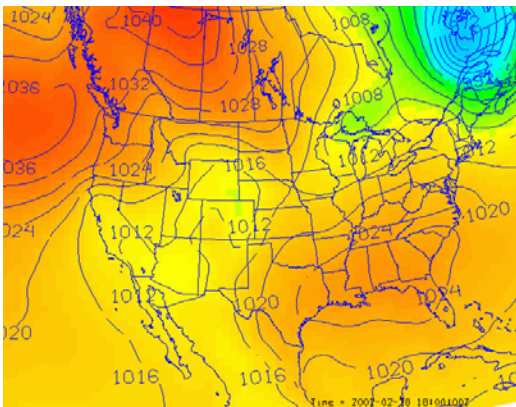
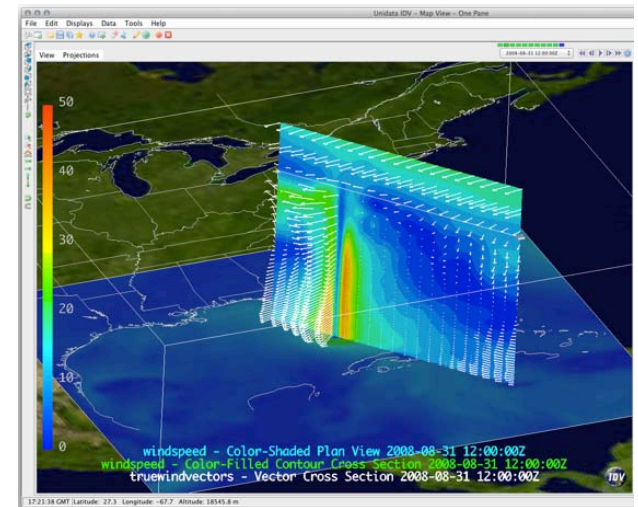


# Scientific Data Formats: NetCDF



Ethan Davis  
4 June 2014  
DCERC Workshop



# What is netCDF?

- Binary data format
- Software libraries / APIs
  - C (Fortran, C++, Python, ...) and Java
  - Upon which many data management, analysis, and visualization tools have been built
- Data model
  - Conceptual model of the data
  - Independent of data format details
  - Independent of programming language

# Why netCDF?

- Developed at UCAR Unidata about 25 years ago
- At the time
  - Most data was stored in an ad hoc manner.
  - Was not portable across computer platforms
  - Required knowledge of the dataset before it could be used
- Other formats also being developed
  - NASA's CDF (a precursor to netCDF)
  - NCSA's HDF4 (which lead to today's HDF5)
  - GRIB, BUFR (still standards for weather data)

# NetCDF format characteristics

- **Binary data format:** Generally used to store large, multi-dimensional arrays
- **Self-Describing:** A netCDF file includes metadata as well as data: names of variables, data locations in time and space, units of measure, and other useful information.
- **Portable:** Data written on one platform can be read on other platforms.
- **Direct-access:** A small subset of a large dataset may be accessed efficiently, without first reading through all the preceding data.
- **Archivable:** Access to all earlier forms of netCDF data will be supported by current and future versions of the software.

