# What is NetCDF ?

***And what are its plans for world domination?***

John Caron

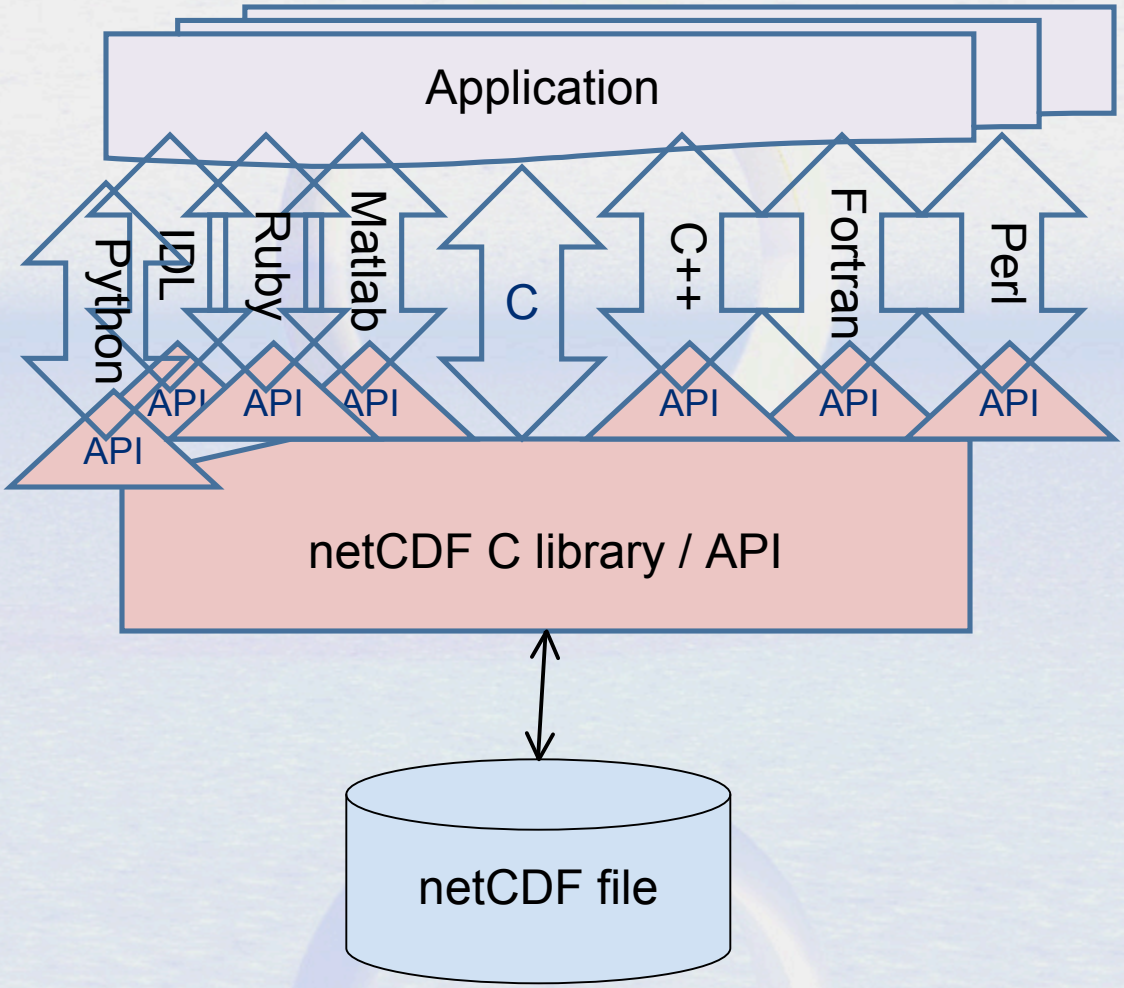
Unidata

August 2009

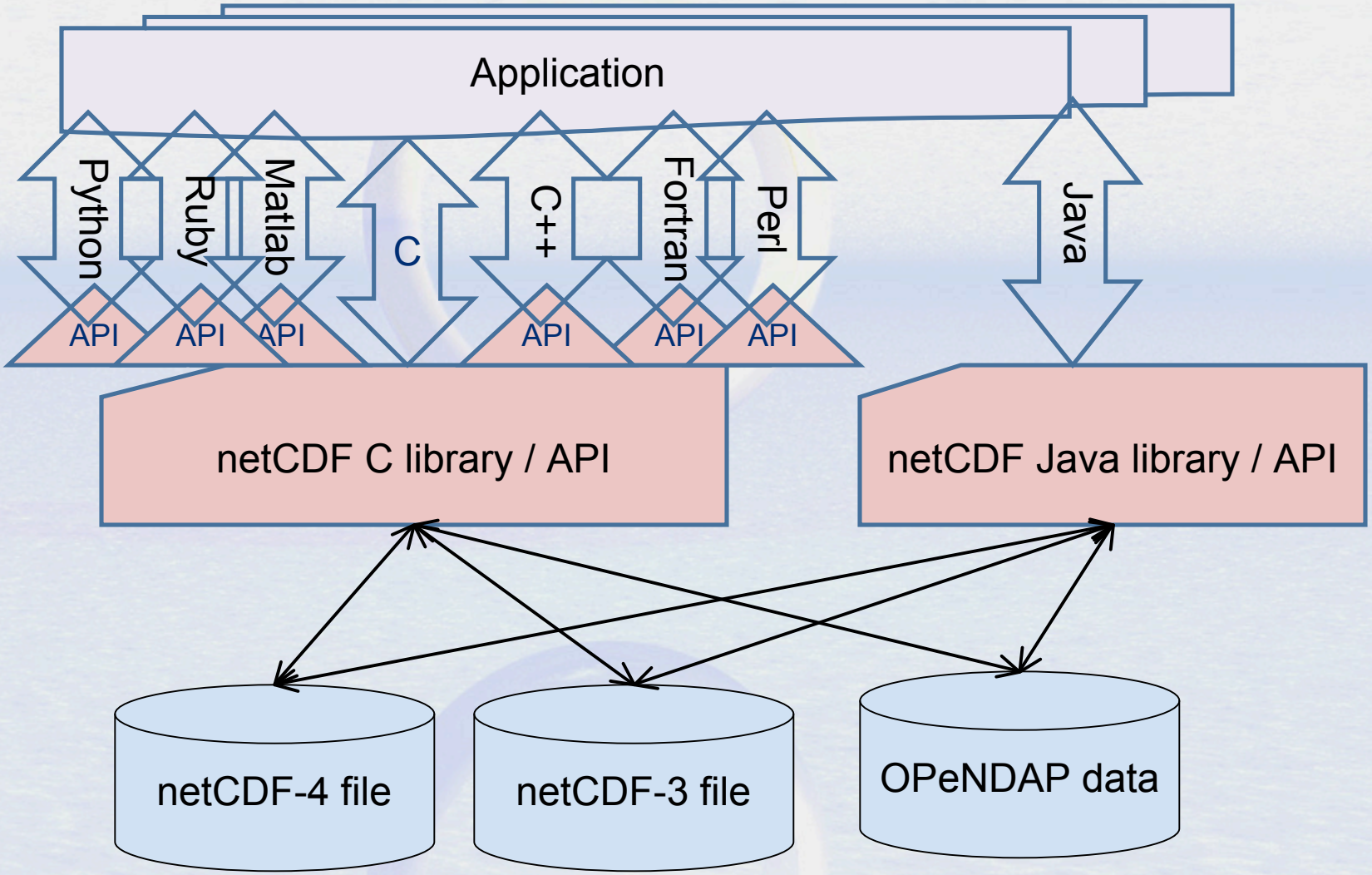
# NetCDF is….

- A file format
- A library
- An Application Programmer's Interface (API)
- A data model
- A dessert topping
- A floor wax

