The Many Facets of THREDDS

Thematic Real-time Environmental Distributed Data Services

For March 2007
Unidata Policy Committee Meeting

Ben Domenico

unidata

Motivation

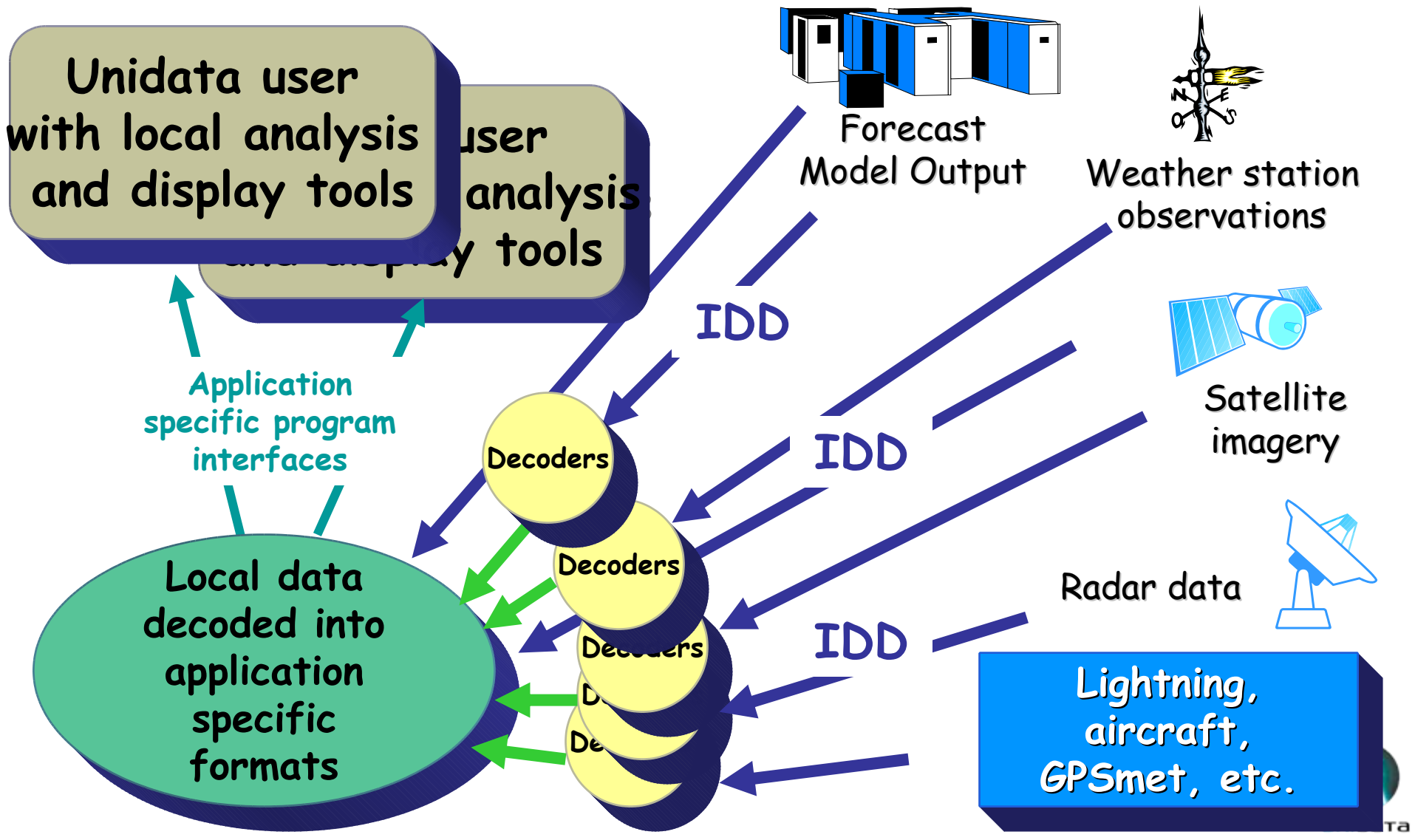
From the ***Unidata 2003*** proposal:

> “utilizing the aggregate data holdings of all Unidata sites as a common resource, accessed via the Internet “(1997).