# What is netCDF?

## NetCDF Disk Formats

NetCDF version  
1.0, 1988

classic format

NetCDF version  
3.6.0, 2004

64-bit offset format

NetCDF version  
4.0, 2008

netcdf4/hdf5 format

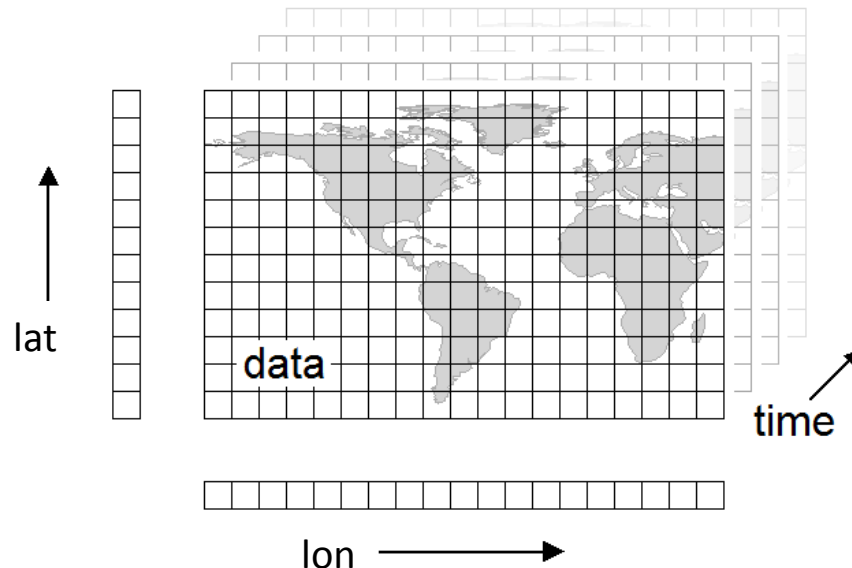
netcdf4/hdf5 classic model format

# What is netCDF?

- Classic netCDF data model
  - Multidimensional arrays of data values
    - Which share dimensions
  - Attributes

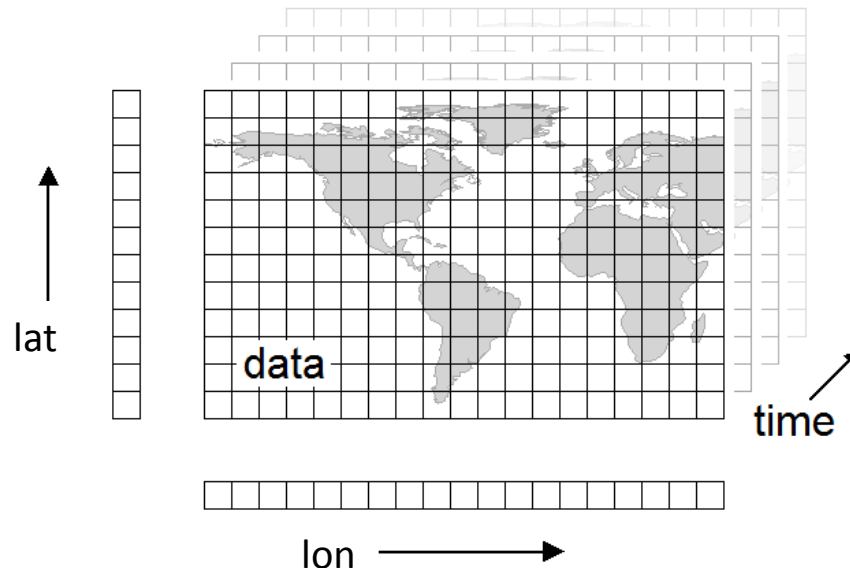
# What is netCDF?

- Classic netCDF data model
  - Multidimensional arrays of data values
    - Which share dimensions
  - Attributes



# What is netCDF?

- Classic netCDF data model
  - Multidimensional arrays of data values
    - Which share dimensions
  - Attributes



```
netcdf mydataset {  
  dimensions:  
    lat = 12 ; lon = 19 ; time = 4 ;  
  variables:  
    float lat(lat) ;  
    float lon(lon) ;  
    float temp(time, lat, lon) ;  
    float rh(time, lat, lon) ;  
  attributes:  
    :Conventions = "CF-1.6";  
}
```



# What is netCDF?

- Classic netCDF data model
  - Multidimensional arrays of data values
    - Which share dimensions
  - Attributes
- Enhanced netCDF data model
  - Adds hierarchical groups
    - Organize and group dimensions and variables
  - Adds structures
  - Only available in netCDF-4 (based on HDF5)

# Self-describing

# Self-describing?

- Not always

```
netcdf myAwesomeData {  
  dimensions:  
    a = 19 ; b = 12 ; c = 4 ;  
  variables:  
    float a(a) ;  
    float b(b) ;  
    float c(c) ;  
    float d(a, b, c) ;  
    float dd(a, b, c) ;  
}
```

# Self-describing?

- Human understandable?
  - Maybe.

```
netcdf mydataset {  
  dimensions:  
    lon = 19 ; lat = 12 ; time = 4 ;  
  variables:  
    float lat(lat) ;  
    float lon(lon) ;  
    float time(time) ;  
    float temp(time, lat, lon) ;  
    float rh(time, lat, lon) ;  
}
```

