

Plenary session summaries

Summary of the CF Standard Names session

The session began with a presentation giving an overview of CF standard names: the purpose of a standard name and how names are constructed; the process for proposing new names (or modifications to existing ones); published versions of the standard name table; likely developments in standard names in the coming year. The slides provided links to all the online resources mentioned in the presentation, such as published versions and the formal rules for discussion and acceptance of new names. View the presentation [here](#).

Following the presentation, a number of questions and comments from workshop participants were discussed:

- A dimensionless (scalar) quantity should have units of “1”. If units are omitted they will be assumed to be “1”. A string valued quantity which would have a standard name such as “region” or “area_type” does not have canonical units and should not have a units attribute inside a data file.
- Standard names descriptions are published in the standard name table - they do not need to appear in the data file.
- There are differing opinions on what terms constitute “jargon” in standard names, e.g. using “tendency” to mean “time_derivative”. It was pointed out that standard names are proposed and agreed by community discussion - we try to avoid using terms that are meaningful only to one scientific community.
- The use of the cell_methods attribute to describe statistical post-processing, e.g. mean, max, versus operations such as “integral” (included in the standard name) was discussed. There was a recognition that the distinction between which operations should appear in the standard name and which in cell methods can be seen as “somewhat arbitrary” but that from a practical point of view the current system has worked well.
- Numerical coordinate information is not included in standard names, but we do use named surfaces such as “surface” (meaning lower boundary of atmosphere) and “tropopause”.
- The question was raised as to whether instrument names should appear in the standard name. The purpose of a standard name is to help data users recognise where data from

different instruments are comparable, so they should use the same standard name. Instrument metadata can be included using other methods. See the following examples:

- [NCEI netcdf templates](#)
 - [ACDD 1.3 instruments](#)
 - [NVS L22](#)
 - There was some discussion as to whether requiring a GitHub account is the best way for new users to propose standard names. There is a recognition that it does involve something of a learning curve, but does not require advanced use of GitHub features such as submitting changes to repositories.
 - There was some discussion about linked data and how standard names can refer to other (non-CF) vocabularies. The possibility of a CF namespace is currently under discussion: <https://github.com/cf-convention/discuss/issues/51>. References to external vocabularies have so far been done for the specific cases of land use and biological taxa. Discussion of how this approach could be applied to other vocabularies continued in the Standard Names breakout session.
 - There was discussion of how to ensure data providers supply CF compliant files for archiving. One approach is for data centres to take on the task of fixing metadata problems after data have been submitted. Another is for data centres to refuse to accept data unless it has been checked as CF compliant prior to submission. Data centre staff are encouraged to submit proposals for new standard names if this will help their user community to supply appropriate metadata.
-

Summary of the Governance session

In this session, an overview of the CF community, CF governance, and the process for changing the CF Conventions document was presented. It described CF as a Community Effort where anyone interested in CF is part of the community and everyone is encouraged to contribute by participating in discussions and proposing changes. CF uses a consensus building method to reach agreement on changes with two governance bodies responsible for development and stewardship of CF:

- The CF Conventions and Standard Names Committee(s) are made up of active community members. They are (self-)nominated by the community and appointed by the CF Governance Panel for 5-year renewable terms. The committee's role is to moderate proposals and contribute to discussion. They are responsible for development of CF.
- The CF Governance Panel consists of representatives of organizations making significant contributions to CF. The Panel promotes the use of CF and is responsible for overall stewardship of CF.

(Find the “CF is a Community Effort” presentation [here](#).)

After the presentation, some needed updates to the CF website were brought up, mainly updates needed due to the move to GitHub from Trac. The idea of a "documentation hackathon" was mentioned. From there, the discussion focused mainly on Copyright, Licensing, and whether a Contributor License Agreement (CLA) would be needed. How this might work was unclear as CF is not a legal entity. It was suggested this discussion should be moved to a GitHub Issue.

(Full discussion/notes are [here](#).)