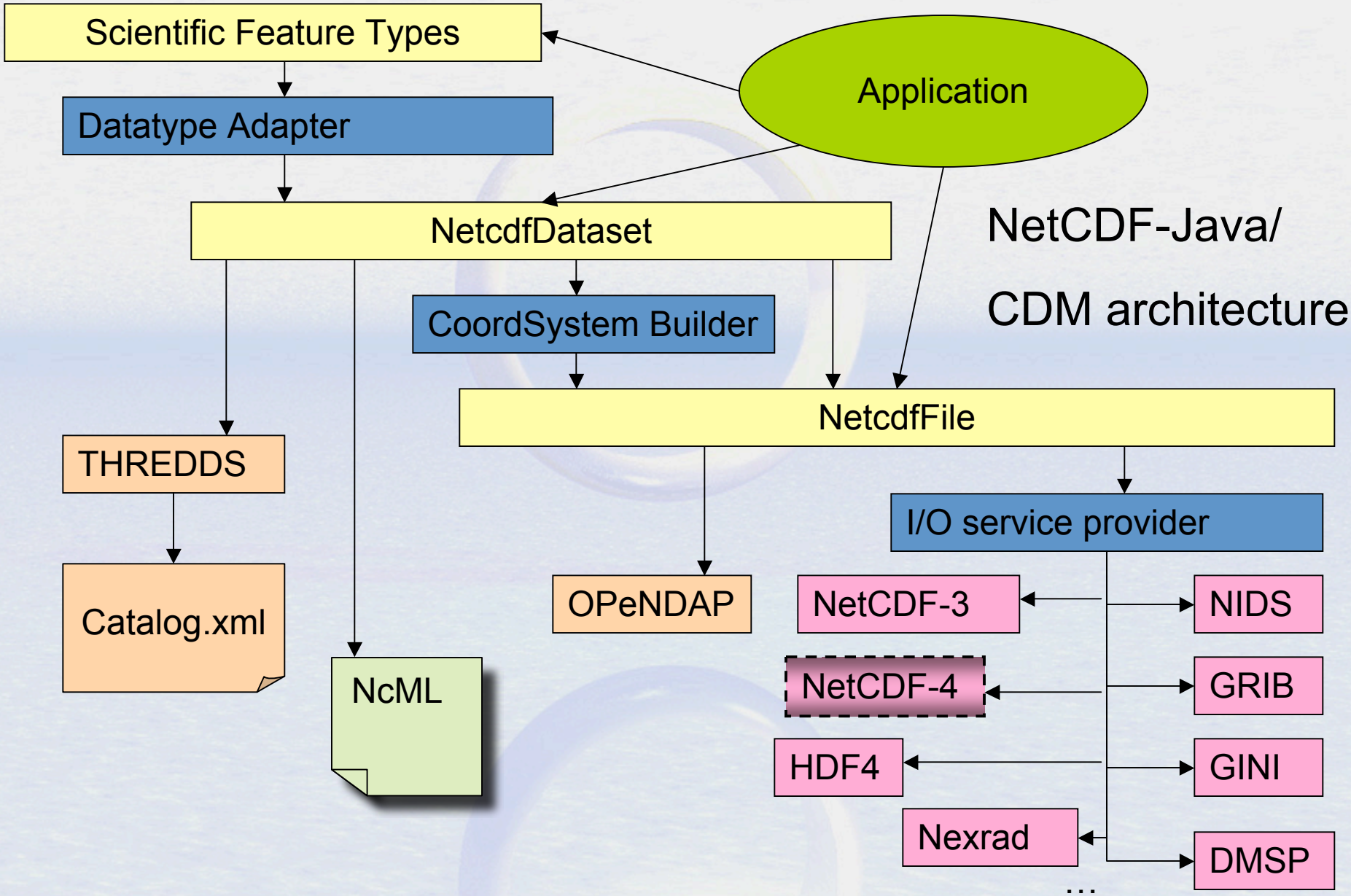
# In the beginning



# Things got more complicated



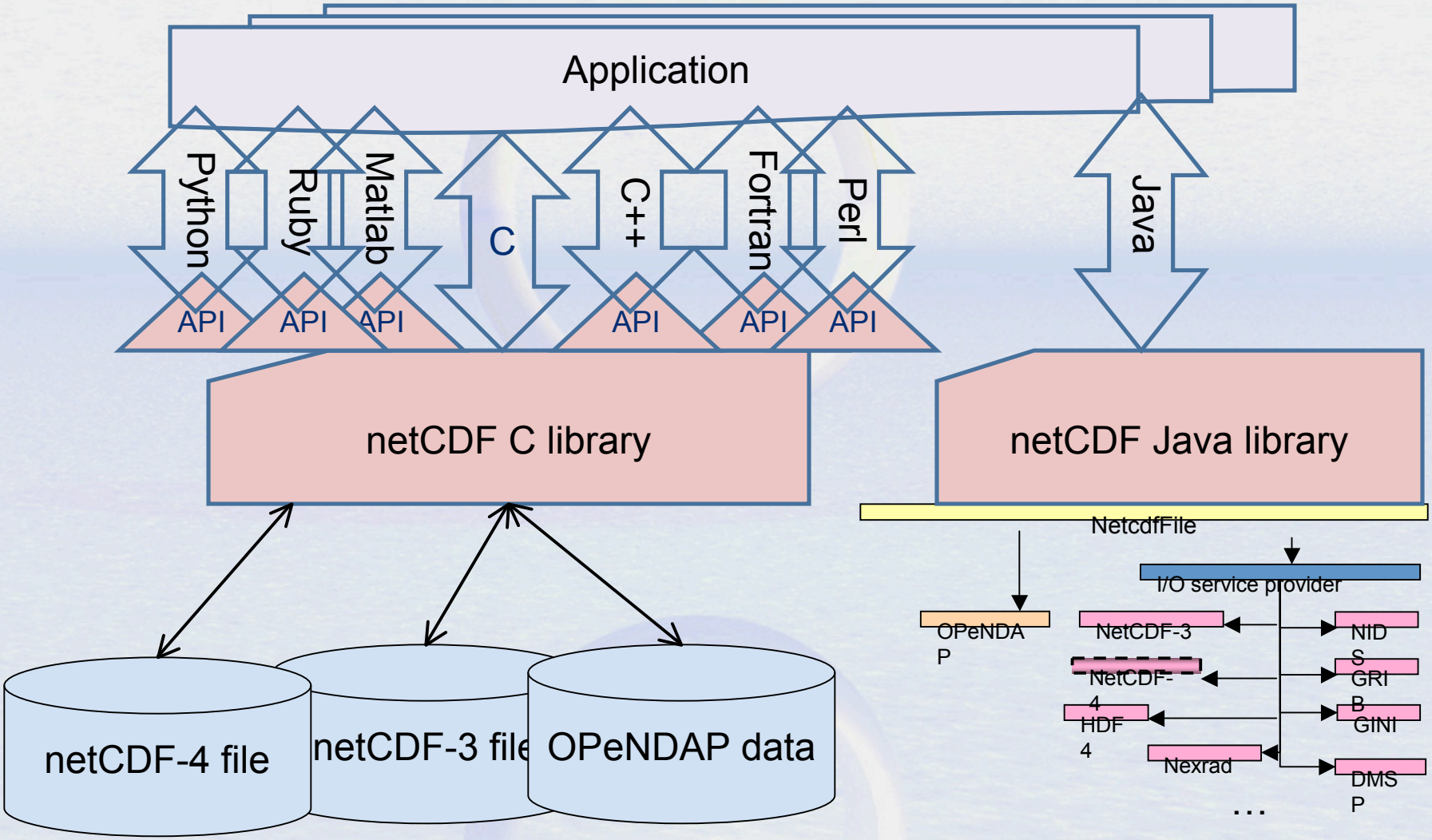
# But wait, there's more!



# Netcdf-Java 4.0 File Formats

- *General*: NetCDF-3, NetCDF-4, HDF5, HDF4, OPeNDAP
- *Gridded*: GRIB-1, GRIB-2, GEMPAK, McIDAS, UAMIV CAMx
- *Point*: BUFR, GEMPAK
- *Radar*: NEXRAD 2&3, DORADE, CINRAD, UF
- *Satellite*: DMSP, GINI, McIDAS, FYSAT
- *Misc*: GTOPO, NLDN, USPLN, etc

# What is netCDF ?



# NetCDF is a…

File format

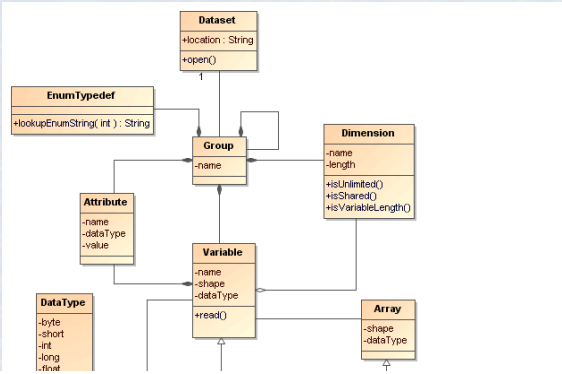
- Store data model objects
- Persistence layer
- NetCDF-3, netCDF-4

Software library

- Implements the API
- C, Java, others

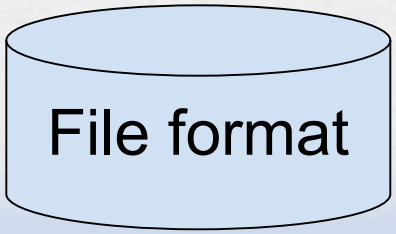
API

An ***API*** is the interface to the Data Model for a specific programming language




An **Abstract Data Model** describes data objects and what methods you can use on them

# NetCDF is a…

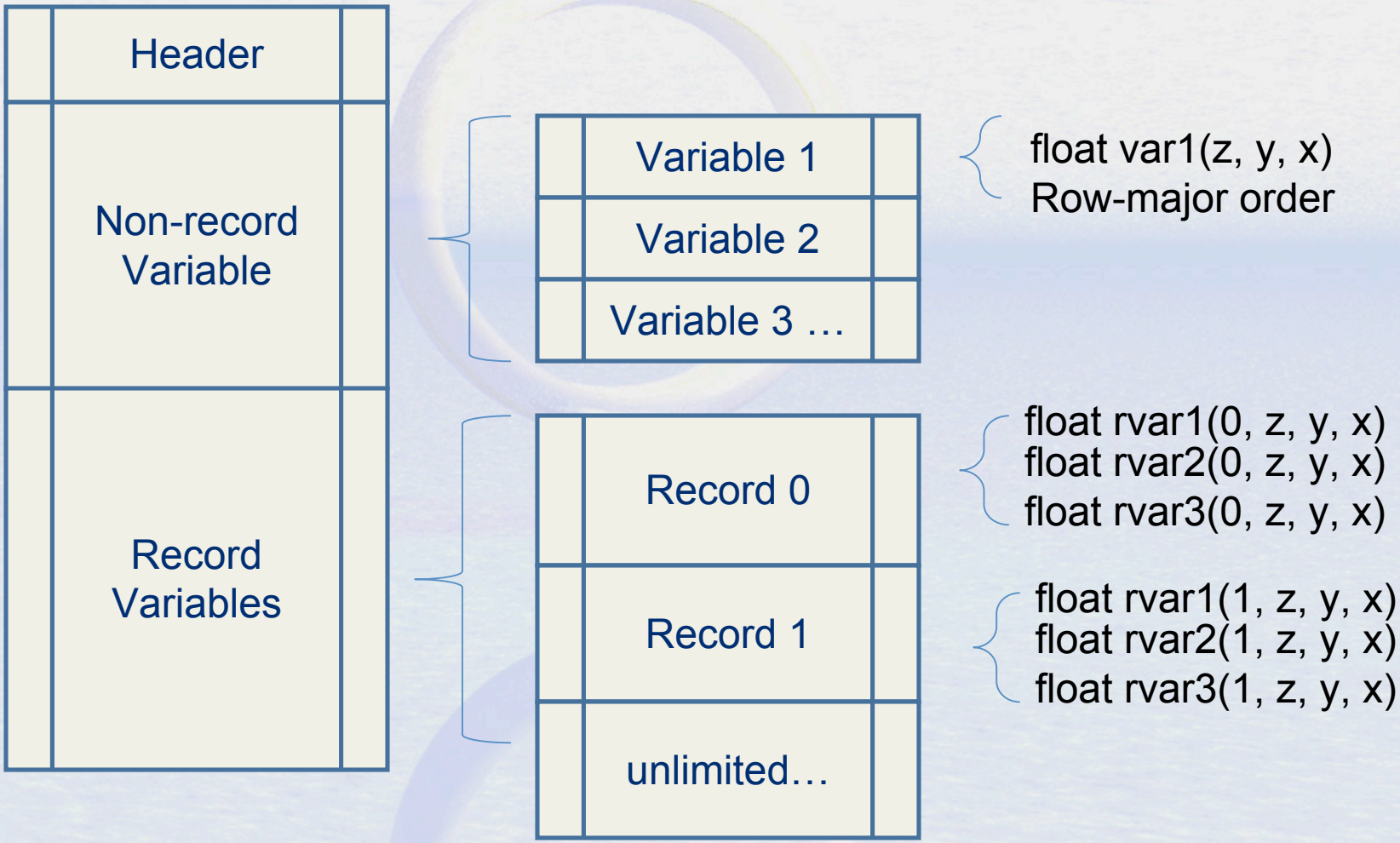



- Stores the objects in the data model
- Persistence layer
- NetCDF-3, netCDF-4

# What you should know about Storage Formats

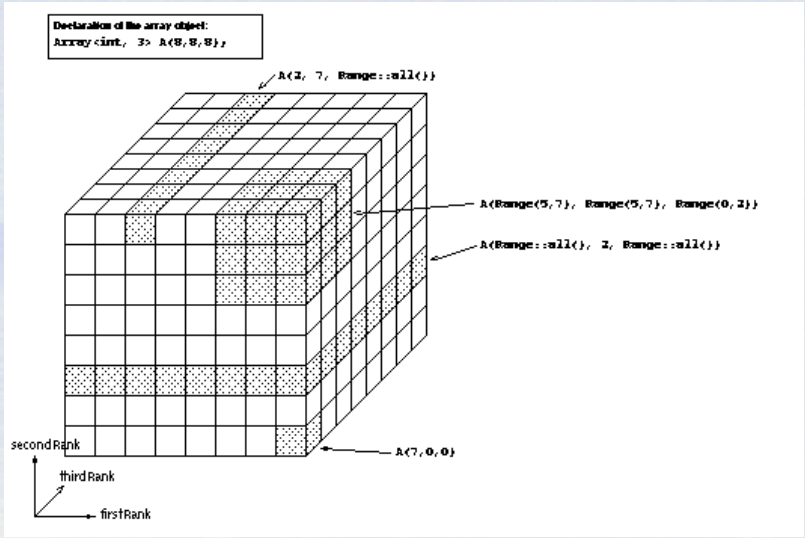
- Locality, locality, locality
- I/O cost is measured in # disk accesses
  - Entire block is read at once
  - Sequential access is 100x faster than random
- Many factors that affect this
  - Local disk, NFS mounted (shared), server RAID
  - The disk is caching sectors
  - The File System / OS is caching pages
  - Library may be caching data
- Applications can try to optimize file layout
  - write, read, common access patterns
  - Only matters for large I/O-bound apps