# Self-describing?

- Computer “understandable”?
  - Possibly

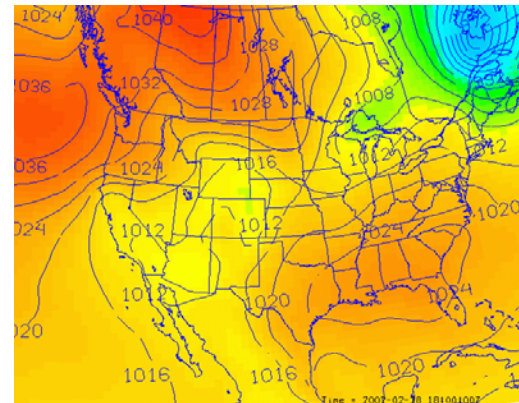
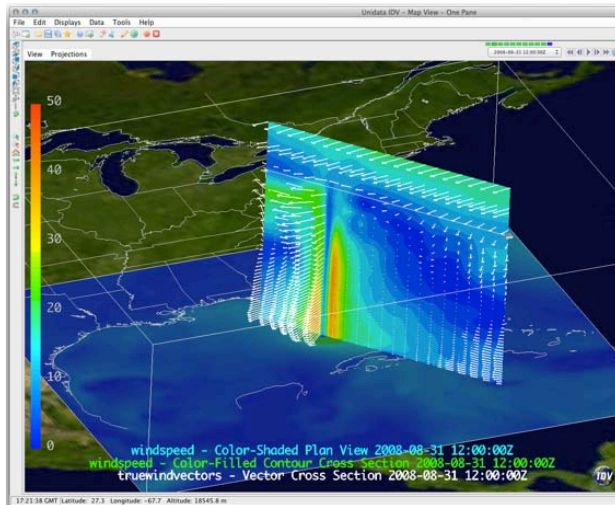
```
netcdf mydataset {  
  dimensions:  
    lat = 12 ;   lon = 19 ;   time = 4 ;  
  variables:  
    float lat(lat) ;  
      lat:units = "degrees_north" ;  
      lat:standard_name = "latitude" ;  
    float lon(lon) ;  
      lon:units = "degrees_east" ;  
      lon:standard_name = "longitude" ;  
    float temp(time, lat, lon) ;  
      temp:units = "Celsius" ;  
      temp:standard_name = "surface_temperature" ;  
    float rh(time, lat, lon) ;  
      rh:units = "percent" ;  
      rh:standard_name = "relative_humidity" ;  
  attributes:  
    :Conventions = "CF-1.6";  
}
```

# Community conventions

- Various community agreed upon attribute conventions have developed over the years
  - NUG, COARDS, NCAR-RAF, ...
  - CF (Climate & Forecast) conventions
    - Gridded data has long been the focus of CF
    - Now moving into observational data
  - ACDD (Attribute Convention for Data Discovery)
    - Originally based on Dublin Core and others
    - Current focus on aligning with ISO 19115

# CF conventions

- Historically dealt with gridded data







# NetCDF standards endorsements

- 2009: NASA ESDS standard for earth science data
- 2010: US FGDC endorsed netCDF-3 and netCDF-4 as “Common Encoding Standards”
- 2010: NASA ESDS endorsed CF Metadata Conventions as a standard for earth science data
- 2011: TDS includes nclSO services providing ISO-19115 metadata (from NOAA/NGDC)
- 2011: OGC approved netCDF as a core binary encoding standard
- 2013: OGC approved CF-netCDF data model

# Declaration of Compatibility

For future access to archives, netCDF development will continue to ensure the compatibility of:

- **Data access:** netCDF software will provide both read and write access to all earlier forms of netCDF data.
- **Programming interfaces:** C and Fortran programs using documented netCDF APIs from previous versions will continue to work after recompiling and relinking (if needed).
- **Future versions:** netCDF will continue to support both data access compatibility and API compatibility in future releases.



uniData

# Data Services

- Data access:
  - OPeNDAP, OGC WCS & WMS, NetCDF Subset Service (NCSS), etc.
- Catalog and discovery:
  - THREDDS, ISO 19115
  - Crawled by GeoPortal and GI-Cat
- Aggregation
  - Reduce the number of datasets by aggregating large collection of homogeneous data as a single, virtual dataset.

