Notes for the Hierarchical data & metadata extensions to CF

Participants: Jim Biard, Tim Whiteaker, Denis Nadeau, Aleksandar Jelenak, Kevin O'Brien, Charlie Zender

Breakout session 1.

Note: OMI level 2 is still hdfeos-5, Sentinel 5 precursor Level 2 uses netCDF with CF (and groups). There are no plans to reprocess OMI into the Sentinel 5 precursor file format (yet).

Note: see the plenary notes for remarks on hierarchical metadata.

For dimensions in hierarchies: should CF limit overriding dimensions under the same name? These are very confusing anyway.

- **Proposal**: If root contains a dimension 'lat', then other groups are not allowed to contain a dimension called 'lat'.

NCEI templates use integer variables to kluge groups. E.g., the platform variable here.

Can we find a way for CF to allow groups that doesn't fundamentally preclude existing CF workflows? E.g., for an attribute which references another variable, how is that variable referenced in the group hierarchy.

Using hdf references seems like a very elegant solution to the problem of referencing ancillary (& other) data. **Should we request adding this functionality to netCDF4?** When added the ncdump tool can translate this as required - however generating the binary file from CDL becomes then very 'interesting'.

One requirement for ancillary data we (KNMI) applied is that they use the same dimension(s). We currently use 'full' path names, using slashes to separate components. Another option is to use 'bare' names, and then search the hierarchy in the file where the same dimensions can be found.

Tools (python):

- Pytables (https://pypi.python.org/pypi/tables)
- Xarray (https://pypi.python.org/pypi/xarray/)

Trajectories

(this went too quickly for me, but those involved with buoys are to answer how hierarchies may help them).

Rich: Need to make sure we're not allowing something with groups that would be too onerous for software development.

Tim Whiteaker (9:45 p): Aleksandar and I discussed groups and CF during dinner and we generally came to the conclusions below. I'll take the blame if I've misrepresented his opinions. But it seems to me that, for CF to serve a goal of enabling discovery, access, interpretation and use, the following ideas may be useful. None of these are my own original ideas as I think they were mentioned in some form during the breakout, but they seem the most pertinent to me. I thought this would help provide something to hammer against, and I'm curious if any current users of groups believe their current use cases would be broken by anything suggested below:

1. Inheritance governs which attributes apply to variables.
2. A referencing scheme, like HDF5 object references, is needed when attribute values point to a variable that is not a sibling of the variable which the attribute describes.
3. With these two points in place, CF need not care or make any other special requirements about groups. This maintains flexibility to arrange whatever hierarchies you need, while also not creating a great amount of new constraints from CF.

Possible simple use cases

Two buoys in different places, measuring variables with different dimensions. Make this file discoverable.

Help for Rich Signell -

The soup ladle is in the Utensil group...

CZ summary notes, Action Plan, Thursday morning:

- Before drafting CF Group extension we must illustrate simplified use-case
    - Storage structure with Groups vs Flat
        - Plain 'Ole Groups (POG)
            - Exploit relationships implied by netCDF4/inheritance
            - Conventions for out-of-groups references
        - New feature type
            - Flat structure
                - Attributes indicate data/metadata relationships
                - CF1-compatible
            - Group structure
                - (As in POG)
    - Discoverability and Workflow Issues/Compatibility with Groups vs Flat
        - DAP4
        - THREDDS
        - Catalog
- Shared use-case with Discrete Sampling Geometries (DSG)
    - Trajectory or TrajectoryProfile-like dataset e.g., ARGO-float style
        - Organization
            - Start with two groups
            - Examine scaling with thousands of groups
        - Shapes:
            - Identical fixed dimension
            - Differing dimensions
            - Ragged vlen() (DSG focus) vs Multiple Record vs Fixed
- Suggest shared repository for CDL templates/issues: http://github.com/diwg