# NetCDF-3 file format



# NetCDF-4 file format

- Built on HDF-5
- Much more complicated than netCDF-3
- Storage efficiency
  - Compression : can optimize chunking for common I/O pattern
  - Compound types



# Row vs Column storage

- Netcdf-3 is a column store
  - All data for one variable is stored together
- Traditional RDBMS is a row store
  - All fields for one row in a table are stored together
- Netcdf-4 allows both row and column store
  - Row: compound type
  - Column: regular variable
- Recent RDBMS research focusing on possible advantages with column oriented storage

# NetCDF is a…

Software library

- Implements the API
- C, Java, third-party

# NetCDF Libraries

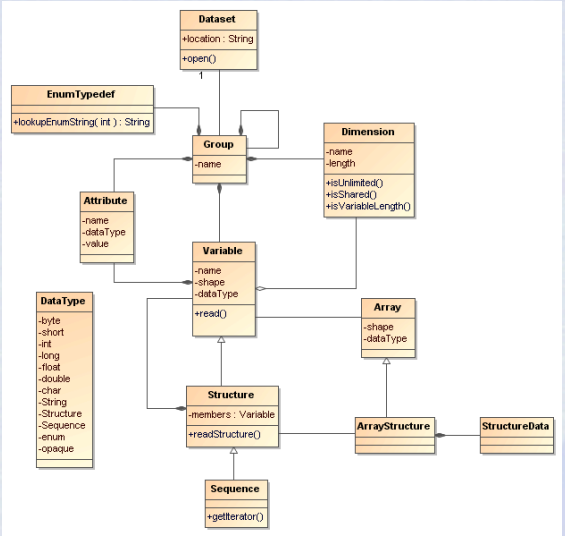
- NetCDF C library – reference implementation
  - Read/write netCDF-3 and netCDF-4
  - Read OPeNDAP (alpha)
- NetCDF  Java Library – exploratory
  - 100% Java == portable
  - Read netCDF-3, netCDF-4, OPeNDAP, many others
  - Only writes netCDF-3 (considering a JNI interface to C library for writing netCDF-4)
  - Thread safe, good for servers, used by the THREDDS Data Server (TDS)

# What you should know about Multicore CPUs

- Commodity CPU's wont get faster – too hot! Lifecycle cost dominated by electricity $$$
- Moores Law -> multiple CPUs on chip
- Multithreaded programs can take advantage of new multicore computer architecture
- Good for servers, harder for client programs to take advantage of this
- New languages (eventually)

# NetCDF is a…


An ***API*** is the interface to the Data Model for a specific programming language




An **Abstract Data Model** describes data objects and what methods you can use on them

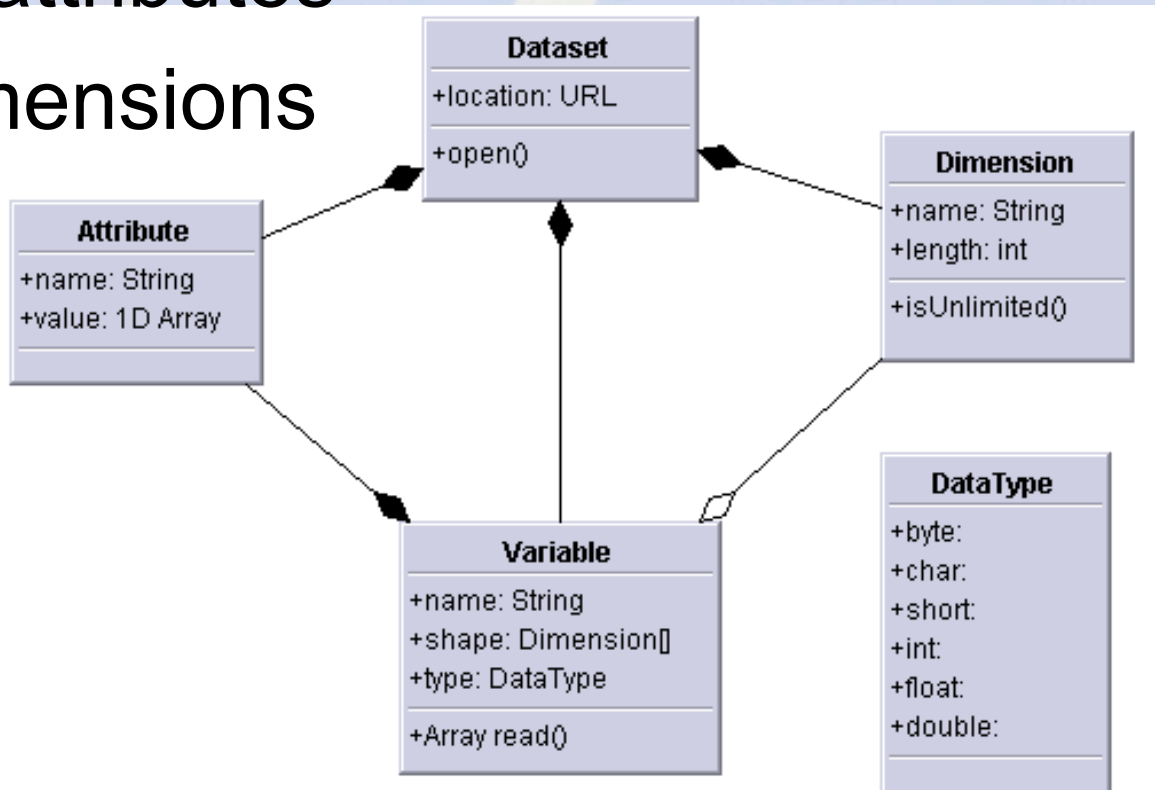
# NetCDF APIs

- Application Programmers Interface
  - Its what you have to deal with
  - Changing this breaks your code
- Lots of language bindings, same data model

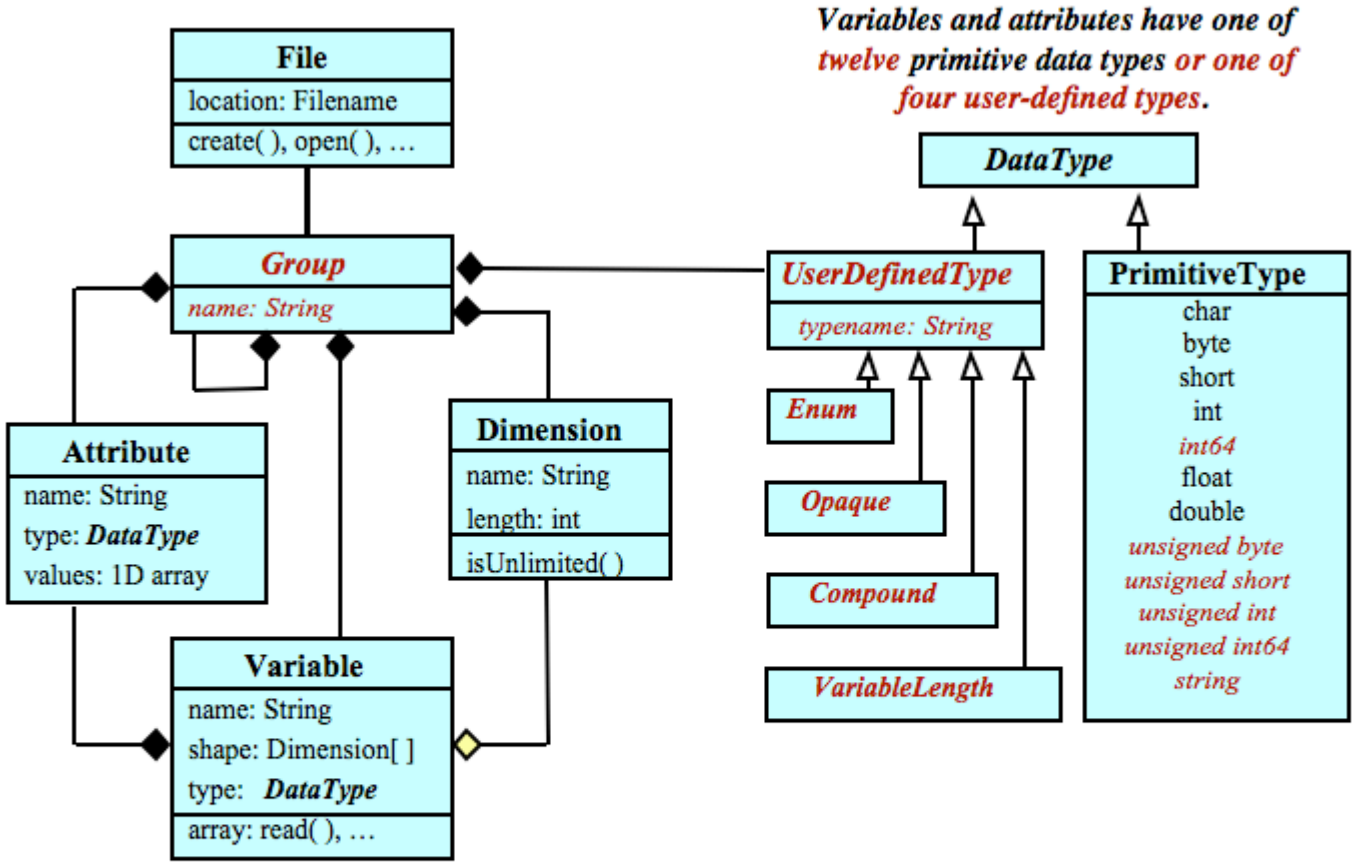
*An **API** is the interface to the Data Model for a specific programming language*

