) is concise and very informative. We did not find a FAQ for CF conventions; this could have saved us time with some of our questions.