

Implementing change in CF: An idea's journey

Daniel Lee

EUMETSAT

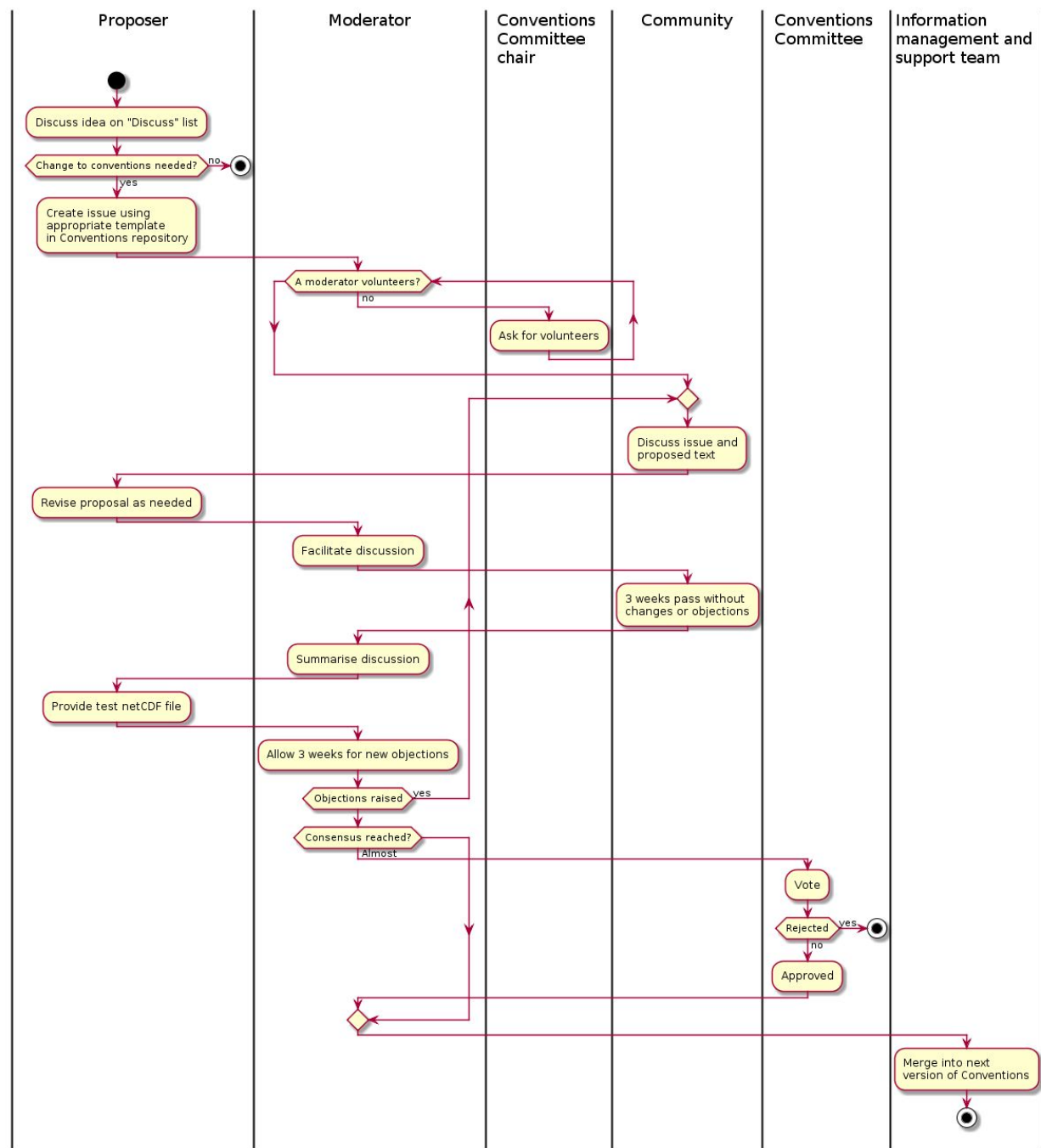


Contents

- How are changes made to the CF Conventions?
- How is this implemented in GitHub?
- What improvements might we introduce?

