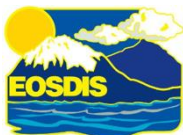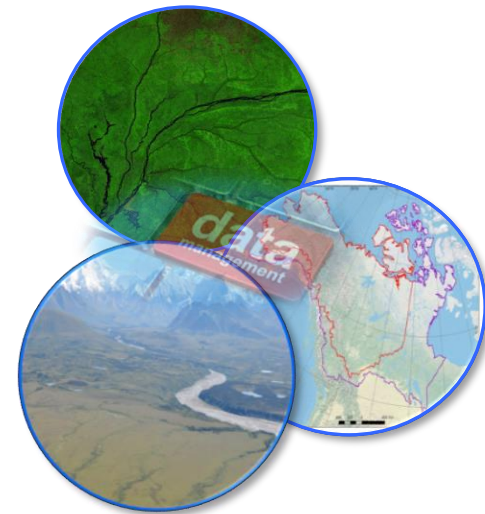# netCDF Standards

Yaxing Wei

ORNL Distributed Active Archive Center (ORNL DAAC)

Environmental Sciences Division

Oak Ridge National Laboratory

Oak Ridge, TN

March 26, 2017

Outline

Data/Metadata Standards and Tools

- *NetCDF and Climate & Forecast Convention*

Data Access/Visualization Standards and Tools

- *OPeNDAP*

- *NetCDF Subset Service (NCSS)*

NetCDF

- Network Common Data Form (NetCDF)
- NetCDF4 v.s. NetCDF3
    - Internal compression
    - Chunking
    - Classic & advanced data models
        - Extended primitive types
        - User-defined types
        - Groups
- A NetCDF4 file is actually an HDF5 file

Structure of a Classic-model NetCDF file

```
dimensions:
    lat = 360 ;
    lon = 720;
    time = UNLIMITED ; // (365 currently)
variables:
    double lat(lat) ……;
    double lon(lon) ……;
    double time(time) ……;
    double var1(time, lat, lon) ;
        var1:standard_name = "some_variable_standard_name";
    double var2(time, lat, lon) ;
        var2:standard_name = "some_variable_standard_name";
// global attributes:    ….
data:
    lat = -89.75, -89.25, -88.75, -88.25, -87.75, ……
    lon = -179.75, -179.25, -178.75, -178.25, -177.75, ……
    time = ……
    var1 = ……
    var2 = ……
```

NetCDF Files in Advanced Data Model
CARVE: L2 Atmospheric CO_2, CO and CH_4 Concentrations, Harvard CRDS, Alaska, 2012-2014

Climate and Forecast (CF) Conventions

- http://cfconventions.org
- Interoperability
  - Self-descriptive
  - Human- and machine-readable
- Widely-used
  - NASA ESDIS-adopted standard
  - IPCC Assessment Report numerical model outputs

Components of CF Conventions

- CF Standard Names

- CF metadata conventions
  - Spatial/temporal coordinates
  - Cell boundaries/shape/methods
  - Missing data/valid range
  - Data units
  - Data quality
  - History
  - Other attributes

Define Standard Variable Names

- CF Standard Name Table (v41)
  - Contain 2799 of (name, description, units) + 386 aliases in 9 categories: *Atmospheric Chemistry, Atmosphere Dynamics, Carbon Cycle, Cloud, Hydrology, Ocean Dynamics, Radiation, Sea Ice, and Surface*.

- What if there is no CF standard name that describes your variable?
  - Make your own: leaf_carbon_content -> leaf_nitrogen_content
  - Make it a CF standard name

Specify Spatial Info in CF Conventions (1)

- Provide cell **center** coordinates in Geographic Lat/Lon SRS and native SRS (*if different*)

```
variables:
    double lat(lat) ;
        lat:standard_name = "latitude" ;
        lat:long_name = "latitude coordinate" ;
        lat:units = "degrees_north" ;
    double lon(lon) ;
        lon:standard_name = "longitude" ;
        lon:long_name = "longitude coordinate" ;
        lon:units = "degrees_east" ;
data:
    lat = -89.75, -89.25, -88.75, -88.25, -87.75, …
    lon = -179.75, -179.25, -178.75, -178.25, -177.75, …
```