# Questions

- Ethan Davis, Unidata
  - [edavis@ucar.edu](mailto:edavis@ucar.edu)

# Questions

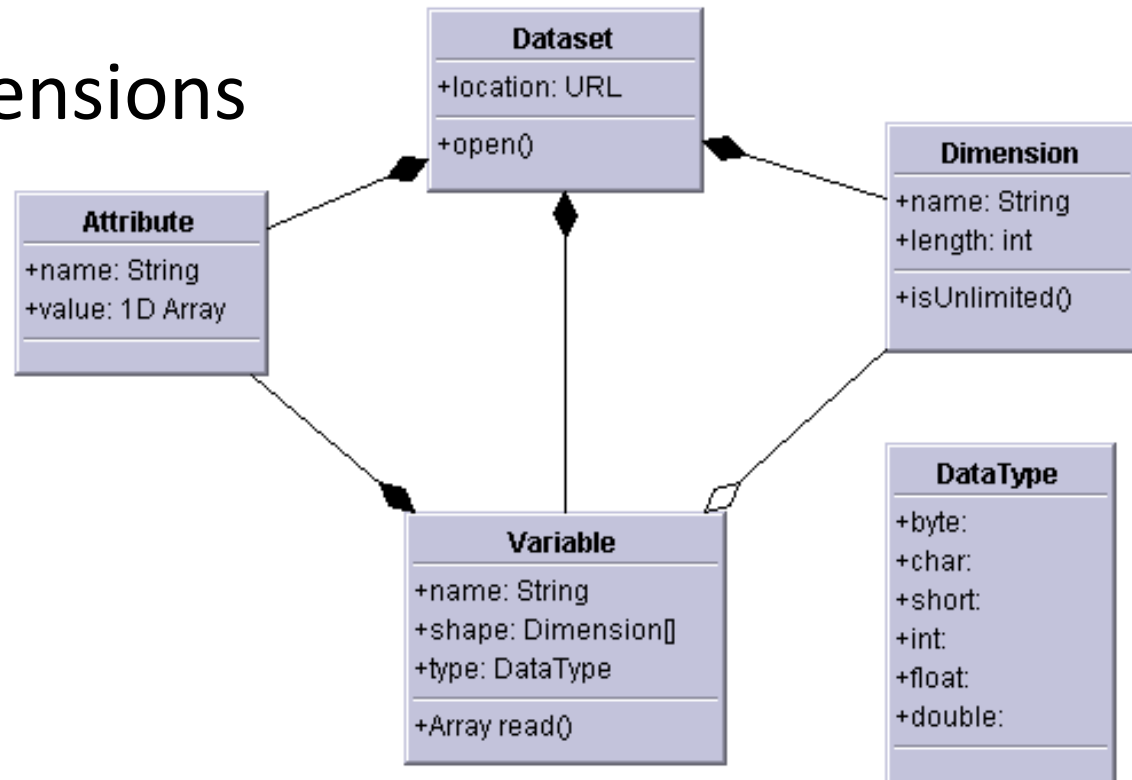
- Ethan Davis, Unidata
  - [edavis@ucar.edu](mailto:edavis@ucar.edu)
- Further reading
  - "Data Intensive Science and Scientific Data Infrastructure"
    - Presented by Russ Rew (Unidata) at 2011 ICTP mtg: [PPT](#), [PDF](#).
  - "NetCDF and Scientific Data Durability"
    - Presented by Russ Rew (Unidata) at 2009 ESIP mtg: [PPT](#), [PDF](#)