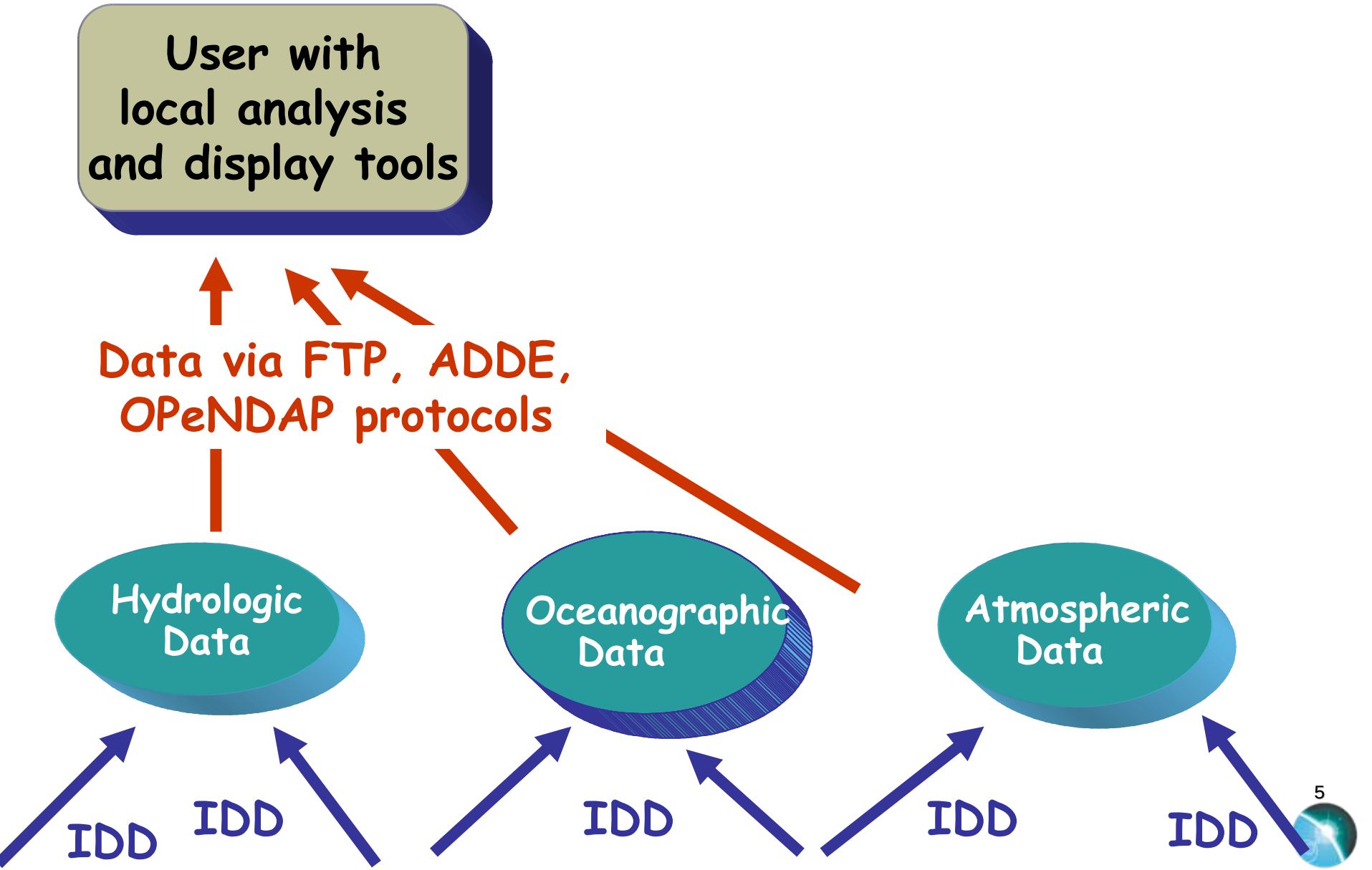
THREDDS Facets Topics

- THREDDS as:
  - A Catalog Service for Remote Data Access via ADDE & DODS/OPeNDAP
  - A Digital Library Collection
  - An Integrated Package for Serving Scientific Metadata and Data
  - A Working Platform for Standards-based Web Services
  - A Highly Collaborative Community Project
- Examples of search systems incorporating THREDDS
- Future options