# NetCDF-3 data model

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

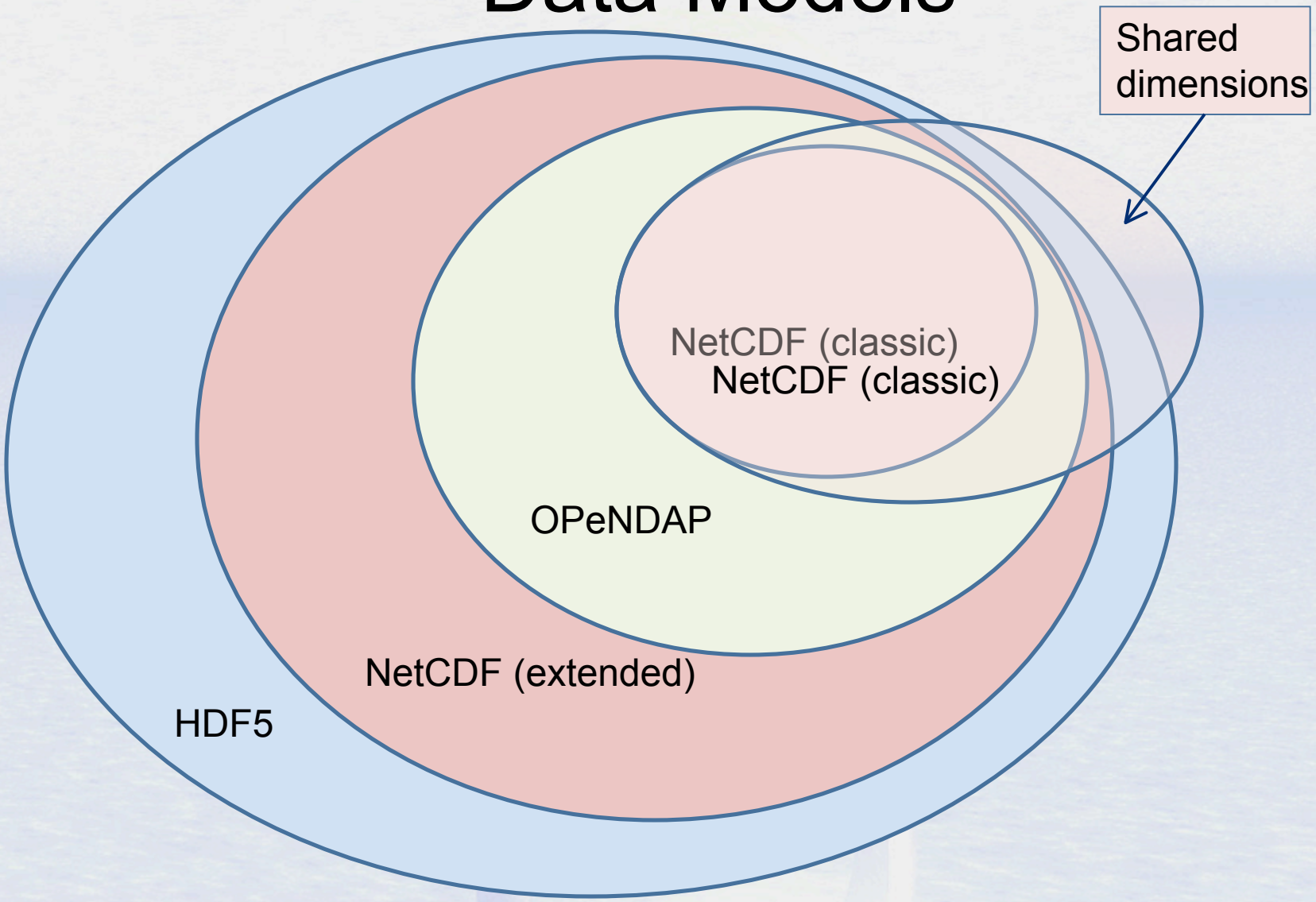


# NetCDF-4 Data Model




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.

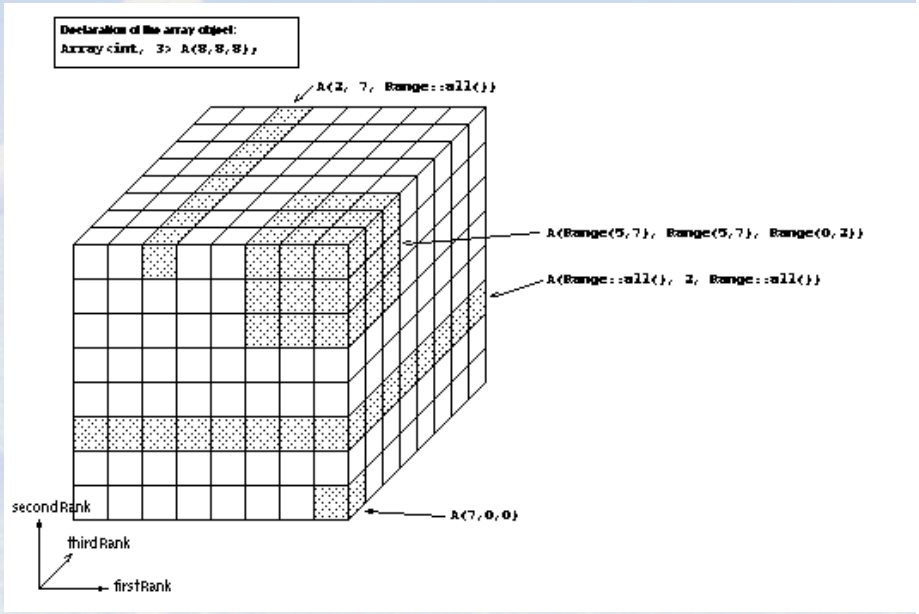
# NetCDF, HDF5, OPeNDAP Data Models



# Gridded Data

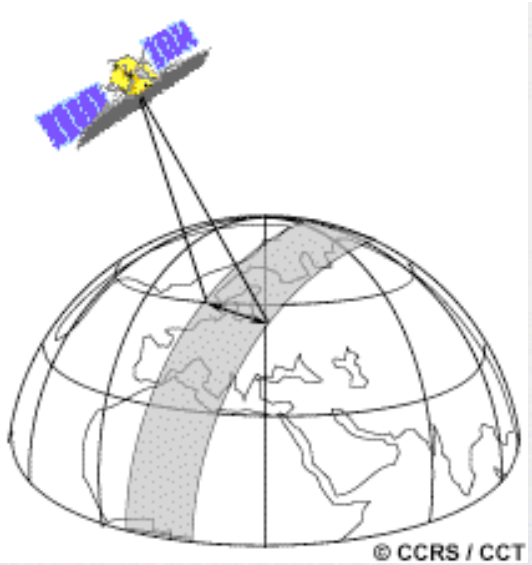
- Cartesian coordinates
- Data is 2,3,4D
- All dimensions have 1D coordinate variables (separable)

```
float gridData(t,z,y,x);
  float t(t);
  float y(y);
  float x(x);
  float z(z);
```



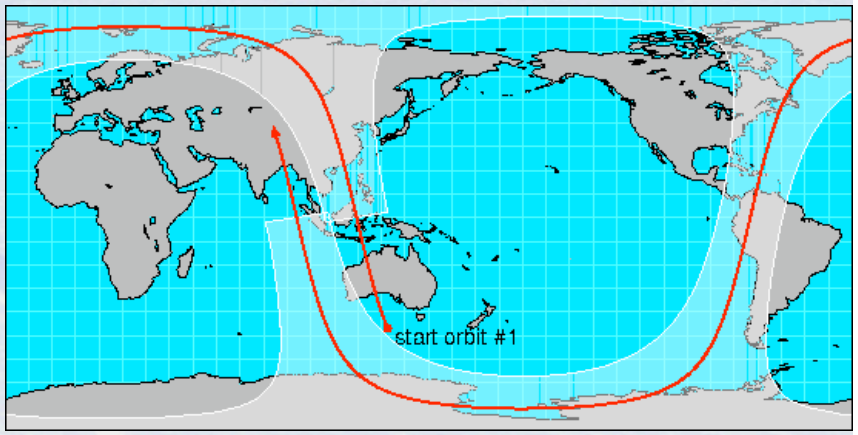

- netCDF: coordinate variables
- OPeNDAP: grid map variables
- HDF: dimension scales

# Swath




- two dimensional
- track and cross-track
- not separate time dimension
- aka *curvilinear coordinates*

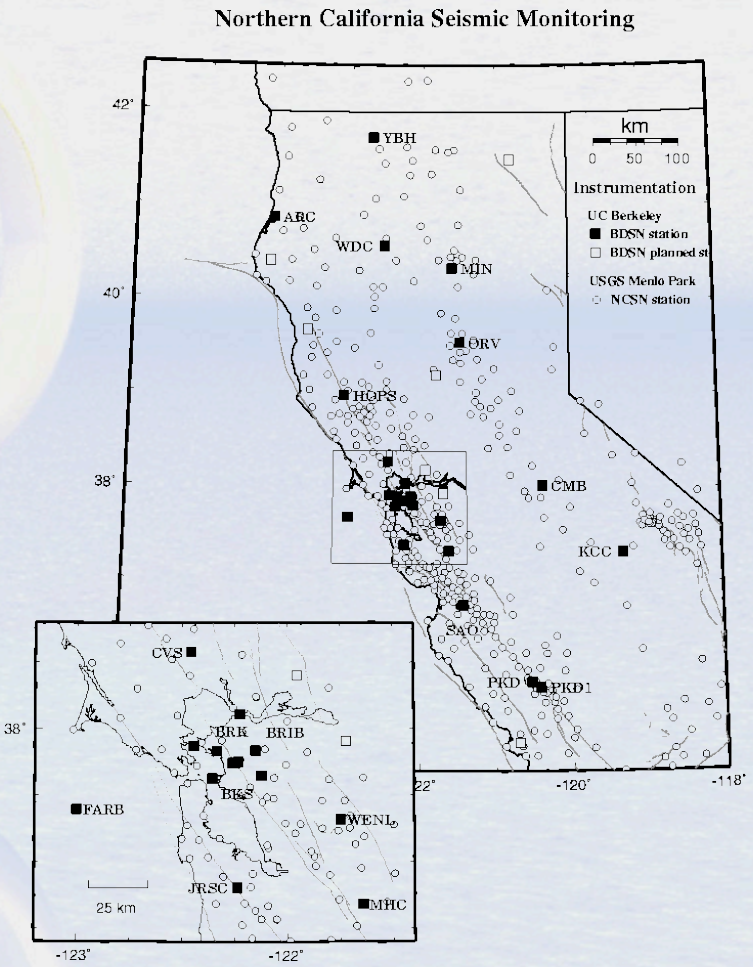
float swathData( track, xtrack)
  float lat(track, xtrack)
  float lon(track, xtrack)
  float alt(track, xtrack)
  float time(track)



# Point Observation Data

- Set of measurements at the same point in space and time = obs
- Collection of obs = dataset
- Sample dimension not connected

