Data File Formats: netCDF

by

Tom Whittaker
University of Wisconsin-Madison
SSEC/CIMSS

2009 MUG Meeting
June, 2009

What it is...and is not...

- netCDF is a model for scientific data

  – Variables, dimensions, attributes, coordinates

- netCDF is <u>not</u> a DBMS

- netCDF provides libraries for data access:

  – C, Fortran, C++, Java, Python, Ruby, IDL, Matlab

- netCDF is <u>not</u> a data transmission format

- netCDF is a format for portable binary data

  – Direct access, metadata, appending new data

Brief History

- Developed at Unidata beginning in 1987
    - Concepts grew out of NASA Common Data Format
    - New Mexico Institute of Mining and Technology
    - SeaSpace Inc.
- Version 2 released in 1999; Version 4 in 2009
- Why not HDF?
    - Too easy to abuse; too complex; no conventions
    - NASA sees the light; funds netCDF4

What is important about netCDF?

- It is a self-contained, self-describing file format

- It is a community-driven project

- Updates are <u>always</u> backward-compatible

- It is extensible

- The Community is active

  – Creating conventions for structure

  – Testing and making suggestions

netCDF Data Model

- Variables: name, shape (list of Dimensions), type, Attributes, values

- Dimensions: name, length

- Attributes: name, type, value(s)

- Coordinate Variables: values associated with Dimensions

Supported File Formats

- One API for all these:

    – NetCDF3 ("classic"), netCDF4

    – HDF5, HDF4, HDF5-EOS, HDF-EOS

    – OpenDAP, GRIB1 and GRIB2, GINI

    – GEMPAK, DMSP, NEXRAD (level 2 & 3)

    – ADDE, BURF, GTOPO, McIDAS AREA

    – TDWR, URF, USPLN and NLDN

Conventions.....not "standards"

- Many, for our varied data

  – NUWG (1992) – observations

  – RAF – aircraft

  – CDC – gridded data

  – COARDS (1995) – grids

  – CF (now....) - grids, etc.

- Why?

  – Sharing data, writing applications

    - Software is NOT omnipotent!

....a word about Units

- International "standards" for Units

    - I say "deg_C"

    - You say "C"

    - I think you mean Coulombs.........

- UDUNITS package provides definitions and ways to combine units

- McIDAS-V (via the VisAD library) enforces legitimate combinations of parameters based on their units

    - e.g., cannot add a temperature to a pressure....

CF Conventions

- CF literally means "Climate and Forecast"
  - Mostly designed for gridded data
  - Good documentation available on-line
  - Defines structure
    - e.g., use of Coordinate variables to describe data variables
- McIDAS-V has a few "requirements"
  - e.g., Time
- What about satellite data?
  - Conventions in the CF-style being developed

What if.....

- ....I cannot read my data?

  – Check for CF compliance

  – Use netCDF Tools to check variables, etc.

  – Use NCML to fix things up

- ....I have one file per time, but want to animate?

  – For some, use the formula

  – For others, use NCML aggregation

- ....I do not know what to do...

  – Write the Help Desk

Who is responsible?

- The MUG Help Desk needs to know

- Don't be afraid to ask for assistance!

Unidata maintains netCDF:
http://www.unidata.ucar.edu/software/netcdf/

CF Conventions Group specifies CF
http://cf-pcmdi.llnl.gov/