Typical IDD Data Handling at a Unidata Site



Remote “pull” Data Access

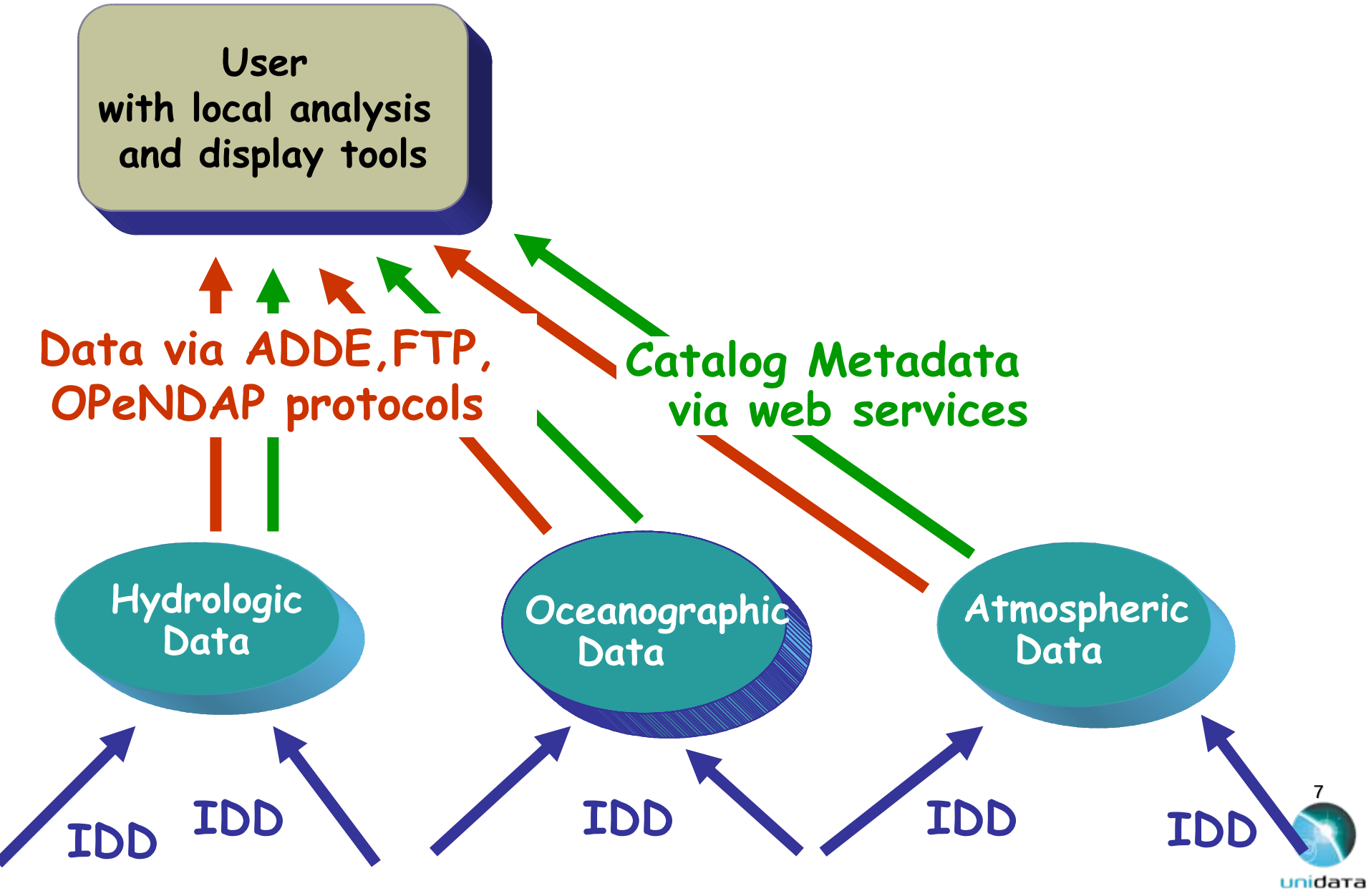


unidata

THREDDS as a Complement to ADDE & OPeNDAP

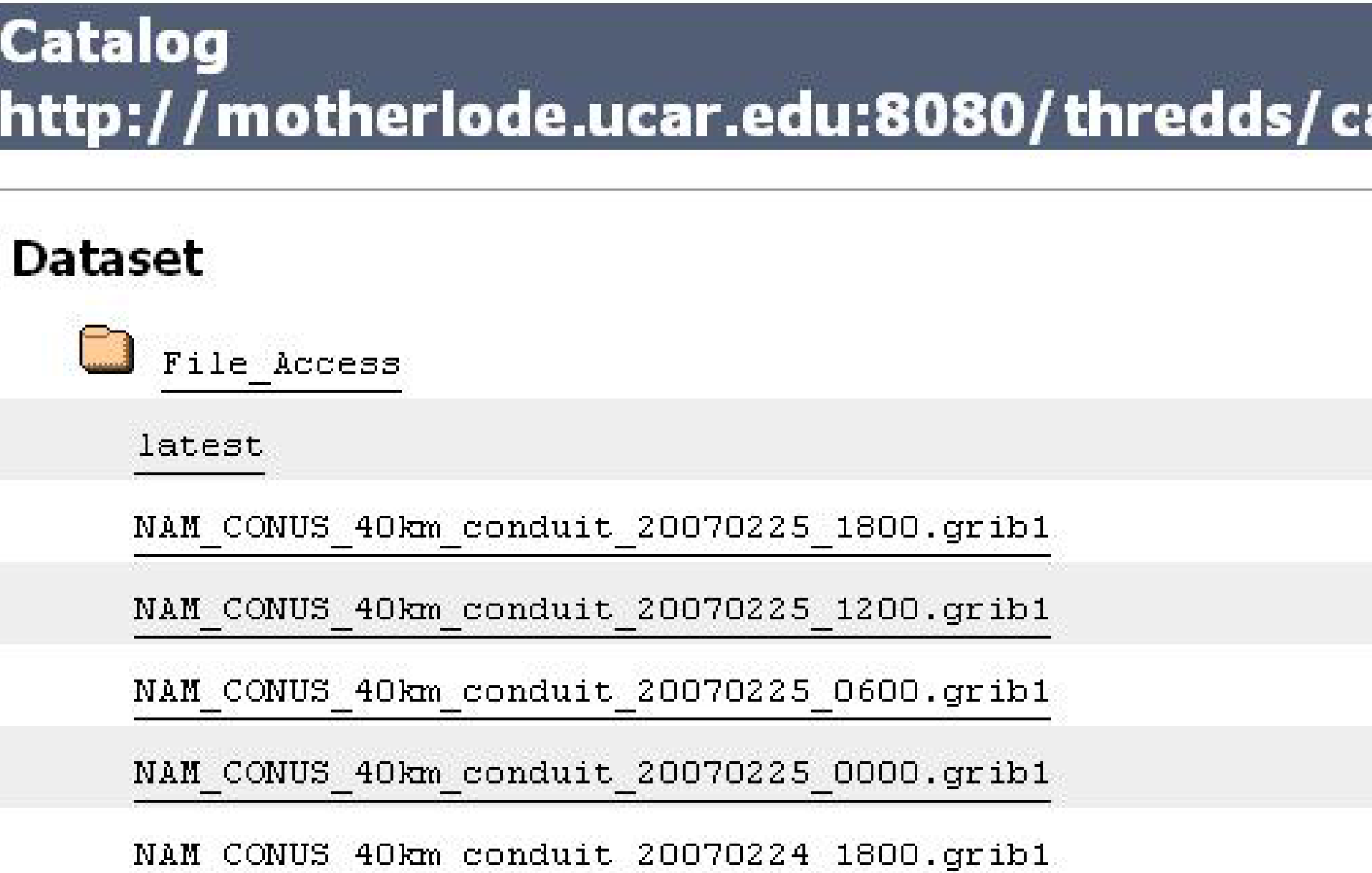
- ADDE and OPeNDAP provide access to remote datasets “as if on local disk”
- Data format transformations (decoding) and subsetting can be done on server side
- THREDDS provides catalogs that can be simple inventory lists of remote datasets
- Need for hierarchical catalogs of catalogs became apparent early on
- End result functions like remote file system