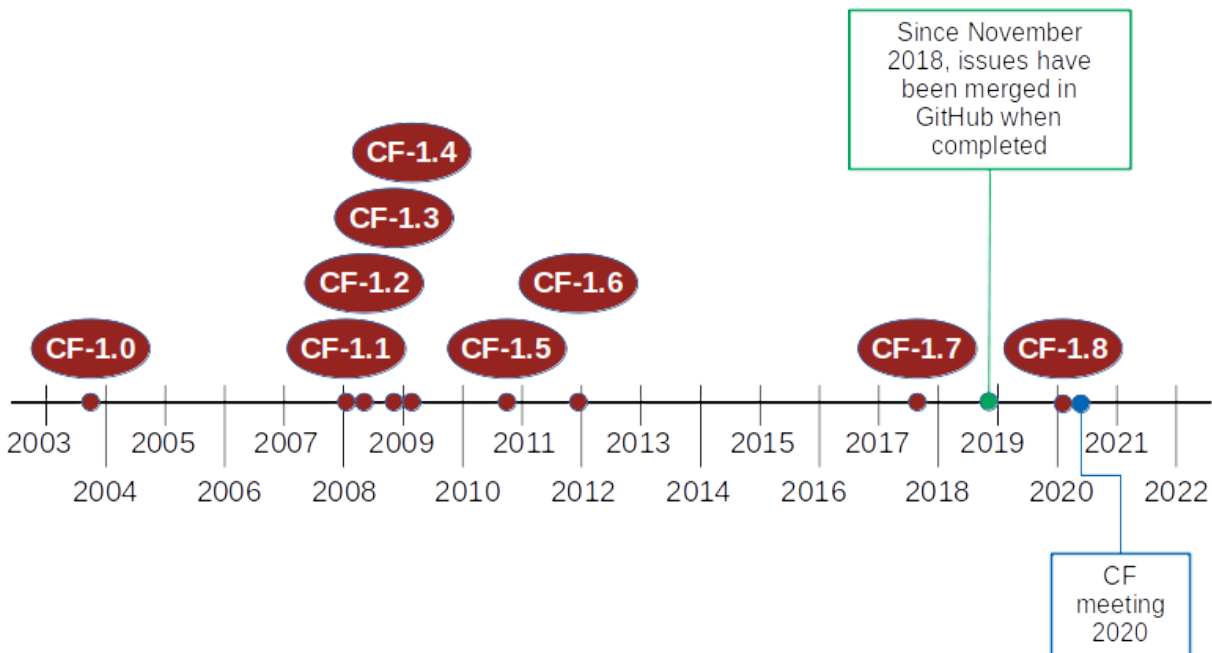
Summary of the Process Demo and Q&A session

This session presented the process of introducing changes to the CF Conventions and the implementation of these processes in GitHub. Potential improvements to these processes were also proposed and discussed. The slides are found [here](#) and the full discussion notes are [here](#).

The presentation and discussion both showed that the migration to GitHub had been largely successful and that the community was able to work in the new environment.

Potential improvements that were discussed included using Kanban to visualise the status of a proposal, as well as proposers keeping the issue summary up to date while posting when they have made these updates so that subscribers to the issue receive emails informing them of the change.

Summary of the Road map: CF 1.8, CF 1.9, and Onward session



In this session a timeline of all CF releases was presented (see above), including some discussion on the reasons behind the non-linear clustering of releases. This was largely due to the availability, or otherwise, of resources for CF editorial work. The changes that contributed to the most recent release (CF-1.8) were described and the impacts of these changes on the CF data model were discussed. Changes already agreed for CF-1.9 were listed. The expected timeline for future releases was proposed, namely one per year (defined as being between annual CF meetings), with the possibility of more frequent releases if needs demanded it.

The CF data model required two changes for CF-1.8, both relating to [geometry cells](#):

- A generalisation of the notion of “cell bounds”
- In particular circumstances, cell bounds may exist without cell coordinates.

