The CF Conventions: Governance and Community Issues in Establishing Standards for Representing Climate, Forecast, and Observational Data

Russ Rew[1], Bob Drach[2], Brian Eaton[3], Jonathan Gregory[4], Steve Hankin[5], Bryan Lawrence[6], Roy Lowry[7], Karl Taylor[2]

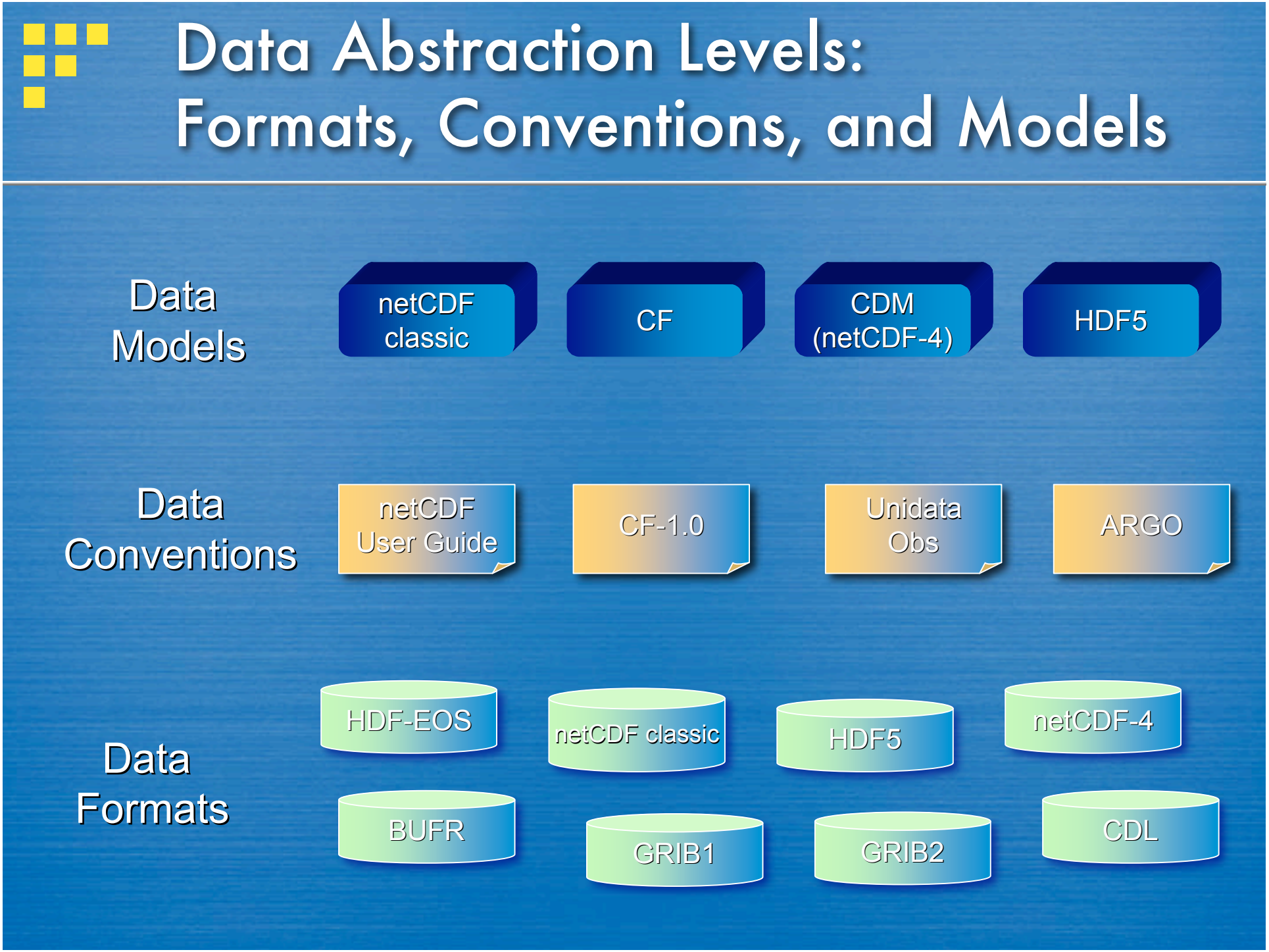
AGU Fall Meeting
San Francisco
December 2007

[1]UCAR Unidata, [2]Lawrence Livermore PCMDI, [3]NCAR, [4]NCAS/University of Reading, [5]NOAA Pacific Marine Environmental Laboratory, [6]NCAS/British Atmospheric Data Centre, [7]British Oceanographic Data Centre

Overview

- What are the CF Conventions?
- How widely used is CF metadata?
- How were the CF Conventions developed?
- What is the current CF governance structure?
- What are some strengths and weaknesses of CF governance?
- Where is CF headed?
- What's on the "Concluding Comments" slide?