THREDDS Metadata & Remote Data



unidata

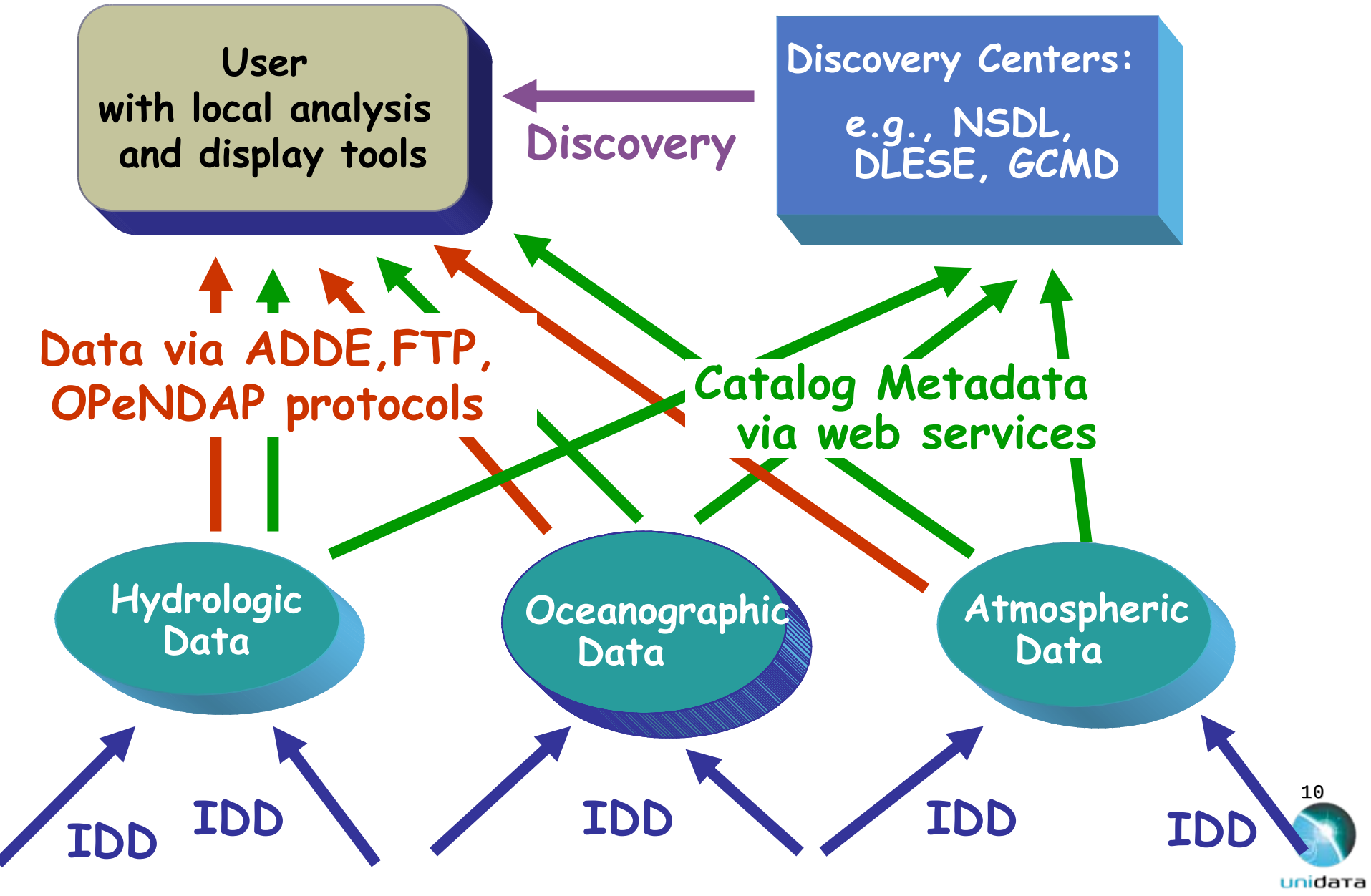
NAM CONUS 40km CONDUIT



THREDDS Catalogs as Collections of Metadata

- Catalogs are more than remote dataset lists:
  - containers for both discovery and use metadata
  - range from simple inventory lists with minimal data descriptions
  - to web documents with embedded interactive data analysis tools
  - mechanism for including datasets into digital libraries and other discovery centers (e.g. GCMD, NSDL, DLESE, NCAR CDP)
- First THREDDS NSDL “collections” grant in 2001

Original THREDDS Concept



unidata

THREDDS Metadata

- **What** property of the Earth, ocean, atmosphere do the numbers in the dataset represent (e.g. temperature, wind speed)?
- **What** are the units of measure?
- **Where** do the measurements or predictions apply?
- **When** were they taken or what time do they forecast?