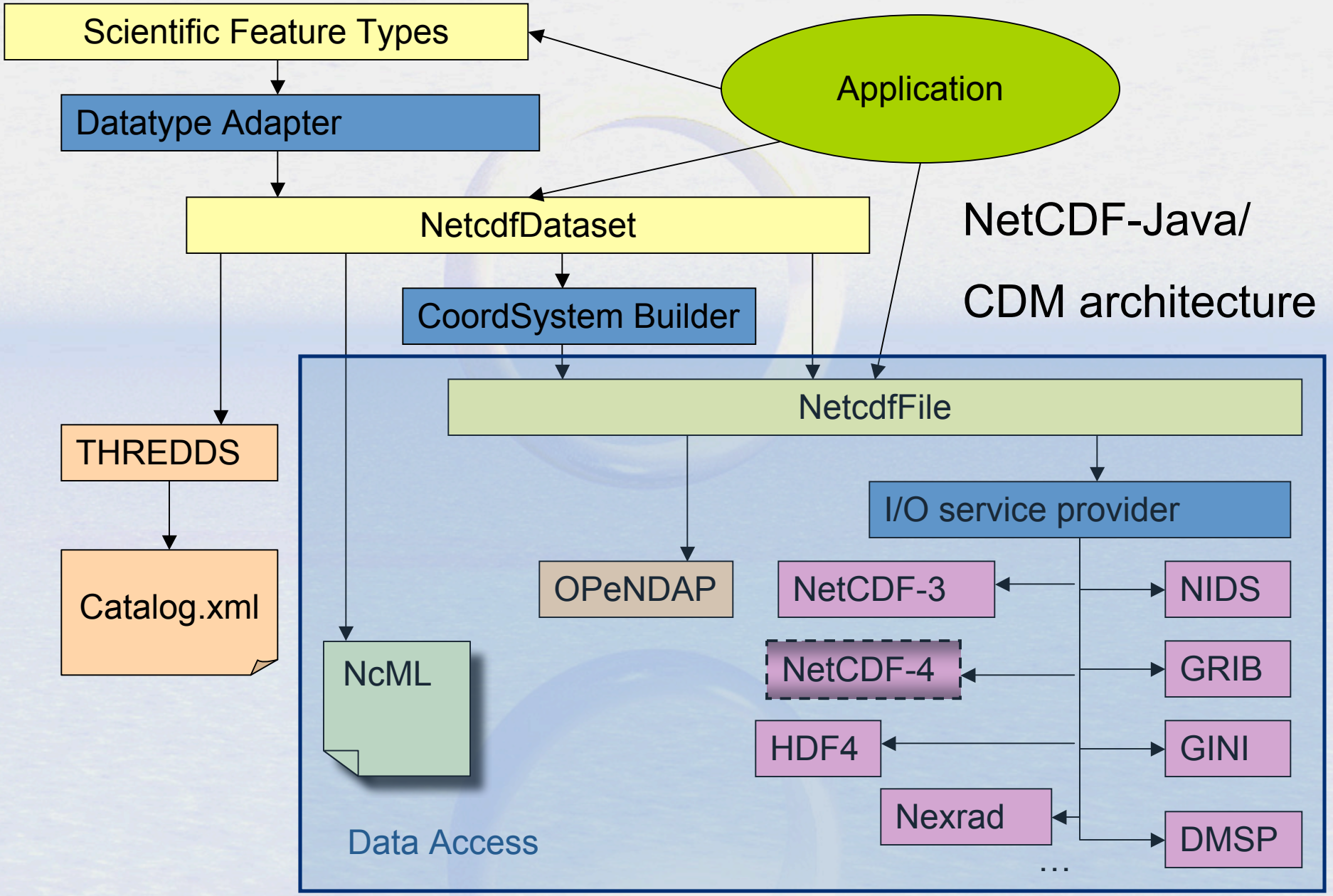
```
float obs1(sample);
float obs2(sample);
  float lat(sample);
  float lon(sample);
  float z(sample);
  float time(sample);
```



# Shared Dimensions Status

- netCDF
  - Shared dimension plus conventions is general solution for coordinates
  - *:coordinates = "lat lon alt time"*
- OPeNDAP
  - No shared dimensions in current data model
  - Shared dimensions will be added to DAP-4
- HDF5
  - No shared dimensions in current data model
  - HDF-EOS added shared dimensions in metadata
  - NetCDF-4 adds a workaround
    - NetCDF-4 not a subset of HDF-5
  - NetCDF-4 does not (yet) read all HDF-5 objects
    - HDF-5 not a subset of NetCDF-4

# Back to API / Data Models



## NetCDF “Index Space” Data Access:

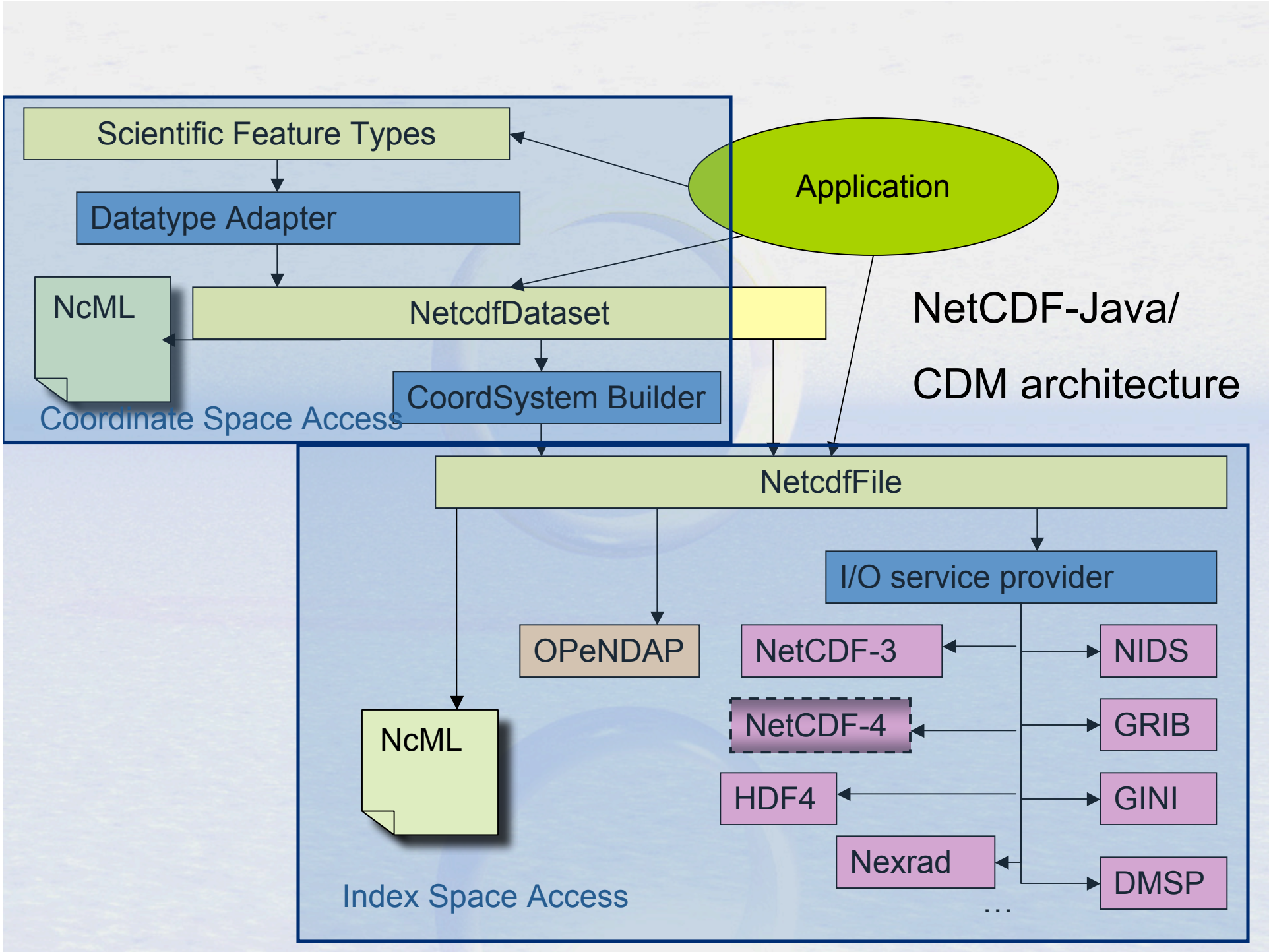
**OPeNDAP URL:**

http://motherlode.ucar.edu:8080/thredds/dodsC/
NAM_CONUS_80km_20081028_1200.grib1.ascii?
Precipitable_water[5][5:1:30][0:1:77]

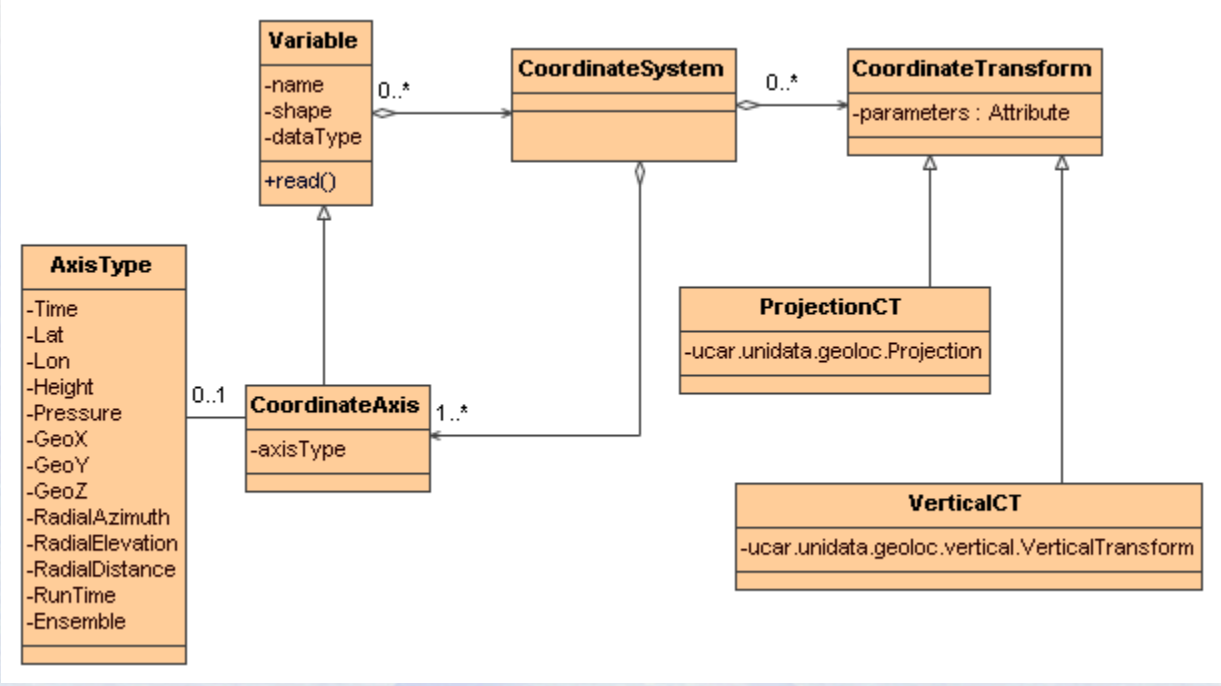
## “Coordinate Space” Data Access:

**NCSS URL:**

http://motherlode.ucar.edu:8080/thredds/ncss/grid/
NAM_CONUS_80km_20081028_1200.grib1?
var=Precipitable_water&
time=2008-10-28T12:00:00Z&
north=40&south=22&west=-110&east=-80


Scientific Feature Types
Datatype Adapter
NcML
NetcdfDataset
CoordSystem Builder
Coordinate Space Access
Application
NetCDF-Java/
CDM architecture
NetcdfFile
I/O service provider
OPeNDAP
NetCDF-3
NIDS
NetCDF-4
GRIB
HDF4
GINI
Nexrad
DMSP
...
NcML
Index Space Access

# Coordinate System UML



# Netcdf-Java Library parses these Conventions

- CF Conventions (preferred)
- COARDS, NCAR-CSM, ATD-Radar, Zebra, GEIF, IRIDL, NUWG, AWIPS, WRF, M3IO, IFPS, ADAS/ARPS, MADIS, Epic, RAF-Nimbus, NSSL National Reflectivity Mosaic, FslWindProfiler, Modis Satellite, Avhrr Satellite, Cosmic, ….
- Write your own *CoordSysBuilder* Java class

# Projections (CF)

- albers_conical_equal_area
- lambert_azimuthal_equal_area
- lambert_conformal_conic
- mcidas_area
- mercator
- orthographic
- rotated_pole
- stereographic (including polar)
- transverse_mercator
- UTM (ellipsoidal)
- vertical_perspective

# Vertical Transforms (CF)

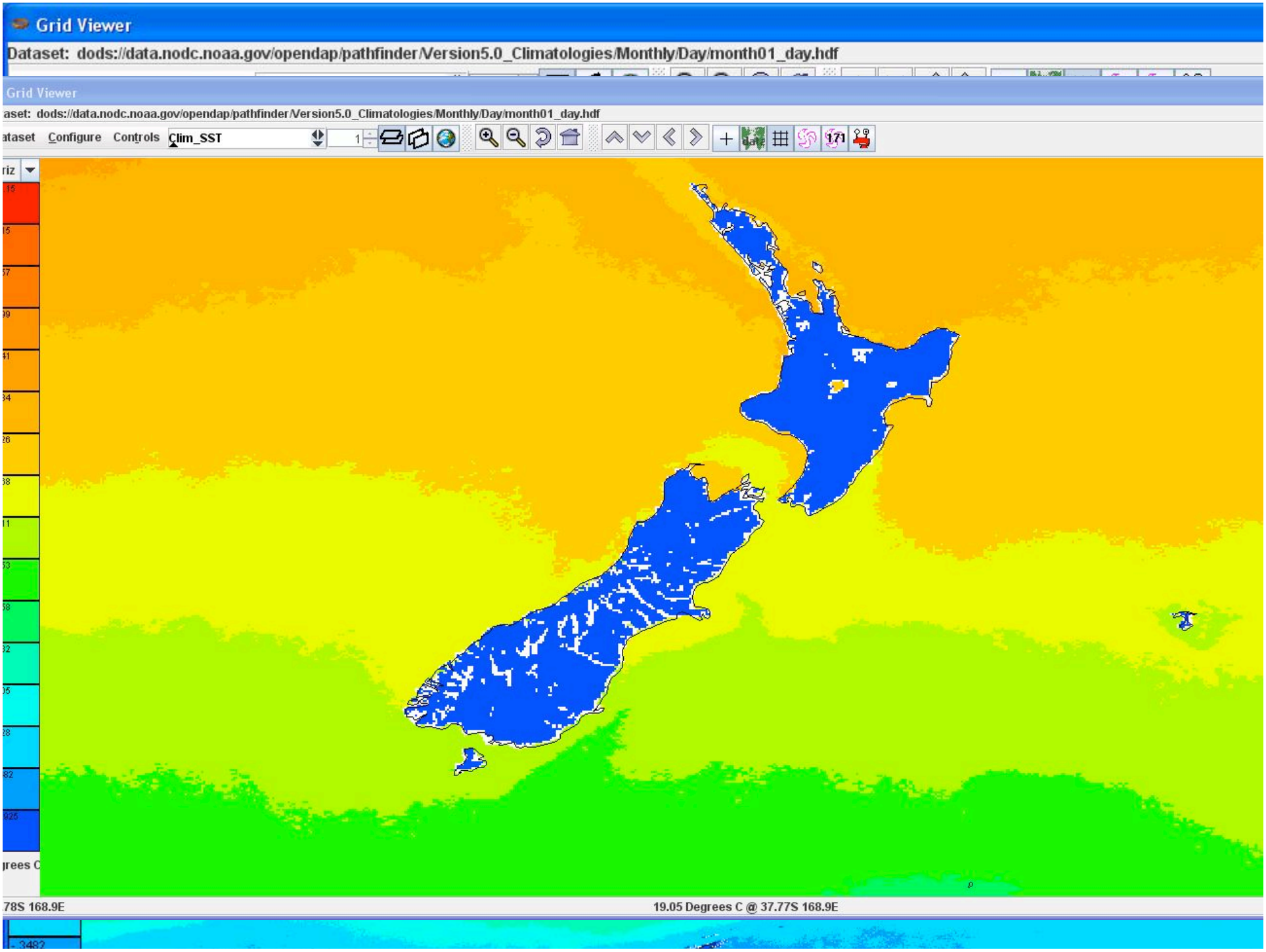
- atmosphere_sigma
- atmosphere_hybrid_sigma_pressure
- atmosphere_hybrid_height
- ocean_s
- ocean_sigma
- existing3DField