Changes: Process spec

- Text form in Rules
- Fear not! This is actually pretty straightforward
- Exception: Errata



Implementation in GitHub

- This is a snapshot, see the repository's CONTRIBUTING.md for current rules
- Issue is discussed in a single issue
- Text proposal is a pull request
- Moderator volunteers to guide discussion
- After content of pull request is agreed, it is merged

Implementation in GitHub: Example

- Blue sky discussion on Discuss
 - Volunteers swarm and begin work via issue, telcon, etc.
- Proposal issue on Conventions – all discussion goes here
- Linked pull request proposing text – review of changes takes place here (diffs and artefacts available)
- After discussion is concluded, a successful proposal is merged for the next release
- Throughout: Guidelines in CONTRIBUTING.md should help you!

What could be improved?

- Process improvements:
 1. Clarify when netCDF sample data is needed
 2. ?: Work packaging a release could be “shifted left”
 3. ?: Streamline the “3 week rules”
- Technical issues based off [#267](#):
 1. Easy to lose track a proposal’s status
 2. Hard to recognize what changed and why
 3. Reduction of technical debt via use of aliases – also potential for general cleanup

Process improvement: 1. Sample data

- Currently sample netCDF files are needed in all cases. This is not always enforced.
 - Specify when it is (not) needed?

Process improvement: 2. Packaging earlier

- When a release is created, the changelog is updated and authors are added to the Conventions. Should this be done as part of the Pull Request?

Process improvement: 3. The long silence

- The Rules specify 2 periods of 3 weeks silence each before a merge
 - Waiting for further contributions
 - Waiting for objections following the moderator's summary
- This inserts 2 time-triggered events into the workflow
- It might be better to request the summary and require silence thereafter only

Technical improvements: 1. Proposal status

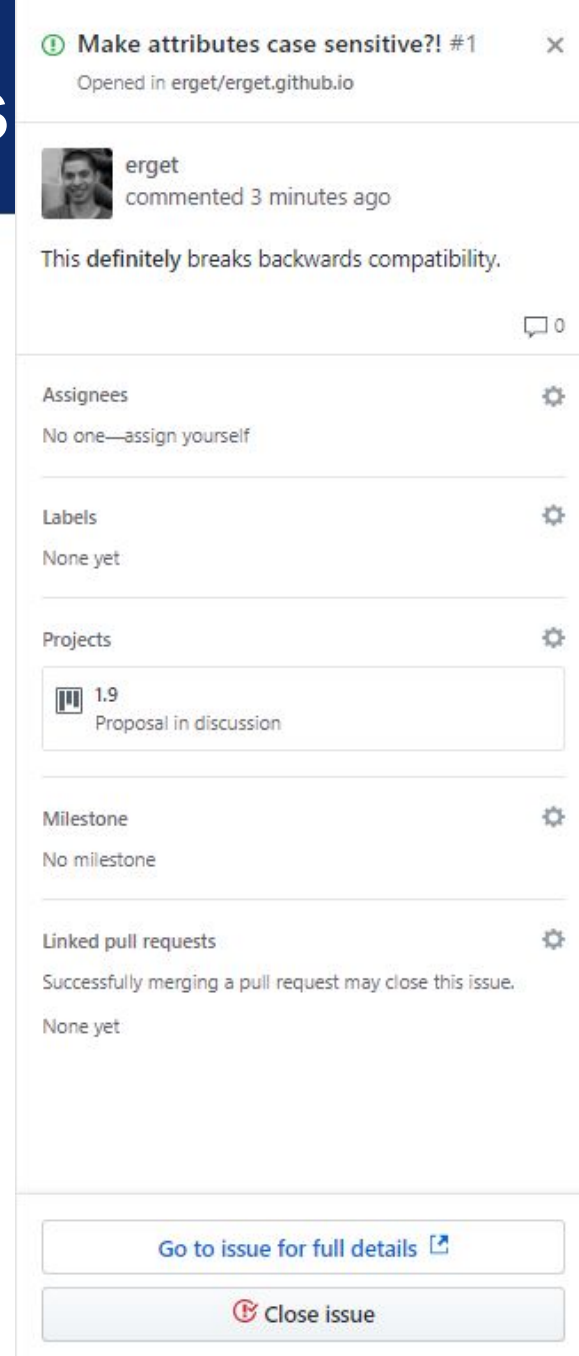
The screenshot shows a GitHub Projects board for the repository 'erget / erget.github.io'. The board is organized into columns representing different stages of proposal status. The 'Projects' tab is selected in the navigation bar. The board contains the following cards:

- Potential improvements** (1 card):
 - Extend CF to HDF? (Added by erget)
- Proposal in discussion** (2 cards):
 - Make attributes case sensitive?! (#1 opened by erget)
 - DOI for each attribute in a file (#2 opened by erget)
- 3-week silence ongoing** (0 cards)
- Post-summary silence** (0 cards)
- Accepted** (0 cards)
- Merged** (0 cards)

Additional UI elements include a search bar at the top, navigation links for Code, Issues (1), Pull requests (0), Actions, Wiki, Security (0), Insights, and Settings. The repository is noted as 'forked from cf-convention/cf-convention.github.io'. Interaction buttons for Watch (0), Star (0), and Fork (24) are visible in the top right.

Technical improvements: 1. Proposal status

- Issues are treated as “cards” that transition status along columns
- This reveals at a glance where an issue is located in its lifecycle without having to parse the comment thread
- Cards are “pulled” along horizontal axis



The screenshot shows a GitHub issue page for the title "Make attributes case sensitive?! #1". The issue was opened in the repository "erget/erget.github.io". A comment from the user "erget" is visible, stating "This definitely breaks backwards compatibility." The issue page includes sections for Assignees (No one—assign yourself), Labels (None yet), Projects (1.9 Proposal in discussion), Milestone (No milestone), and Linked pull requests (None yet). At the bottom, there are two buttons: "Go to issue for full details" and "Close issue".

Technical improvements: 2. Following changes 🙄

```
==== IDENTIFICATION OF CONVENTIONS
154
155 - We recommend that netCDF files that follow these conventions
    indicate
156 + this by setting the NUG defined global attribute
    **`Conventions`** to
157 - the string value "***CF-1.8***". The string is interpreted as a
158 - directory name relative to a directory that is a repository of
    documents
159 - describing sets of discipline-specific conventions. The
    conventions
160 - directory name is currently interpreted relative to the
    directory
161 - **`pub/netcdf/Conventions/`** on the host machine
162 - **`ftp.unidata.ucar.edu`**. The web based versions of this
    document are
163 - linked from the
164 -
    link:$$http://www.unidata.ucar.edu/netcdf/conventions.html$$[netc
    DF
165 - Conventions web page].
166
167 It is possible for a netCDF file to adhere to more than one set
    of conventions,
168 even when there is no inheritance relationship among the
    conventions. In this case,
169 @@ -177,7 +169,6 @@ of the data-writer to ensure that all common metadata is used with consistent me
170
171
172
173
174
175
176
177
178
179
180 -
181 [[description-of-file-contents, Section 2.6.2, "Description of
    file contents"]]
182 ==== Description of file contents
183
```

```
==== IDENTIFICATION OF CONVENTIONS
154
155 + Files that follow this version of the CF Conventions must
    indicate this by setting the NUG defined global attribute
    **`Conventions`** to a string value that contains "***CF-1.8***".
156 + The Conventions version number contained in that string can be
    used to find the web based versions of this document are from the
    link:$$http://cfconventions.org/$$[netCDF Conventions web page].
157 + Subsequent versions of the CF Conventions will not make invalid
    a compliant usage of this or earlier versions of the CF terms and
    forms.
158
159 It is possible for a netCDF file to adhere to more than one set
    of conventions,
160 even when there is no inheritance relationship among the
    conventions. In this case,
161
162
163
164
165
166
167
168
169
170
171
172
173
174
```

```
==== IDENTIFICATION OF CONVENTIONS
154
155 + Files that follow this version of the CF Conventions must
    indicate this by setting the NUG defined global attribute
    **`Conventions`** to a string value that contains "***CF-1.8***".
156 + The Conventions version number contained in that string can be
    used to find the web based versions of this document are from the
    link:$$http://cfconventions.org/$$[netCDF Conventions web page].
157 + Subsequent versions of the CF Conventions will not make invalid
    a compliant usage of this or earlier versions of the CF terms and
    forms.
158
159 It is possible for a netCDF file to adhere to more than one set
    of conventions,
160 even when there is no inheritance relationship among the
    conventions. In this case,
161
162
163
164
165
166
167
168
169
170
171
172
173
174
```

Identification of Conventions

We recommend that netCDF files that follow these conventions indicate this by setting the NUG defined global attribute `Conventions` to the string value " **CF-1.8** ". The string is interpreted as a directory name relative to a directory that is a repository of documents describing sets of discipline-specific conventions. The conventions directory name is currently interpreted relative to the directory `pub/netcdf/Conventions/` on the host machine `ftp.unidata.ucar.edu` . The web based versions of this document are linked from the [netCDF Conventions web page](#).