GCMD “thredds” item

- Platforms/Sources
- Projects
- Geospatial One Stop Projects
- Free text Search
- Portal Collaborations
- Help Center
- Questions?

Data Center

Data Center Name: UCAR/UNIDATA >Unidata, University Corporation for Atmospheric Research
Data Center URL: http://www.unidata.ucar.edu/

Personnel

Name: UNIDATA USER SUPPORT
Email: support at unidata.ucar.edu
Contact Address:
UCAR Office of Programs
Unidata Program Center
P.O. Box 3000
City: Boulder
Province or State: CO
Postal Code: 80307-3000
Country: USA

Related URL

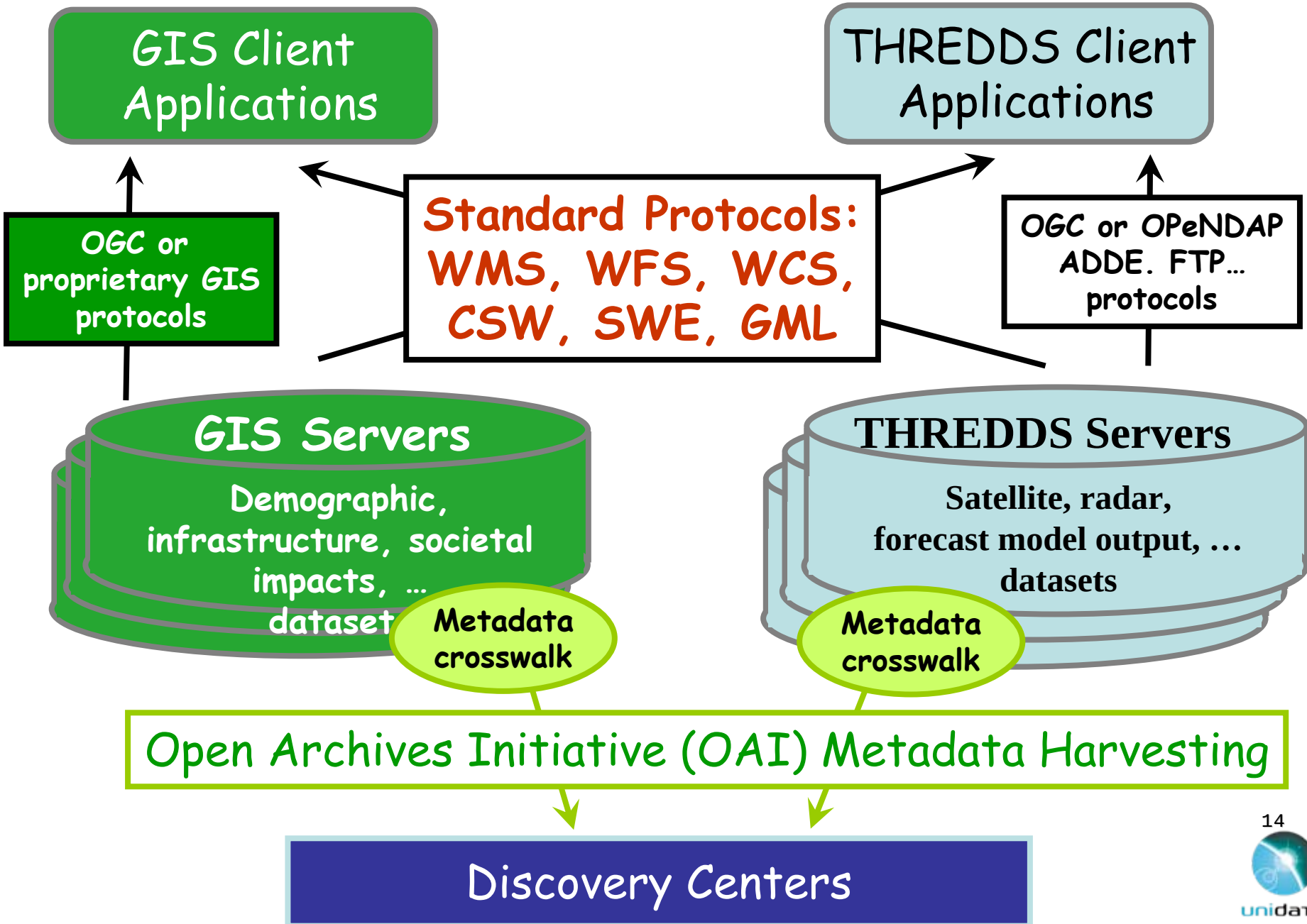
Content Type: GET DATA > **THREDDS** CATALOG
URL: http://motherlode.ucar.edu:8080/thredds/catalog/fmrc/NCEP/DGEX/CONU...

Content Type: GET DATA > **THREDDS** DIRECTORY
URL: http://motherlode.ucar.edu:8080/thredds/catalog/fmrc/NCEP/DGEX/CONU...