These changes have been accepted for CF-1.9, the asynchronous addition being a one-off artifact of the the delay in merging the CF data model into an appendix of the CF conventions, which will also appear in CF-1.9.

It was agreed that there is not currently a need for “CF2”, as the extended netCDF features (strings and groups) that it meant to implement are now incorporated into CF-1.8. The discussion touched on semantic versioning, and version 2.x will be reserved for major backwards incompatible changes, should they arise. The use of release candidate versions was proposed as a beneficial part of the release workflow.

Summary of the CF Software tools session

In this session, we had three presentations that focused on CF compliance checkers and the implementation of the CF data model. Our three presenters were Robert Fratantonio (RPS group), Rosalyn Hatcher (NCAS-CMS) and David Hassell (NCAS-CMS).

Bob was first up and provided an overview of the various compliance checkers and tools that the RPS group has built in support of the US Integrated Ocean Observing System (IOOS). This collection of fine tools includes web-based compliance checkers that test compliance with CF (up to v1.6), the Attributes Conventions for Dataset Documentation (ACDD), and various other IOOS-specific NetCDF data formats. More information for these highly recommended tools are available at:

<https://compliance.ioos.us/index.html>

<https://github.com/ioos/compliance-checker>

Next, Rosalyn led us through the recent development she has been working on in regards to the CF compliance checker which is available through CEDA and a mirror in Reading. This checker can test compliance with CF up to version 1.7. Many noted these were the original CF checkers that they were used to using.

[CEDA compliance checker](#)

[Reading compliance checker](#)

It was noted that both of these checkers are extremely valuable to the community. Suggestions from the session included:

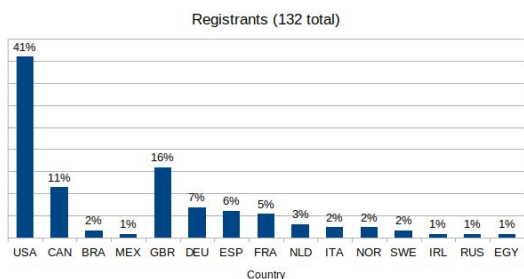
- As more compliance checkers are developed, perhaps there is a need to “certify” a checker as being verified
 - Results of various checkers could be listed to ensure certification status is transparent
- Perhaps the CF community could develop and maintain a set of test files with known errors to ensure all compliance checkers arrive at the same results
 - This could include a curated set of special cases that may be especially tricky
- There is a [CF Software page](#) that lists such software. Please update if things are missing.

The last presentation of this session was by David Hassell, who walked us through the CF data model python library, cfdm. This is a python package which has implemented the CF data model. David was able to demonstrate the extremely nice usability of this module interactively through an example Jupyter notebook.

[CFDM documentation](#)
[CFDM tutorial](#)
[CFDM source code](#)
[Jupyter Notebook example](#)

Summary of the Wrap-up and Conclusions session

In this session, a summary about the 2020 CF Workshop registration and participation was made. There were 132 registrants from 15 countries most of them from North America (71) and Europe (59). The participation in plenary sessions was between 65 and 90 people. These figures made evident the success of the decision to make the workshop fully remotely.



With respect to topic interest and participation the *Standard Name* sessions was the most popular one with 37 participants, followed by the *Metadata handling through processes* with 32 participants. The third one was the *Cell Methods* with 21 participants. The rest of the topics were followed by 10 or less people. A wrap-up about plenary sessions was presented, and in particular for the CF Governance and Committees, its process to propose changes and improvements to CF Conventions and Standard Names. For details on plenary and topic sessions see the rest of this document.

Discussion topic summaries

Standard names: general discussion on any aspect (Alison Pamment)

Introduction

- Publish CF Standard Names as an Ontology (Alison Pamment)
<https://github.com/cf-convention/discuss/issues/51>
- International effort to create a data standard for chemical oceanographic data?
- The latest standard name table:
<http://cfconventions.org/Data/cf-standard-names/72/build/cf-standard-name-table.html>
- The CEDA standard name editor: <http://cfeditor.ceda.ac.uk/proposals/1>

- A useful reference for background reading is Jonathan Gregory's presentation at the 2018 CF meeting: <http://www.met.rdg.ac.uk/~jonathan/talks/CF180620.pdf>

Summary of discussion

1. Q: How do we get people from other scientific domains involved in CF (given that it has expanded beyond the original Climate-Forecast community)?