Files that follow this version of the CF Conventions must indicate this by setting the NUG defined global attribute `Conventions` to a string value that contains " **CF-1.8** ". The Conventions version number contained in that string can be used to find the web based versions of this document are from the [netCDF Conventions web page](#). Subsequent versions of the CF Conventions will not make invalid a compliant usage of this or earlier versions of the CF terms and forms.

It is possible for a netCDF file to adhere to more than one set of conventions, even when there is no inheritance relationship among the conventions. In this case, the value of the Conventions attribute may be a single text string containing a list of the convention names separated by blank space (recommended) or commas (if a convention name contains blanks). This is the Unidata recommended syntax from NetCDF Users Guide, Appendix A. If the string contains any commas, it is assumed to be a comma-separated list.

[The NetCDF Data Model: Groups](#)

Technical improvements: 2. Following changes 😊

```
@@ -85,7 +85,7 @@ vertical dimension:: A dimension of a netCDF variable that has an associated ver  
85  
86 No variable or dimension names are standardized by this  
convention. Instead we follow the lead of the NUG and standardize  
only the names of attributes and some of the values taken by those  
attributes. Variable or dimension names can either be a single  
variable name or a path to a variable. The overview provided in  
this section will be followed with more complete descriptions in  
following sections. <<attribute-appendix>> contains a summary of  
all the attributes used in this convention.  
87  
88 - We recommend that the NUG defined attribute **Conventions** be  
given the string value "CF-1.8" to identify datasets  
that conform to these conventions.  
89  
90 The general description of a file's contents should be contained  
in the following attributes: **title**, **history**,  
**institution**, **source**, **comment** and  
**references** ( <<description-of-file-contents>> ). For  
backwards compatibility with COARDS none of these attributes is  
required, but their use is recommended to provide human readable  
documentation of the file contents.  
91
```

```
2 ch01.adoc  
Overview  
No variable or dimension names are standardized by this convention. Instead we follow the lead of the NUG and standardize only the names of attributes and some of the values taken by those attributes. Variable or dimension names can either be a single variable name or a path to a variable. The overview provided in this section will be followed with more complete descriptions in following sections. [attribute-appendix] contains a summary of all the attributes used in this convention.  
We recommend that Files using this version of the CF Conventions must set the NUG defined attribute Conventions be given to contain the string value "CF-1.8" to identify datasets that conform to these conventions.  
The general description of a file's contents should be contained in the following attributes: title , history , institution , source , comment and references ( [description-of-file-contents] ). For backwards compatibility with COARDS none of these attributes is required, but their use is recommended to provide human readable documentation of the file contents.
```

- Line flow
- Trailing whitespace

Technical improvements: 3. Maintainable links

- Recent PR adjusted the same link. 60 lines touched.

```
126 - __Notes__: Notes on using the PROJ.4 software package
    for computing the mapping may be found at
127 -
    link:$$http://proj4.org/projections/laea.html$$[http://proj4.org/
    projections/laea.html]
128 -
129 -
131 + __Notes__: Notes on using the PROJ software package for
    computing the mapping may be found at
132 + link:$$https://proj.org/operations/projections/laea.html$$[http
    s://proj.org/operations/projections/laea.html]
```

- An alias would have reduced that by a lot
- We have a lot of links to maintain!

<http://cfconventions.org/Data/cf-conventions/cf-conventions-1.8/cf-conventions.html>
100% scanned - 131/131 URLs checked, 111 OK, 20 failed
Scan completed with 20 errors. [Full report](#) [Retry dead links](#)

- Check multiple sites at the same time - free - [login](#) or [sign up](#) for a free account.
- Check sites automatically on a regular basis with the Auto-Checker - [see options](#).

Way forward

- Big bang?
- Adjust when touched?
- Leave as-is?