Interoperability with Other Disciplines: Different Ways of Thinking about Data

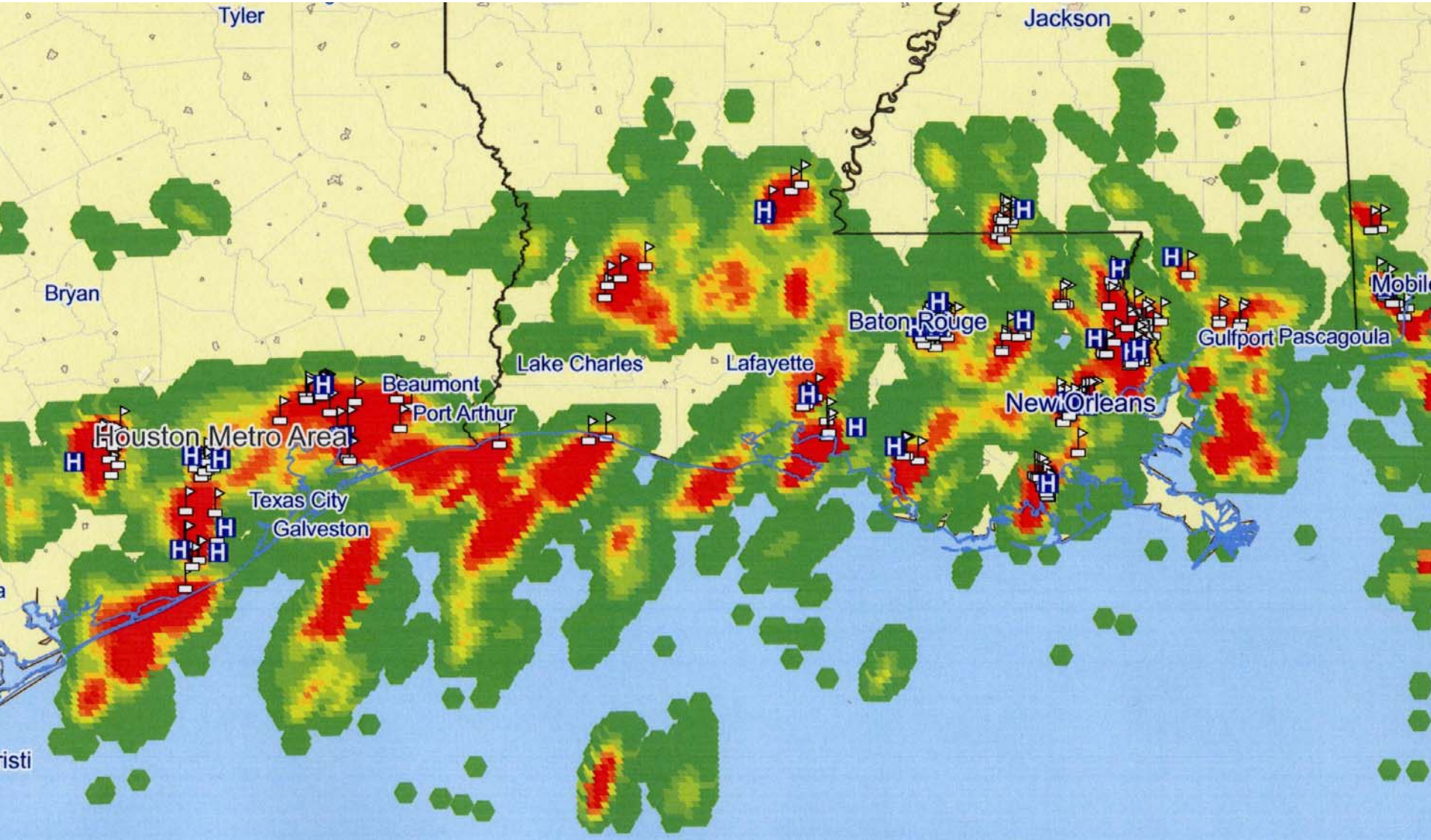
- To the GIS (solid earth, hydrology, and societal impacts) community, the world is:
  - A collection of static *features* (e.g., roads, lakes, plots of land) with geographic footprints on the Earth (surface).
  - The *features* are discrete objects with attributes which can be stored and manipulated conveniently in a **database.**
- To the fluids (atmosphere and oceans) communities, the world is:
  - A set of *parameters* (e.g., pressure, temperature, wind speed) which vary as continuous functions in 3-dimensional space and time.
  - The behavior of the *parameters* in space and time is governed by a set of **equations.**
  - Data are simply discrete points in the mathematical function space.