A: We should try to have a CF presence at more scientific conferences e.g. AGU, EGU, ocean conferences. It would be good to have some pre-prepared CF materials available for posters, presentations, training that we can all use as needed.

2. Q: Should we decide to include large sets of new standard names, e.g. to serve the ocean chemistry community, in the standard name table or should we point to external vocabularies as we are now doing for biological taxa?

A: If a 'suitable' external vocabulary already exists then we should use it. By 'suitable' we mean something that passes FAIRness tests. See e.g. the work being done under FAIRsFAIR [essential criteria\(WIP\)](#) and RDA Interest Group [VSSIG](#) (Vocabulary and Semantic Services Interest Group). Every term in external vocabularies should have an individual URI that resolves to something!

If there is no suitable external vocabulary then the CF community should do the work of agreeing the new standard names so that data producers can get on and write their data.

3. Q: We discussed the issues currently being raised in <https://github.com/cf-convention/discuss/issues/51> ("Publish CF Standard Names as an Ontology").

A: We think it is possible to have URIs for each standard name in CF "namespace", i.e. to include "CF" somewhere in the URI. This could be based on the terms served in NVS and would probably need redirection at DNS level. We will arrange a Zoom meeting to discuss the technical details and take this forward in the near future. Having a Turtle file on the CF website for use with ontology tools is a good idea, and would be even more useful if we can set up the redirection.

4. Q: Is it possible to view diffs between standard name versions in a simple way?

A: At the moment there isn't an easy way to do this on the CF website or in GitHub because of the way the files are organised in the repository. A Python tool has been developed (by Harry Singh?) for finding diffs between XML versions of the standard name table and this could be advertised on the CF website. Also we could tidy up the

way the standard names files are stored in the website repo so that it would be easier to see diffs using the features of GitHub.

- EUMETSAT has a tool for viewing vocabularies that is currently restricted to WMO formats. It's here: <https://vocabulary-manager.eumetsat.int/>. You can use it to view e.g. the differences between WMO BUFR & GRIB table versions, which are in the end pretty similar to CF Standard Names. - user driven tool - could request CF to be added

Cell methods: "within"|"over" "days"|"months" and time axis (Section 7.4)
(Lars Barring)

<https://github.com/cf-convention/cf-conventions/issues/197>

Introduction

There are two components:

- One is to clarify, and agree on a solution/interpretation of what the "climatology" attribute means and its relation to the cell methods "within" and "over": How are they connected? What does it mean for different typical cases? Can the "climatology" attribute be disconnected from the cell methods? Is in fact "climatology" necessary at all? What to do vs. existing CMIP6 data and what to do for CMIP7? What is the status quo, and what minimal changes may we want to make? Once these matters have been resolved this may come down to rather small changes to Section 7.4.
- The other one, is the more far-reaching need for a new or alternative mechanism that allows for a more flexible description of more complex and/or multi-step temporal processing of data.

Summary of discussion

Standard way to define subsampled coordinates (Daniel Lee)

<https://github.com/cf-convention/discuss/issues/37>

Introduction

The volume of data products is increasing exponentially with the resolution of models and sensing systems. Increased spatial granularity, in particular, has led coordinates to contribute significantly to the cost of encoding, storing, and transmitting data products. As a result, data transfer and storage has become prohibitively expensive in some fields, even when state of the art data compression methods are used.

Data reduction has proven successful in significantly reducing data volumes, and thus costs. The greatest potential is in the field of remote sensing, where observational data is often not on a regular grid and thus explicit coordinates are needed for every observation point. Subsampling coordinate data while providing sufficient information to allow the user to fully recover it can reduce data volumes by up to 40% - *after* off-the-shelf compression techniques have been used.