Specify Spatial Info in CF Conventions (2)

- Specify cell boundaries
  - Left-right boundary
  - Bottom-top boundary

```
double lat_bnds(lat=360, nv=2);
    :units = "degrees_north";
double lon_bnds(lon=720, nv=2);
    :units = "degrees_east";
double lat(lat=360);
    :bounds = "lat_bnds";
    :units = "degrees_north";
double lon(lon=720);
    :bounds = "lon_bnds";
    :units = "degrees_east";
```

Specify Temporal Info in CF Conventions (1)

- Specify calendar and time coordinates

dimensions:

   time = UNLIMITED ; // (365 currently)

variables:

   float time(time) ;

     time:standard_name = "time" ;

     time:calendar = "standard" ;

     time:units = "days since 1980-01-01 00:00:00 UTC" ;
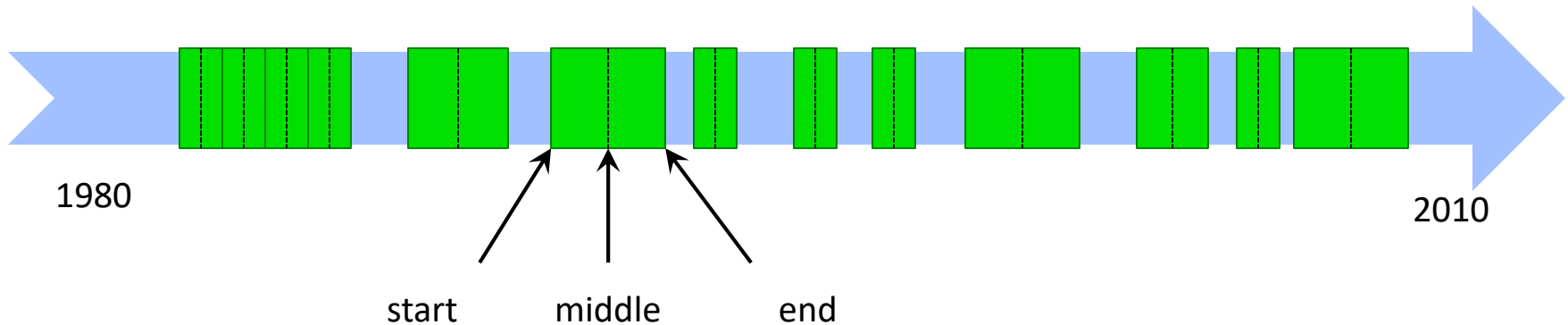
     time:bounds = "time_bnds" ;

data:

   time = 7305.5, 7306.5, 7307.5, …, 7668.5, 7669.5 ;

   // time = "2000-01-01 12", "2000-01-02 12", "2000-01-03 12",
…, "2000-12-29 12", "2000-12-30 12" ;

Calendar and time coordinates for Daymet data in 2000

Specify Temporal Info in CF Conventions (2)

- Specify time boundaries

1980   2010

start   middle   end

Specify Temporal Info in CF Conventions (3)

- Specify time boundaries

dimensions:
 time = UNLIMITED ; // (365 currently)
 nv = 2 ;
variables:
 float time_bnds(time, nv) ;
  time:long_name = "time boundaries" ;
  time:calendar = "standard" ;
  time:units = "days since 1980-01-01 00:00:00 UTC" ;
data:
 time_bnds = {7305, 7306; 7306, 7307; ...; 7669, 7670; }
 // time_bnds = {"2000-01-01 00:00:00", "2000-01-02 00:00:00"; "2000-01-02 00:00:00", "2000-01-03 00:00:00"; ...; "2000-12-30 00:00:00", "2000-12-31 00:00:00"; }

Time boundaries for Daymet data in 2000

Cell Methods

- To describe the characteristic of a variable that is represented by grid cell values