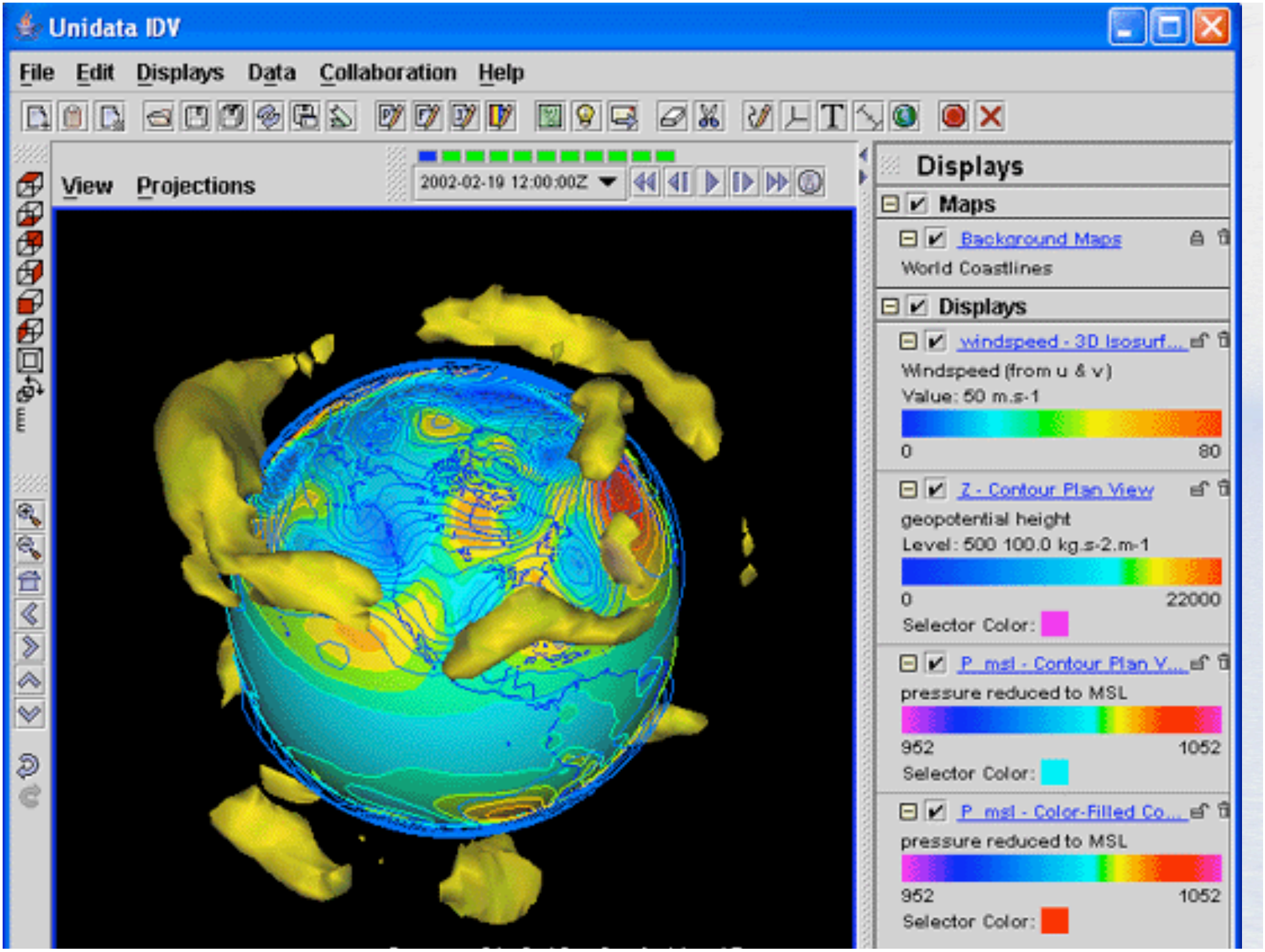
# Add your own Transform

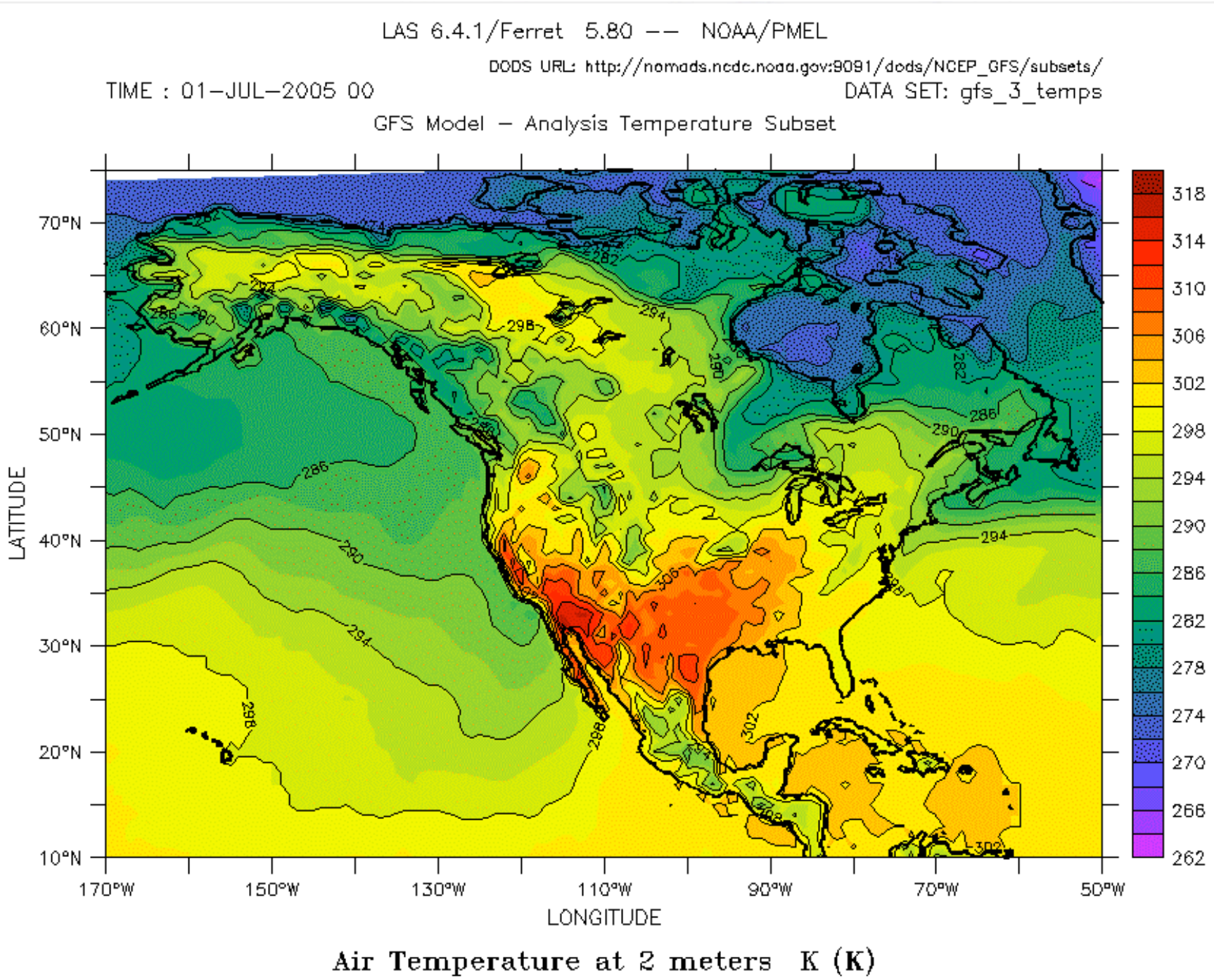
- Pluggable framework
  - Add at runtime
  - CoordTransBuilder.registerTransform()
- Implement *CoordTransBuilderIF*

# Coordinate Systems Summary

- How?
  - Write your own Java code, plug into CDM
  - Write your files using CF Conventions
- Why?
  - Standard visualization, debugging, and data manipulation tools
  - Standard servers to make your data remotely accessible


Grid Viewer
Dataset: dods://data.nodc.noaa.gov/opendap/pathfinder/Version5.0_Climatologies/Monthly/Day/month01_day.hdf
Grid Viewer
aset: dods://data.nodc.noaa.gov/opendap/pathfinder/Version5.0_Climatologies/Monthly/Day/month01_day.hdf
ataset Configure Controls Clim_SST
78S 168.9E
19.05 Degrees C @ 37.77S 168.9E


Unidata IDV
File Edit Displays Data Collaboration Help
View Projections
2002-02-19 12:00:00Z
Displays
Maps
Background Maps
World Coastlines
Displays
windspeed - 3D Isosurf...
Windspeed (from u & v)
Value: 50 m.s-1
0
80
Z - Contour Plan View
geopotential height
Level: 500 100.0 kg.s-2.m-1
0
22000
Selector Color:
P_msl - Contour Plan V...
pressure reduced to MSL
952
1052
Selector Color:
P_msl - Color-Filled Co...
pressure reduced to MSL
952
1052
Selector Color:


LAS 6.4.1/Ferret 5.80 -- NOAA/PMEL
DODS URL: http://nomads.ncdc.noaa.gov:9091/dods/NCEP_GFS/subsets/
TIME : 01-JUL-2005 00
DATA SET: gfs_3_temps
GFS Model - Analysis Temperature Subset
70°N
60°N
50°N
40°N
30°N
20°N
10°N
LATITUDE
170°W
150°W
130°W
110°W
90°W
70°W
50°W
LONGITUDE
318
314
310
306
302
298
294
290
286
282
278
274
270
266
262
Air Temperature at 2 meters K (K)

cGIS Desktop Help 9.2 - About netCDF data - Mozilla Firefox

Edit View History Bookmarks Tools Help

http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?TopicName=About_netCDF_data

ESRI netcdf

kipedia GooMaps gnews 538 mlode motest modev tdsLocal esupport Jira TeamCity Ramadda

Unidata Pro... ArcGI... Free (LGPL) ... OPeNDAP D... riana.si famfamfam.... http:/...ml.dds CXA-81624... Dance Floor... Catalog /thr... Catalog /thr... http:/...

ESRI Support Center | Customer Service | Training | More ESRI Sites...

# ArcGIS 9.2 Desktop Help

Search

search options


lapse all
lcome to ArcGIS Desktop Help
yright information
Dictionary
at's new in ArcGIS
ting started
p projections and coordinate systems
pping and visualization
ting and data compilation
processing
processing tool reference
databases and ArcSDE
a management with ArcCatalog
a support in ArcGIS
An overview of data support in ArcGI
Data formats supported in ArcGIS
CAD data
Coverages
NetCDF: multidimensional, time series
An overview of netCDF data
About netCDF data
The storage of netCDF data
Organizations that use netCD
Adding netCDF data in ArcGI
Understanding netCDF data
Working with netCDF data in ArcG
Raster data
Shapefiles
Tables and attribute information
Terrains
ensions
coding and address management
ear referencing
work analysis
bile GIS
eroperability and standards support


**Show Navigation** | Hide Navigation

**You are here:**
Data support in ArcGIS > NetCDF: multidimensional, time series data > An overview of netCDF data

## About netCDF data

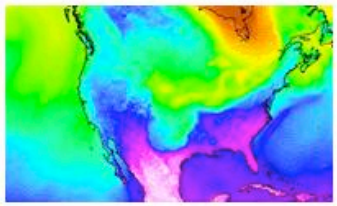
Release 9.2

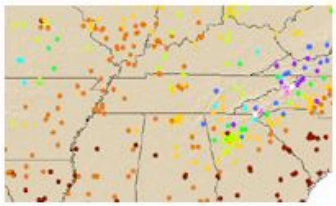
Last modified November 9, 2006

E-mail This Topic | Print This Topic | Give Us Feedback

Print all topics in : "An overview of netCDF data"

NetCDF (network Common Data Form) is a file format for storing multidimensional scientific data (variables) such as temperature, humidity, pressure, wind speed and direction. Each of these variables can be displayed through a dimension (such as time) in ArcGIS by making a layer or table view from the netCDF file.





Learn more about netCDF

Learn more about how to add netCDF data to ArcGIS

### Related Topics

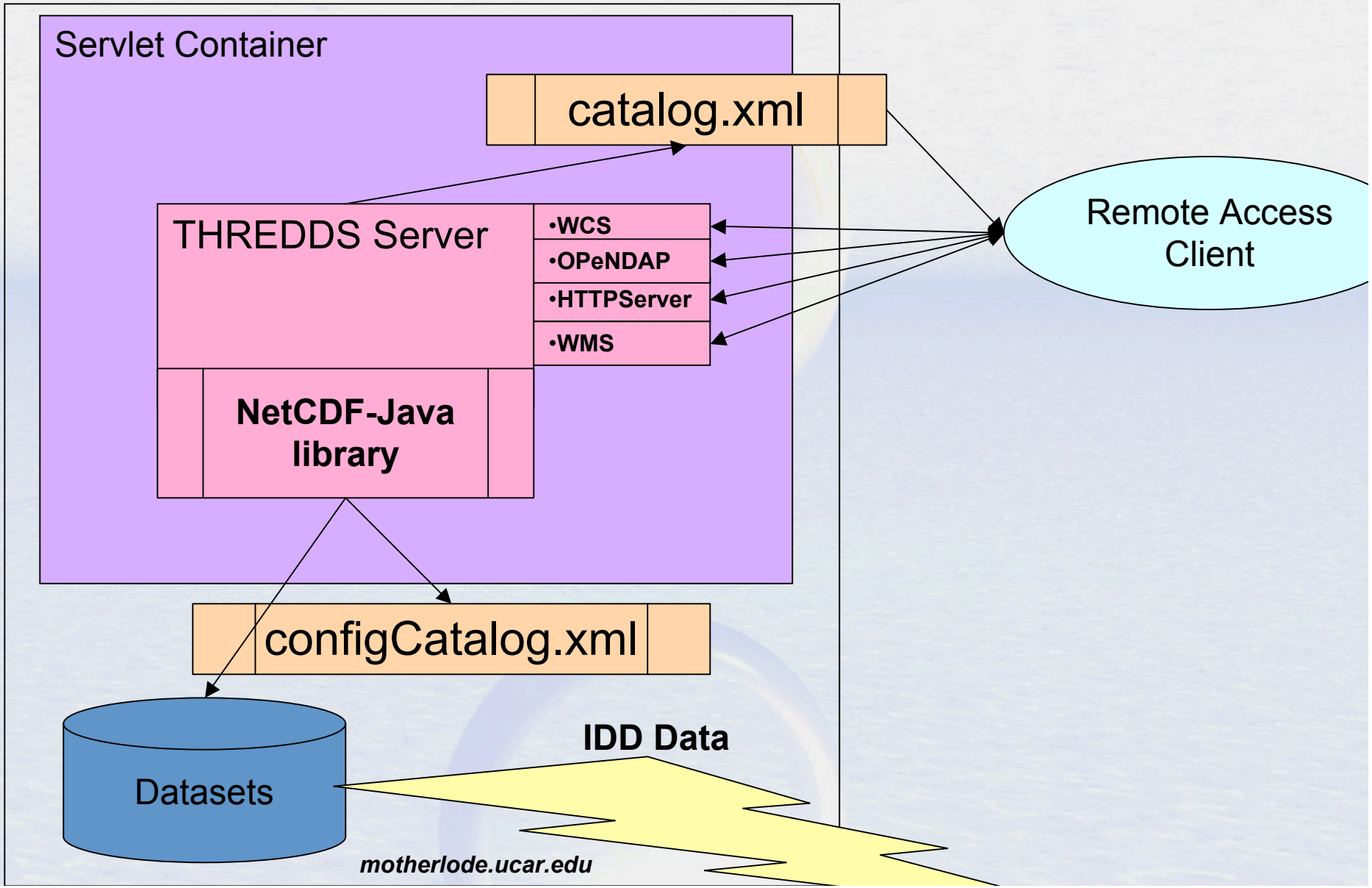
- The storage of netCDF data
- Organizations that use netCDF data
- Adding netCDF data in ArcGIS

Please visit the Feedback page to comment or give suggestions on ArcGIS Desktop Help.