THREDDS/GIS Interoperability



unidata

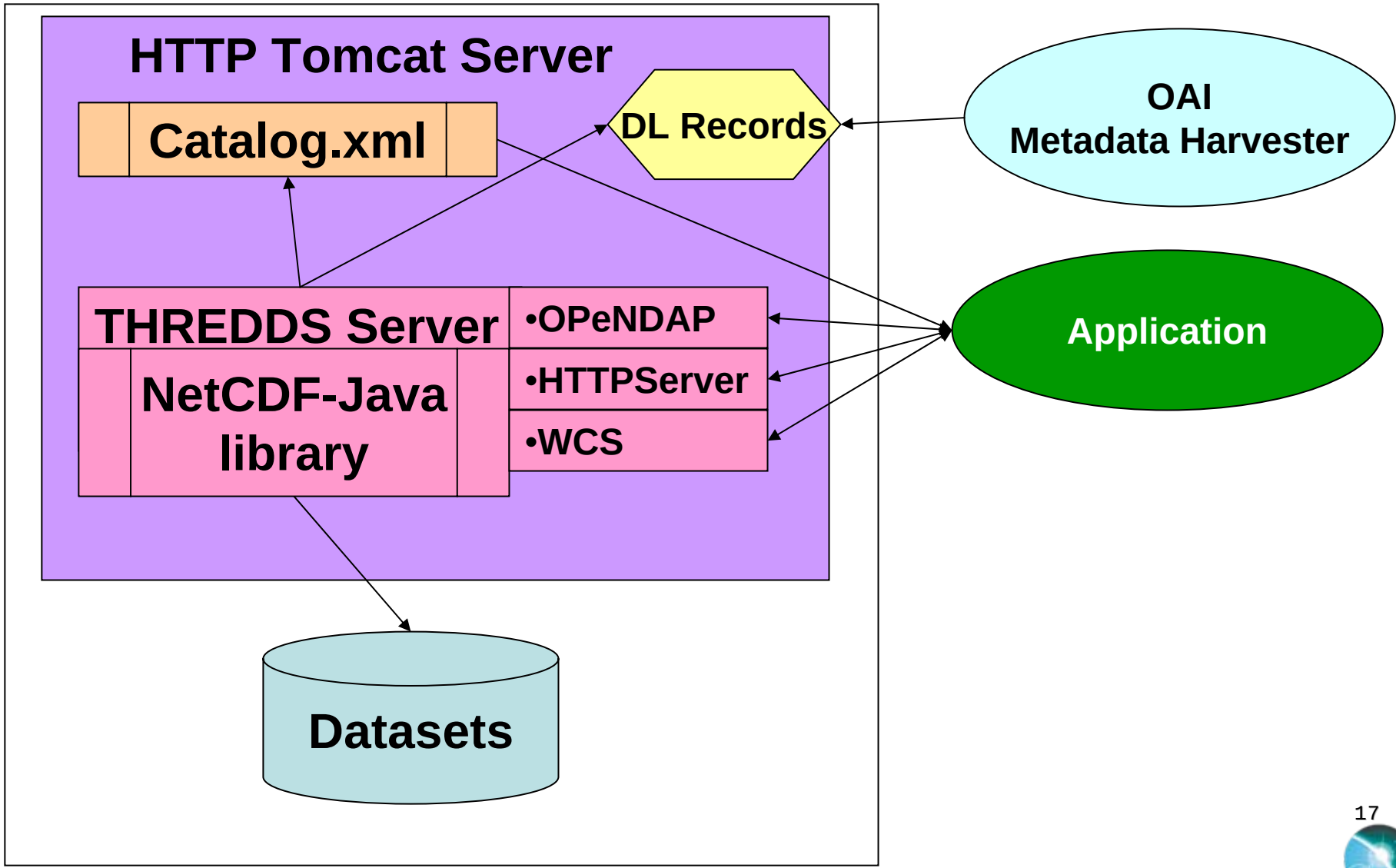
ESRI arcGIS THREDDS Integration:
Schools in LEAD Forecast Rainstorm Area



TDS: THREDDS Data Server

- Integrated package of services
- Data access
  - OPeNDAP
  - Web Coverage Service (WCS)
  - netCDF subset service
- Rudimentary processing
  - Format transformations
  - Subsetting variables, space, time
- THREDDS catalogs
- Open Archives (OAI) harvesting

THREDDS Data Server Architecture



unidata

Common Data Model

- Combine best characteristics of netCDF, OPeNDAP, and HDF data models
- Maintain simple, elegant data access interface of netCDF
- Add useful capabilities from OPeNDAP and HDF such as:
  - Character strings
  - Structures
- Provides common model for data exchange

THREDDS as a Team Effort

- **UPC**
  - coordinates collaborative activities
  - develops and supports catalog generation and service technologies
- **Community** implements servers
  (e.g., several dozen ADDE servers made available as part of Unidata McIDAS distribution)
- **Partners**
  - as data providers, tool builders, interoperability experts from **academia, government, industry**
  - represent **several disciplines**

THREDDS Data Provider Partners

- **University of Alabama Huntsville**
- **ARM** Atmospheric Radiation Measurement)
- **CDC**, the Climate Diagnostic Center
- **COLA**, Center for Oceans Land Atmosphere
- **University of Florence**
- **GMU**, George Mason University
- **IRI/LDEO**, International Research Institute/Lamont Doherty Earth Observatory
- **ESG**, the Earth System GRID (NCAR/SCD)
- **IRIS DMC**, Incorporated Research Institutes for Seismology Data Management Center
- **NCAR**, the National Center for Atmospheric Research
- **NCDC**, the National Climatic Data Center
- **NGDC**, National Geophysical Data Center
- **NOMADS,**NOAA Operational Model Archive and Distribution System,
- **University of Oklahoma**
- **PMEL**, the Pacific Marine Environment Laboratory
- **FNMOC**, Fleet Numerical Meteorological and Oceanographic Center
- **SSEC**, the Space Science and Engineering Center., U. of Wisconsin-Madison
- **Unidata Community ADDE servers**
- ***CIESIN*** Consortium for International Earth Science Information Network)
- ***CUAHSI*** Consortium of Universities for Advancement of Hydrologic Science
- ***ESIG*** NCAR Environmental Societal Impacts Group
- ***Earthscope*** UCAR UNAVCO
- ***GEON*** GEOphysical Network UCSD San Diego Supercomputer Center
- ***ESRI*** GIS Community