  - NARR dswrf: 3-hourly average, average across a 32km by 32km region

  - NARR precip: 3-hourly accumulated, average across a 32km by 32km region

- cell_methods
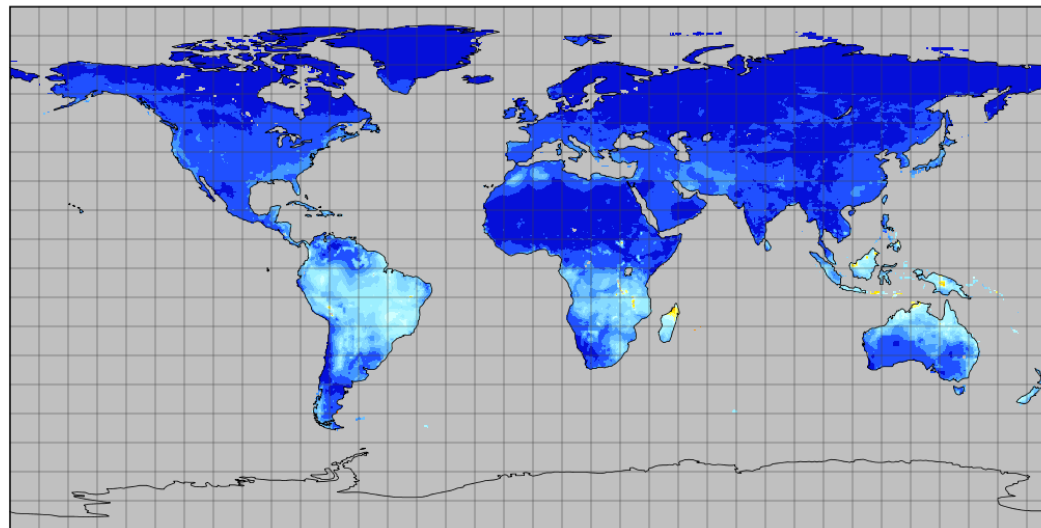
  - "time: mean area: mean"

  - "time: sum area: mean"

point
sum
maximum
median
mid_range
minimum
mean
mode
standard_deviation
variance

Missing Data

- Use _FillValue, valid_min, valid_max, and valid_range to indicate what values in a variable are considered to be valid or what values shall be ignored.

  float nbp(time=20, lat=74, lon=120);

  :_FillValue = -9999.0f; // float

Total Evaporation (kg m−2 s−1)

−4.2E-06   1.3E-05   3.0E-05   4.7E-05   6.5E-05   8.2E-05

Data Min = −4.2E-06, Max = 8.2E-05

Data Units

- UDUNITS
  - Based on International System of Units
  - Support conversion of unit specifications
  - Support arithmetic manipulation of units
  - conversion of values between compatible scales of measurement

Follow the rules and computers can then do a lot of work for you and others.

kg m-2 s-1

Kg/m2/month

kgC m-2 s-1

Units for Gross Primary Productivity (GPP)

Data Quality (1)

- Ancillary Data
    - one data variable provides metadata about the individual values of another data variable

```
float q(time) ;
    q:standard_name = "specific_humidity" ;
    q:units = "g/g" ;
    q:ancillary_variables = "q_error_limit q_detection_limit" ;
float q_error_limit(time)
    q_error_limit:standard_name = "specific_humidity standard_error" ;
    q_error_limit:units = "g/g" ;
float q_detection_limit(time)
    q_detection_limit:standard_name = "specific_humidity detection_minimum" ;
    q_detection_limit:units = "g/g" ;
```

Source: Example 3.2. of CF Conventions Document

Data Quality (2)

- Flags
    - The attributes **flag_values**, **flag_masks** and **flag_meanings** are intended to make variables that contain flag values self describing.

```
byte current_speed_qc(time, depth, lat, lon) ;
    current_speed_qc:long_name = "Current Speed Quality" ;
    current_speed_qc:standard_name = "sea_water_speed status_flag" ;
    current_speed_qc:_FillValue = -128b ;
    current_speed_qc:valid_range = 0b, 2b ;
    current_speed_qc:flag_values = 0b, 1b, 2b ;
    current_speed_qc:flag_meanings = "quality_good sensor_nonfunctional
outside_valid_range" ;
```

Source: Example 3.3. of CF Conventions Document

General Attributes

- **title**: description of what is in the dataset
- **institution**: specifies where the data was produced
- **references**: references that describe the data
- **source**: the method of production of the original data
- **history**: provenance of the data
- **comment**: miscellaneous information

// global attributes:

    :title = "CLM4 monthly GPP for MsTMIP RG1 global simulation" ;

    :source = "CLM-CN v4.0" ;

    :institution = "Oak Ridge National Laboratory" ;

    :references = "Oleson K. W. et al., 2010, Technical description of version 4.0 of the Community Land Model (CLM) NCAR Tech. Note NCAR/TN-478+STR (Boulder, CO: National Center for Atmospheric Research)" ;

Attributes for Global Monthly GPP Output from CLM4 Model

CF Conventions for Discrete Data

- Point Data

- Time Series Data

- Profile Data

- Trajectory Data

- Time Series of Profiles

- Trajectory of Profiles

CF-compliance Checker

- Check if a netCDF file complies with the CF comformance requirements and recommendations.

- http://cfconventions.org/compliance-checker.html

CF Conventions for HDF

- CF Conventions were originally tied to NetCDF
- The HDF Group and NASA are working together to make HDF work with CF
  - NASA ESDSWG Data Interoperability WG
  - Consider converting/augmenting HDF/HDF-EOS to netCDF
    - HDF4/HDF-EOS2 to CF Conversion Toolkit
    - HDF-EOS5 augmentation tool
  - Example from The HDF Group: Adding CF Attributes to an HDF file

Outline

Data Access/Visualization Standards and Tools

- *OPeNDAP*

- *NetCDF Subset Service (NCSS)*

- *OGC Web Services*

OPeNDAP

- Open-source Project for a Data Access Protocol
  - On-the-fly remote data access: provides users simple, remote access to large collections of data sets via the internet through HTTP
  - Subset data along any dimension(s)
  - Get data in ASCII (ascii), DAP (dods), and netCDF formats (nc/nc4)

OPeNDAP Data Request Syntax

http://HOSTNAME/PATH/TO/DATA.format?VAR1[start:stride:end],
VAR2[start1:stride1:end1][start2:stride2:end2][start3:stride3:end3]

An OPeNDAP Quick Start Guide

OPeNDAP Servers and Clients

- **Servers**
  - THREDDS Data Server
  - Hyrax
  - ERDDAP
  - PyDAP
  - GrADS Data Server (GDS)
  - ...

- **Clients**
  - NCL
  - CDAT
  - Ferret
  - GrADS
  - Panoply
  - ArcGIS Desktop
  - ArcGIS Pro
  - R
  - IDV
  - Matlab
  - Python
  - ...

NCSS – NetCDF Subset Service

- Similar to OPeNDAP

- Subset based on real coordinates and time ranges instead of matrix indices

- Get data in netCDF format

NCSS Data Request Syntax

http://HOSTNAME/PATH/TO/DATA?var=VAR1&var=VAR2
&north=70&west=10&east=50&south=20&horizStride=1
&time_start=1988-01-01T12:00:00Z&time_end=1988-12-30T12:00:00Z
&timeStride=1 &accept=format

Spatial and Temporal Subsets of Daymet Data Using the
THREDDS NetCDF Subset Service (NCSS) for Grids

OGC

OGC – Open Geospatial Consortium

"The Open Geospatial Consortium, Inc.® (OGC) is a non-profit, international, voluntary consensus standards organization that is leading the development of standards for geospatial and location based services"

http://www.opengeospatial.org/

26

OGC Standards

- OGC Web Coverage Service (WCS)
  - Web service for geospatial raster data access

- OGC Web Feature Service (WFS)
  - Web service for geospatial vector data access

- OGC Web Map Service (WMS)
  - Web service for geospatial data (raster or vector) visualization

- Many others ...

WCS Data Request Syntax

http://HOSTNAME/PATH/TO/WCS?
service=WCS&version=1.0.0&request=GetCoverage
&coverage=data_name&crs=EPSG:4326&bbox=west,south,east,north&resx=0.2
&resy=0.2&format=GeoTIFF&time=1972-01&interpolation=AVERAGE

Data Available Through These Web Services

- Most NASA data centers set up OPeNDAP/NCSS/OGC services to distribute their data holdings
  - Daymet, NACP, ABoVE, …
  - Land surfaces
  - Atmosphere
  - Ocean
- Coupled Model Intercomparison Project (CMIP5) model outputs