We propose a method of providing subsampled coordinates to users that would significantly reduce the required data volume while still allowing a full reconstruction of all coordinate data. Our approach is a synthesis of methods used throughout the Earth Observation community and generalised for use in applications beyond remote sensing. It is possible to apply it along an arbitrary number of dimensions and on data that is regularly spaced or warped, as is the case e.g. in satellite observations with increasingly oblique viewing angles across the product. The integration of this method into the CF Conventions is inspired by grid mappings and compression by gathering.

At this stage the proposed method is still being refined and any input we can gather from the Community is greatly appreciated.

Summary of discussion

The need for this proposal to progress was confirmed and reiterated. Most significantly, the group needs to decide where to position what information related to viewing geometries. What information should be associated directly with a data variable, as opposed to the description of the interpolation method required in order to reconstruct all coordinates? These are questions that will be pursued in the coming weeks as progress is made toward being able to make a formal proposal to extend CF.

"mesh variable" instead of "boundary variable" for contiguous grid cells
(Ryan Abernathey)

<https://github.com/cf-convention/discuss/issues/5>

Introduction

Most numerical atmospheric and ocean models use some form of staggered curvilinear grid discretization (e.g. Arakawa grids), in which variables are located at some specific point (e.g. cell center, cell face) with respect to a finite volume quadrilateral (grid cell). CF conventions currently recommend encoding the geometry of these cells via the “boundary variable” convention. Boundary variables are extremely general, accommodating arbitrary unstructured, non-contiguous cell geometry, requiring N boundary vertices to be specified for each grid point for N-dimensional geometry. For contiguous curvilinear coordinates, this representation is inefficient in terms of memory usage, due to the redundancy in the vertices of neighboring contiguous cells. For very large grids, this inefficient representation can become a bottleneck for analysis and visualization.

We propose to extend CF conventions with a new mechanism to represent such cell geometries which lies closer to how models represent their grids internally. We wish to incorporate the concept of a “mesh variable,” which describes the boundaries of contiguous grid cells in terms of cell vertices. The [sgrid conventions](#) provide an excellent template for such representation, so we will propose to incorporate some or all of sgrid into CF proper. However, as discussed in the github issue, several thorny technical issues would benefit from community discussion.

Summary of discussion

The attendees of the "Mesh or boundary variable" breakout session reviewed a presentation submitted by Ryan Abernathy. There was general agreement that it would be good for CF to provide a way to describe cell coordinates in terms of boundaries alone rather than points and boundaries, both because it can allow grid coordinate data to be represented more compactly and because this is a more natural representation of the information in certain cases. As an example, a number of disciplines make exclusive use of grid vertices for their coordinates.

The attendees agreed that the SGRID convention provides a full-fledged system for representing coordinates via boundaries, and they agreed that SGRID and UGRID conventions should be brought more fully into the CF "family", but there was concern about the complexity of SGRID and its novel use of the 'cf_role' attribute. The attendees sketched out a simpler convention that is included in the breakout notes.

Embedding provenance information - Metadata handling through processes (David Huard)

<https://github.com/cf-convention/discuss/issues/33>

Introduction

When multiple netCDF files are aggregated to compute ensemble statistics, or when chains of algorithms are applied to netCDF files, recording operations and source metadata in the

"history" attribute can become very messy. Recording such provenance information is necessary in applications where traceability and reproducibility are critical.

There are standards (e.g. [PROV](#)) to encode such provenance or data lineage information in machine readable formats. netCDF files could include a "provenance" attribute in which a machine-readable representation of the provenance would be stored. An example of this can be found at <http://metaclip.org/>, where provenance information is embedded in figure metadata.