# NetCDF Classic data model

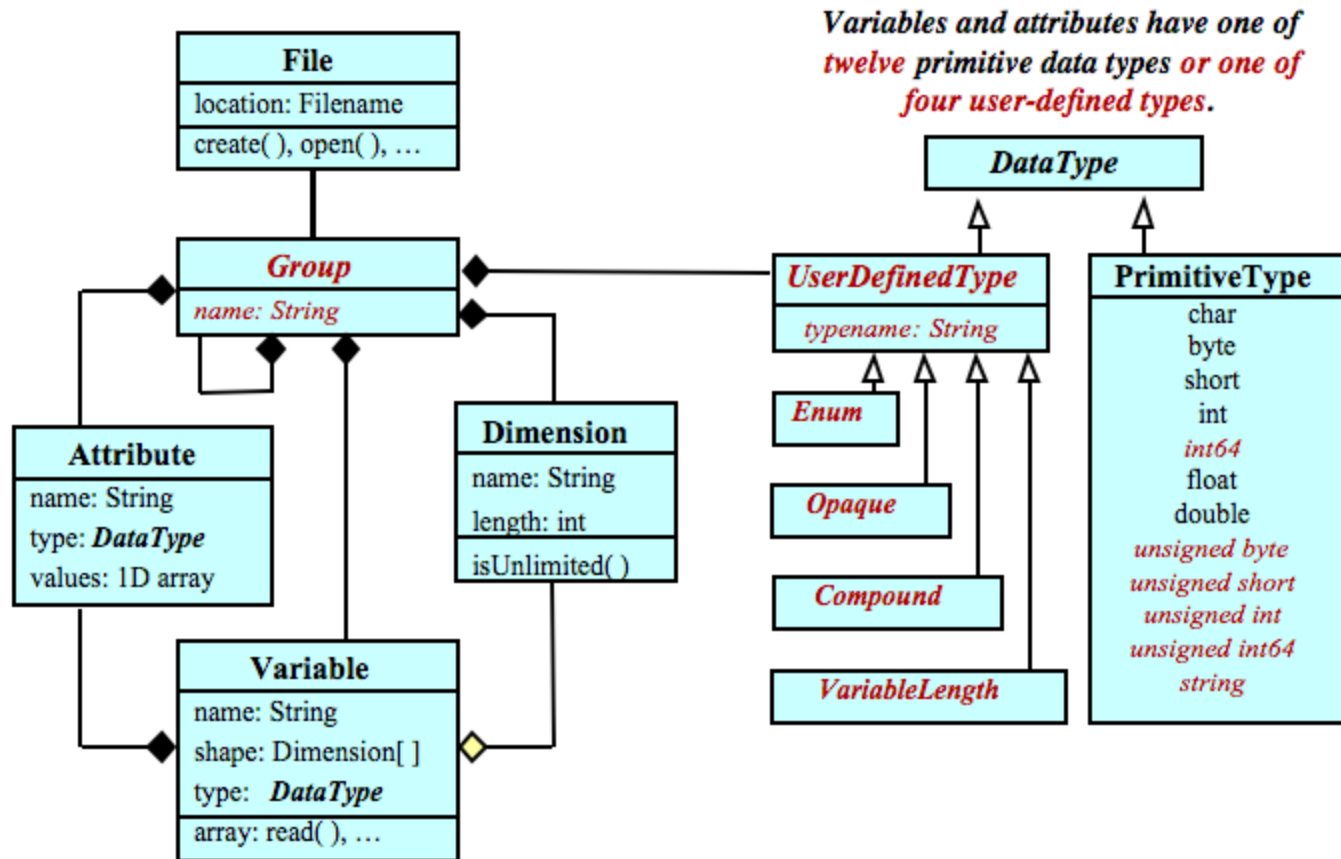
(aka netCDF-3)

- Multidimensional arrays of primitive values
  - byte, char, short, int, float, double
- Key/value attributes
- Shared dimensions
- Fortran77



# NetCDF Enhanced Data Model

(aka netCDF-4)



*A file has a top-level unnamed group. Each group may contain one or more named subgroups, user-defined types, variables, dimensions, and attributes. Variables also have attributes. Variables may share dimensions, indicating a common grid. One or more dimensions may be of unlimited length.*



# Background: What is Unidata?

- Where netCDF is developed and maintained
- Funded primarily by US National Science Foundation through UCAR
- Staff of about 22, including 13 developers
- Mission: data services, tools, and community leadership to advance Earth system science, enhance educational opportunities, and broaden participation
- Open source software for data access and distribution, analysis and visualization, community advocacy, workshops, and software support