Data Abstraction Levels: Formats, Conventions, and Models



What are the CF Conventions?

- A standard for encoding Climate and weather Forecast metadata in netCDF files: cfconventions.org
- Metadata conventions supporting interoperability for earth science data from different sources
- Intended for both model output and observational datasets
- Examples of CF metadata
  - Coordinate information needed to locate data in space and time
  - Standard names for quantities – to determine whether data from different sources are comparable
  - Additional grid information (e.g., grid cell bounds, cell averaging methods)

Where is CF metadata used?

- Widely used and accepted in the climate community
  - World Climate Research Programme's (WCRP's) Coupled Model Intercomparison Project phase 3 (CMIP3) multi-model dataset, used by Intergovernmental Panel on Climate Change (IPCC) Working Group 1
  - Paleoclimate Modeling Intercomparison Project (PMIP), Hemispheric Transport of Air Pollution (HTAP), regional groups, EU-funded ENSEMBLES prediction system for climate change, …
  - Planned use in model archives for next IPCC cycle
- Widely adopted in other netCDF archives for atmosphere, oceans, and surface data: ESMF, GFDL, Hadley Centre, NCAR, NOAA, …
- Supported by various software packages with facilities for analyzing, visualizing, subsetting, regridding, and aggregating data

Guiding principles of CF

1. Data should be self-describing, without external tables needed for interpretation.
2. Conventions should only be developed for things we know will be needed.
3. Conventions should not be onerous to use for either data-writers or data-readers.
4. Metadata should be readable by humans as well as easily interpretable by programs.
5. Redundancy should be minimized to avoid inconsistencies when writing data.

A brief history of CF

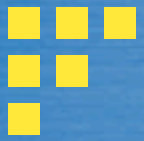
- Evolved from simple netCDF User Guide conventions (1989), COARDS standard (1995), GDT (1999), and NCAR CSM (1999) conventions
- 2000-2003: Developed by volunteer efforts (Brian Eaton, Jonathan Gregory, Bob Drach, Karl Taylor, and Steve Hankin)
- 2003: CF 1.0 released
- 2005: CF white paper discussing future governance circulated
- 2006: Revised white paper presented to WCRP WGCM
- 2007: Rules for community-initiated changes to CF conventions agreed upon

Governance structure

- CF Governance Panel established
- Control turned over to two working committees:
  - CF Conventions
  - CF Standard Names
- Committee work done via email and archived web discussion at cfconventions.org
- WCRP/WGCM has been asked to assume responsibility for stewardship
- WCRP/WGNE has been invited to appoint representation on CF Governance Panel

Some strengths of CF governance

- Successful international collaboration to codify best practices into a community standard
- Proven record of achieving interoperability
- Engagement of diverse communities to capture expertise for standard names
- Agreement on open process for evolving conventions and reaching consensus
- Commitment of organizational infrastructure and resources
  - BADC: Standard names – (50% FTE)
  - LLNL PCMDI: Web site support – (20% FTE)
  - UCAR Unidata: Library development (libcf) – (10% FTE)
- Discussion of CF issues at annual GO-ESSP (Global Organization for Earth System Science Portals) meetings



CF governance issues

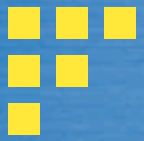
- How to get volunteers from community to help with
  - Creating and reviewing proposals to address new technical issues
  - Testing adequacy of proposed extensions
- How to balance desired simplicity versus necessary complexity?
- How to balance immediate needs of data providers versus stability needed by application developers?
- How to resist temptation to tinker, oversimplify, or over-generalize?

Future directions for CF

- Implementing CF metadata conventions for other file formats (besides netCDF)
- Supplying both data providers and application developers with library support for using CF
- Providing improvements for representing observational data and metadata
- Supporting more types of grids (staggered, curvilinear, nested)
- Supporting mappings between CF and other metadata standards and conventions
- Use of netCDF-4 data model and format

Concluding comments

- CF has undergone a two-year transition from informal maintenance by its authors to community governance.
- The CF Conventions transition seems moderately successful so far, but needs more active engagement by community volunteers.
- The CF Standard Names transition is also successful, with over 50 contributors and 900 standard names.
- Wide usage and real-world experience suggests CF metadata conventions are highly suitable for a broad community of data providers and users.
- To guarantee maintenance and ensure persistence as an internet resource, CF will need either
  - a single recognized authoritative organization to provide stewardship, or
  - a continued supply of interested and knowledgeable volunteers



For more information

- CF Conventions web site:

  cfconventions.org
- CF Conventions governance:

  cfconventions.org/governance