Summary of discussion

Keeping track of the history of a netCDF file can become messy when it involves multiple source datasets going through complex workflows. In a bid to complement existing human-readable CF *provenance* attributes (i.e. history, source, comment), the break-out group discussed the idea of including in the CF-Convention a variable or global attribute recording formal provenance information using a machine-readable data format. Using a provenance language allows data providers to include detailed information without sacrificing understandability, facilitates interoperability and automated parsing. While this is not meant as a tool to enable out-of-the-box reproducibility, it can help diagnose errors and improve traceability for high-stake applications. For instance, participants described two projects embedding provenance information in the figures or netCDF files they generate for the IPCC report. During the break-out session, a consensus quickly emerged for netCDF files to include this provenance information as a URL (e.g. a Digital Object Identifier) linking to a provenance document, rather than embedding the provenance document within the attribute. This would allow multiple netCDF files to link to the same provenance document. The convention would recommend using an existing standard to describe provenance information (e.g. PROV), as this would help focus user training efforts and software development.

Allow CRS WKT to represent the CRS without requiring comparison with grid mapping Parameters (Alan Snow)

<https://github.com/cf-convention/cf-conventions/issues/222>

Introduction

CF guide reference:

<http://cfconventions.org/cf-conventions/cf-conventions.html#use-of-the-crs-well-known-text-form-at>

Proposed change:

There will be occasions when a given CRS property value is duplicated in both a single-property grid mapping attribute and the `crs_wkt` attribute. In such cases the onus is on data producers to ensure that the property values are consistent. *If both a `crs_wkt` and grid mapping attributes exist, the attributes must be the same. As such, information from either one (or both) may be used to represent the CRS of the file, recognizing that the grid mapping parameters should always be completed as fully as possible. If conflicts exist between the representations, you should inform the provider so they can be addressed.* However, in those situations where two values of a given property are different, then the value specified by the single-property attribute shall take precedence. For example, if the semi-major axis length of the ellipsoid is defined by the grid mapping attribute `semi_major_axis` and also by the `crs_wkt` attribute (via the WKT SPHEROID[...] element) then the former, being the more specific attribute, takes precedence. Naturally if the two values are equal then no ambiguity arises.

Benefits:

1. The CRS could originate from several different formats such as WKT, PROJ, or SRS Authority Code. If there are errors in the conversion process to the CF or WKT representation, only the provider would have the original CRS representation. As such, if there are conflicts, the provider would be the best source to go to in order to resolve the conflicts.
2. Making this change will simplify the lives of software developers so they can just read in the WKT or grid mapping CF parameters for the CRS without a need to compare the two.

Summary of discussion

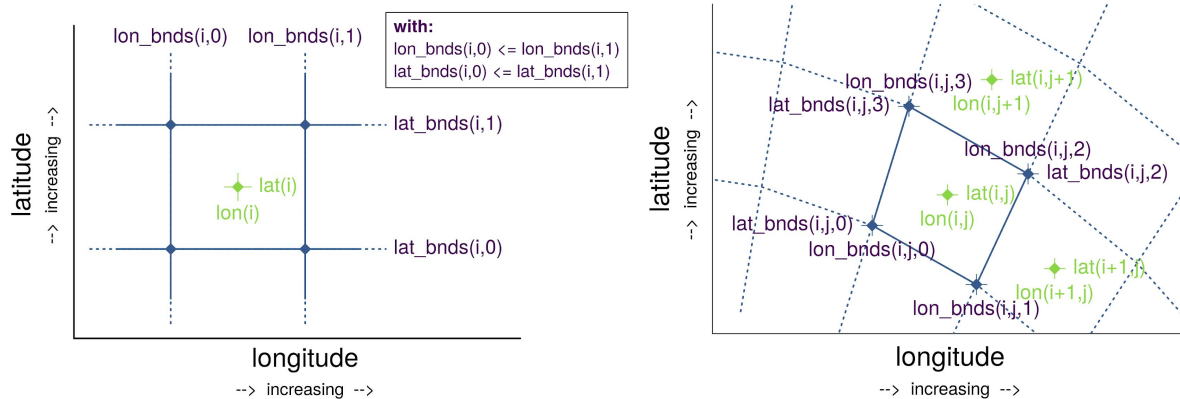
We finalized the text changes and added it to the issue ([comment link](#)). Next, we discussed changes needed to reduce the need for users to use the `crs_wkt` as a standalone attribute. In order to make this feasible. It needs to be possible to generate the CF grid mapping version from the WKT programmatically. The changes needed for this to occur include adding the units used by EPSG into the UDUNITS database as well as potentially adding supplementary units into the CF conventions. The next step is to identify the grid mappings that are currently not included. Lastly, a convention to represent the direction of projected coordinates is needed ([issue link](#)).