# NetCDF-Java library

- Used as a component in other software (partial)
  - Integrated Data Viewer, ToolsUI (Unidata)
  - Panoply (NASA)
  - ncBrowse (EPIC/NOAA)
  - Java NEXRAD Viewer (NCDC/NOAA)
  - MyWorld GIS (Northwestern)
  - EDC for ArcGIS, ERRDAP (SFSC/NOAA)
  - Live Access Server (PMEL/NOAA)
  - ncWMS (Reading)
  - Matlab plug-in (USGS)

# THREDDS Data Server



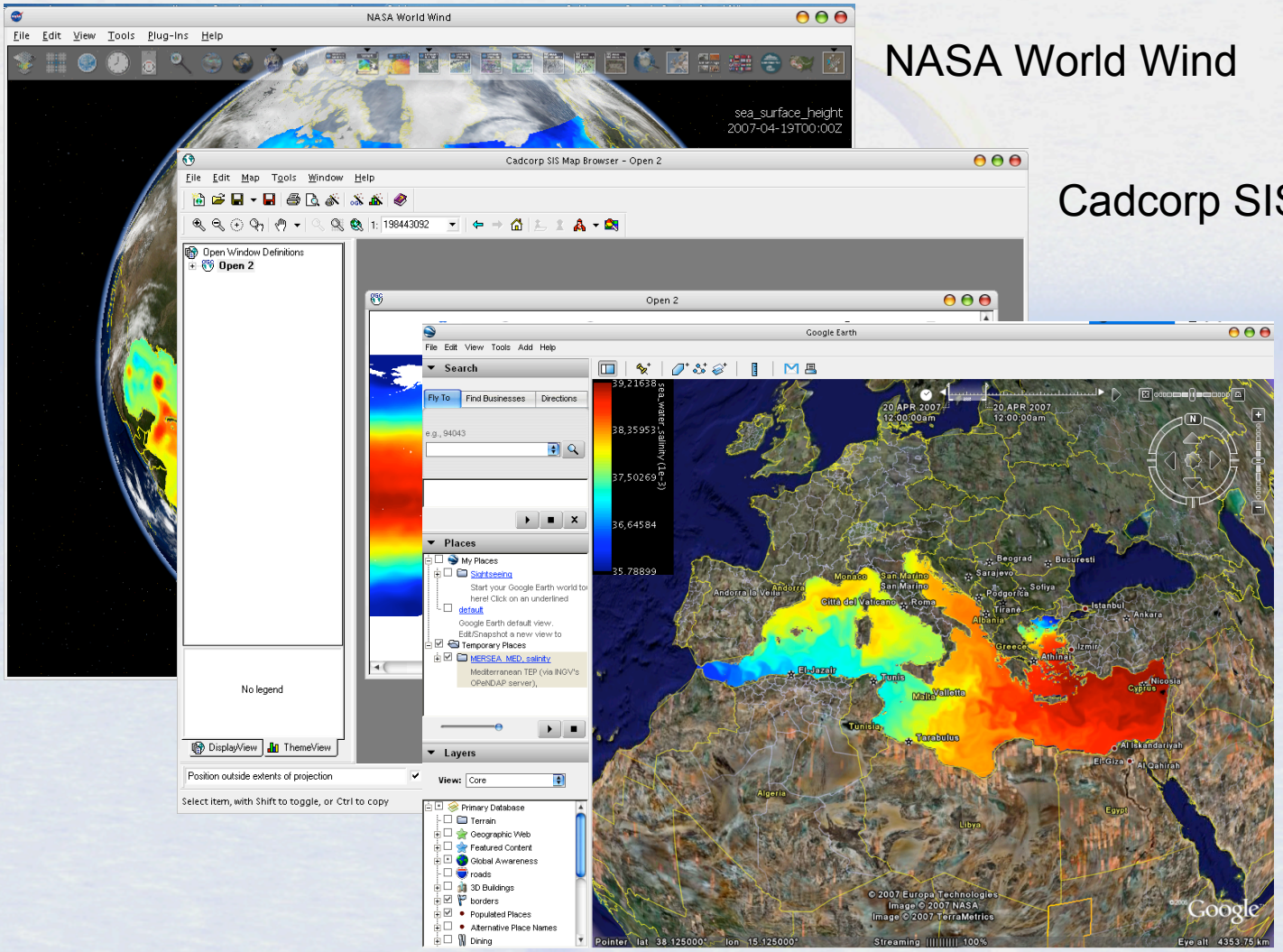
# THREDDS Data Server (TDS)

- Web server for scientific data
- From Unidata
- Provides remote data access
  - OPeNDAP
  - Open Geospatial Consortium (OGC) WMS and WCS
  - HTTP file transfer
  - Experimental data access protocols.

# OGC Web Map Service

- Jon Blower's (Reading, UK) ncWMS integrated with TDS
- Coordinate Space subsetting
- Produces JPEG output
- Fast generation of images
- Reproject images into large number of coordinate systems

# WMS Clients




NASA World Wind

Cadcorp SIS

Google Earth

3rd-party clients can't use the custom WMS extensions

# Web Coverage Service

- Coordinate Space subsetting
- Return formats
  - GeoTIFF floating point, grayscale
  - NetCDF/CF
- No reprojections, resamplings
- Restricted to CDM files that have Grid coordinate system
  - evenly spaced x,y

# NetCDF Markup Language (NcML)

- XML representation of netCDF metadata (like ncdump -h)
- Create new netCDF files (like ncgen)
- Modify (“fix”) existing datasets without rewriting them
- Create virtual datasets as aggregations of multiple existing files.
- Integrated with the TDS

# NcML
# Modify and serve through TDS

```
<dataset name="Polar Orbiter Data" urlPath ="idd/sat/PolarData" >

  <netcdf location="/data/sat/P02393.hdf">
    <attribute name="Conventions" value="CF-1.4"/>
    <variable name="Reflectivity" orgName="R34768">
      <attribute name="units" value="dBZ" />
      <attribute name="coordinates" value="time lat lon" />
    </variable>
  </netcdf>

</dataset>
```

# TDS / NcML
# Modify all files in datasetScan

```
<datasetScan name="Polar Orbiter" path="/data/sat/"
     location= "/data/hdf/polar/">

   <netcdf>
     <attribute name="Conventions" value="CF-1.4"/>
     <variable name="Reflectivity" orgName="R34768">
       <attribute name="units" value="dBZ" />
       <attribute name="coordinates" value="time lat lon" />
     </variable>
   </netcdf>

</datasetScan>
```

# TDS / NcML aggregation

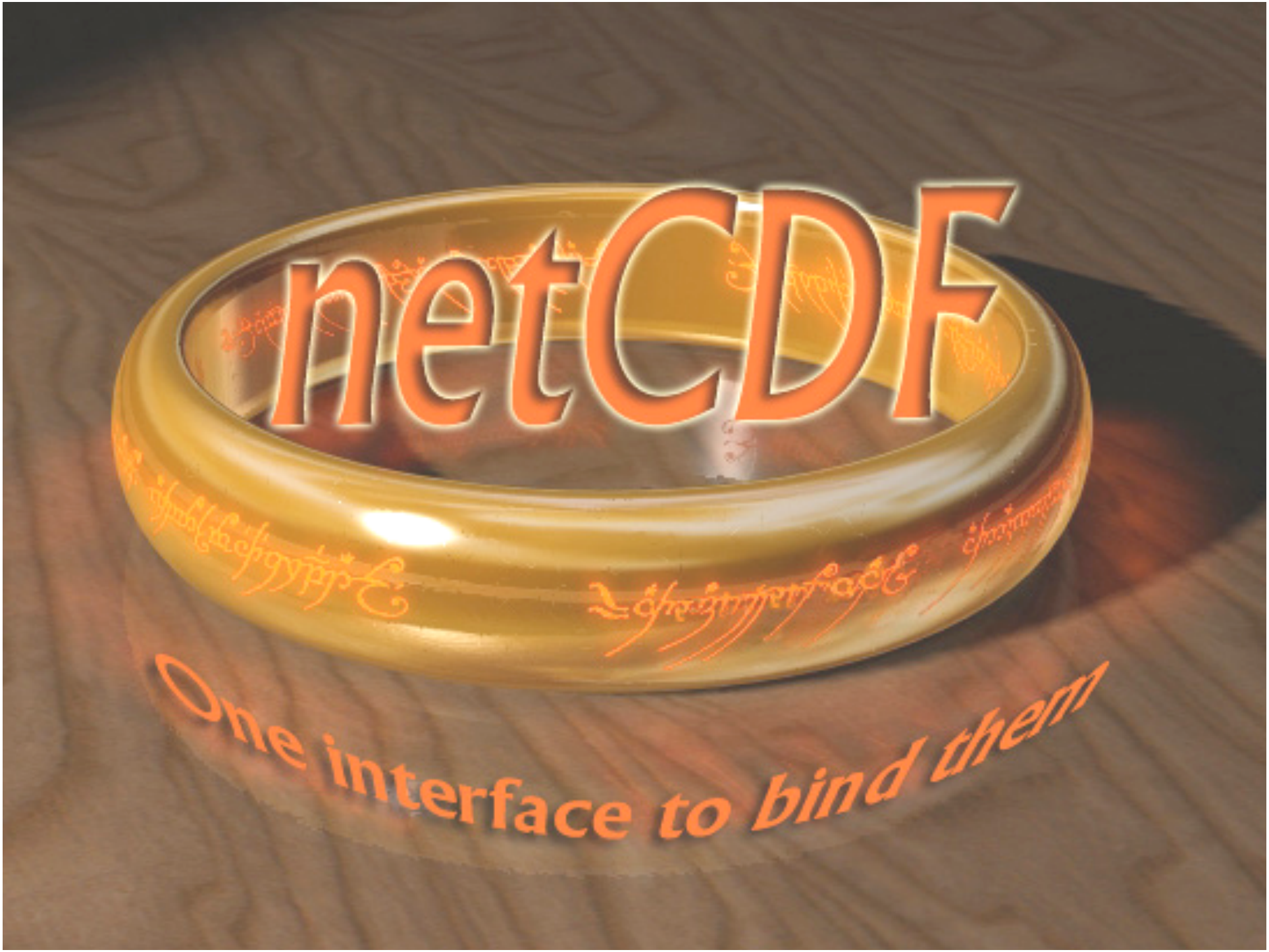
```
<dataset name="WEST-CONUS_4km Aggregation"
    urlPath="satellite/3.9/WEST-CONUS_4km">

  <netcdf>
   <aggregation dimName="time" type="joinExisting">
     <scan  location="/data/satellite/WEST-CONUS_4km/" suffix=".gini"
     />
   </aggregation>
  </netcdf>

</dataset>
```

# Conclusions

- NetCDF is a floor wax and a dessert topping
- A data model is a good way to see the forest through the trees
- We now have a useable merger of netCDF, OPeNDAP, HDF5 technologies
- Add Coordinate information to allow “coordinate space subsetting”
  - NcML/TDS can help
  - But the right way to do this is….

# Conclusion

- I will use CF Conventions
- I will use CF Conventions
- I will use CF Conventions
- I will use CF Conventions
- I will use CF Conventions
- I will use CF Conventions
- I will use CF Conventions
- I will use CF Conventions
- I will use CF Conventions
- I will use CF Conventions


netCDF
One interface to bind them

# NetCDF-Java Common Data Model (Data Access Layer)