Adding figure to paragraph "Bounds for 2-D coordinate variables with 4-sided cells" in Section 7.1 on bounds (Daniel Heydebreck)

<https://github.com/cf-convention/cf-conventions/issues/193>

Introduction

The CF Conventions [Section 7.1 "Cell Boundaries"](#) contains a description on how to represent grid cell boundaries in `bounds` variables when the grid is defined by two one-dimensional coordinate variables (e.g. $\text{lon}(x), \text{lat}(y)$) and by two-dimensional auxiliary coordinate variables (e.g. $\text{lon}(x,y), \text{lat}(x,y)$). In the latter case (subsection "Bounds for 2-D coordinate variables with 4-sided cells"), the coordinates of the four vertices consist of four distinct values along each spatial axis (e.g. $\text{lon_bnds}(x,y,4), \text{lat_bnds}(x,y,4)$) and not of two distinct values as in the first case (e.g. $\text{lon_bnds}(x,y,2), \text{lat_bnds}(x,y,2)$). The order in which the coordinates have to be provided in the bounds variables is described in the text. However, it might be easier and faster to grasp by the reader by looking at a figure. The proposed figure contains this information.



Summary of discussion

Past discussion (before the CF annual meeting):

A modification of the plot colors had been requested to make the figure understandable for color-blind viewers. The color were modified accordingly. An additional question on the direction of x/y and lon/lat axis came up and was resolved.

Presentation and Discussion at the CF Annual meeting:

Number of participants: three

The figures and the reason for creating the figures were briefly presented including cdl-Examples. A short version of the aim of the figures was: A figure to indicate the order of grid

cell vertices in bounds variables (e.g. `lon_bnds(x,y,4)`) of 2dim coordinate variables (e.g. `lon(x,y)`) would be useful.

Past comments in the respective GitHub Issue to the figures were checked and were found to be fulfilled/resolved by the presented figures. No further improvements with respect to the figures were suggested. It was agreed that a pull request was to submit as next step (#276) and Daniel Lee volunteered to become moderator of the issue.

Moderation of proposals? (Chris Barker)

<https://github.com/cf-convention/cf-conventions/issues/151>

Introduction

Both due to a maturing project and new tools it's a good time to discuss and perhaps update and/or clarify how proposals for changes to CF are managed. The CF conventions have been managed in gitHub for a while now, so that we now have a bit of experience with the system. It has worked fairly well in some cases, not so well in others.

We will review the current rules for convention changes:

<http://cfconventions.org/rules.html>, and discuss what's worked well, what has not, and mull over possible ideas from improvements.

This discussion relates to the following issues in gitHub:

[#151](#) [#130](#) [#172](#)

(most of which are closed, but provide context)

Summary of discussion

Problem: A minority of discussions have grown up to a point which makes it difficult to follow or summarise.

- Should rules encourage use of separate summary document/position paper?
- Question: Does the initial issue description get updated as the conversation evolves?

Problem: Getting moderators for long or complex issues

- Do we need to grow the number of people on committees?
- Add a GitHub label "Needs Moderator": this could be automatically added to each new issue
- High-level overview of open issues (Daniel's suggested Kanban board e.g.) would allow us to notice when issues get stuck.

- How do people communicate with committees? For instance, to request a moderator. GitHub teams are not publicly visible. Not sure if people can @-mention teams they can't see.
 - Klaus - Conda-forge project uses @-mentions for non-visible teams (<https://github.com/conda-forge>)
-