THREDDS partners: Analysis/Display Tool Builders

- **Data Discovery Toolkit and Foundry** (New Media Studio).
- **GDS,** GrADS/DODS Server (COLA)
- **IDV**, Integrated Data Viewer (Unidata Program Center)
- **INGRID** (**IRI/LDEO**, International Research Institute/Lamont Doherty Earth Observatory,)
- **LAS**, Live Access Server (PMEL, the Pacific Marine Environment Laboratory)
- **VGEE**, Virtual Geophysical Exploration Environment (NCAR, DLESE, U. of Illinois, Unidata, many collaborators)
- **WXWISE Applets (SSEC**, the Space Science and Engineering Center., U. of Wisconsin-Madison, Tom Whittaker)

- ***ArcGIS*** (ESRI GIS Clients (ESRI, Inc.)
- ***OGC*** Clients (Open GIS Consortium)
- ***MyWorld*** (Northwestern educational GIS Client)

THREDDS Interoperability Partners

- **ADDE,** Abstract Data Distribution Environment (University of Wisconsin – Madison,)
- **DIMES**, DIstributed MEtadata System (George Mason University)
- **DODS/OPeNDAP/Aggregation Server**, Distributed Oceanographic Data System/Open source Project for a Network Data Access Protocol (University of Rhode Island, Unidata)
- **DLESE**,(Digital Library for Earth System Education)
- **ESML**, Earth System Markup Language (University of Alabama-Huntsville)
- **GCMD**, (Global Change Master Directory)
- **OGC and ISO Standards** (University of Florence)

- ***ADL*** (Gazetteer Services The University of California, Santa Barbara)
- ***DLESE Evaluation Services*** (The University of Colorado CIRES)
- ***DLESE Data Services*** (TERC)
- ***DLESE Program Center*** Digital Library for Earth System Education
- **ESRI**
- ***OPeNDAP*** (The University of Rhode Island Open source Project for a Network Data Access Protocol -- formerly DODS)
- ***LAITS*** (Laboratory for Advanced Information Technology and Standards, George Mason University)
- ***NSDL Evaluation Services*** (University of Colorado)
- ***OGC*** (Open GIS Consortium,)
- ***SWEET*** (Semantic Web for Earth and Environmental Terminology)

unidata

THREDDS in Action

- Portal interfaces
  - Unidata Motherlode
  - NCDC
  - PMEL Live Access Server
- Programmatic web services interfaces
  - IDV
  - arcGIS
  - IDL
  - Gi-GO
- Discovery
  - NCAR CDP
  - UAH NOESIS
  - George Mason

Projects Related to THREDDS

http://www.unidata.ucar.edu/projects/THREDDS/GALEON/Reports/RelatedTechnologies.html

- **GALEON/WCS**
- **NASA/GMU ACCESS Geosciences Catalogs CSW (Catalog Services for the Web)**
- **GI-Go (Catalog browser and data access cliient)**
- **netCDF (Java netCDF is basis for THREDDS Data Service)**
- **HDF (HDF5 access via TDS)**
- **ADDE (Client/server protocol, part of McIDAS)**
- **CF (and COARDS) Conventions for netCDF**
- **Live Access Server (LAS) (Incorporates THREDDS catalog system, deployed at hundreds of sites)**
- **Community Data Portal at NCAR**
- **MMI (Marine Metadata Interoperability project)**
- **SWEET (Semantic Web for Earth and Environmental Terminology)**
- **ESIP Federation Service Collaboration Demos (includes TDS/WCS services)**
- **ESRI ArcGIS 9.2 netCDF Interface (works with TDS datasets)**

Summary

- “Seamless” access to historical data at remote sites: *Client/server with ADDE and OPeNDAP*
- Discovery and use of remote data: *THREDDS catalogs*
- Interoperability with GIS data systems: Web Coverage (and other) standard services
- UPC develops core technologies, coordinates team projects, supports integrated packages
- Community deploys
- Partners as data provider, discovery centers, client implementers

Recent (last week) Developments

- Gi-GO catalog metadata and data access client from U of Florence partners
- Standards-based catalog generator and client from George Mason University/NASA ACCESS project partners
- Invitation to participate in ESIP Federation proposal to NSF NSDL data-centri Pathway. (Google likely partner.)

Resulting Policy Issue

Roles of UPC, Unidata community, and partnerships in data search systems

Until now, the UPC has worked with partners for data discovery. Should we become more active?

Future Possibilities

- UPC continues developing infrastructure and capitalizing on partnerships with search experts
- UPC takes more active role in providing search services
- UPC and Unidata community focus more resources on creating metadata content for metadata catalogs

One Approach

- New slot or redirection of staff to establish a “data czar” position with responsibilities to:
- Know what datasets are important, and document their characteristics: size, format, data types, availability, etc.
- Work with community to add metadata to THREDDS catalogs.
- Create a datapedia, and get others in the community to contribute to it.
- Be a resource for developers needing to access the data.

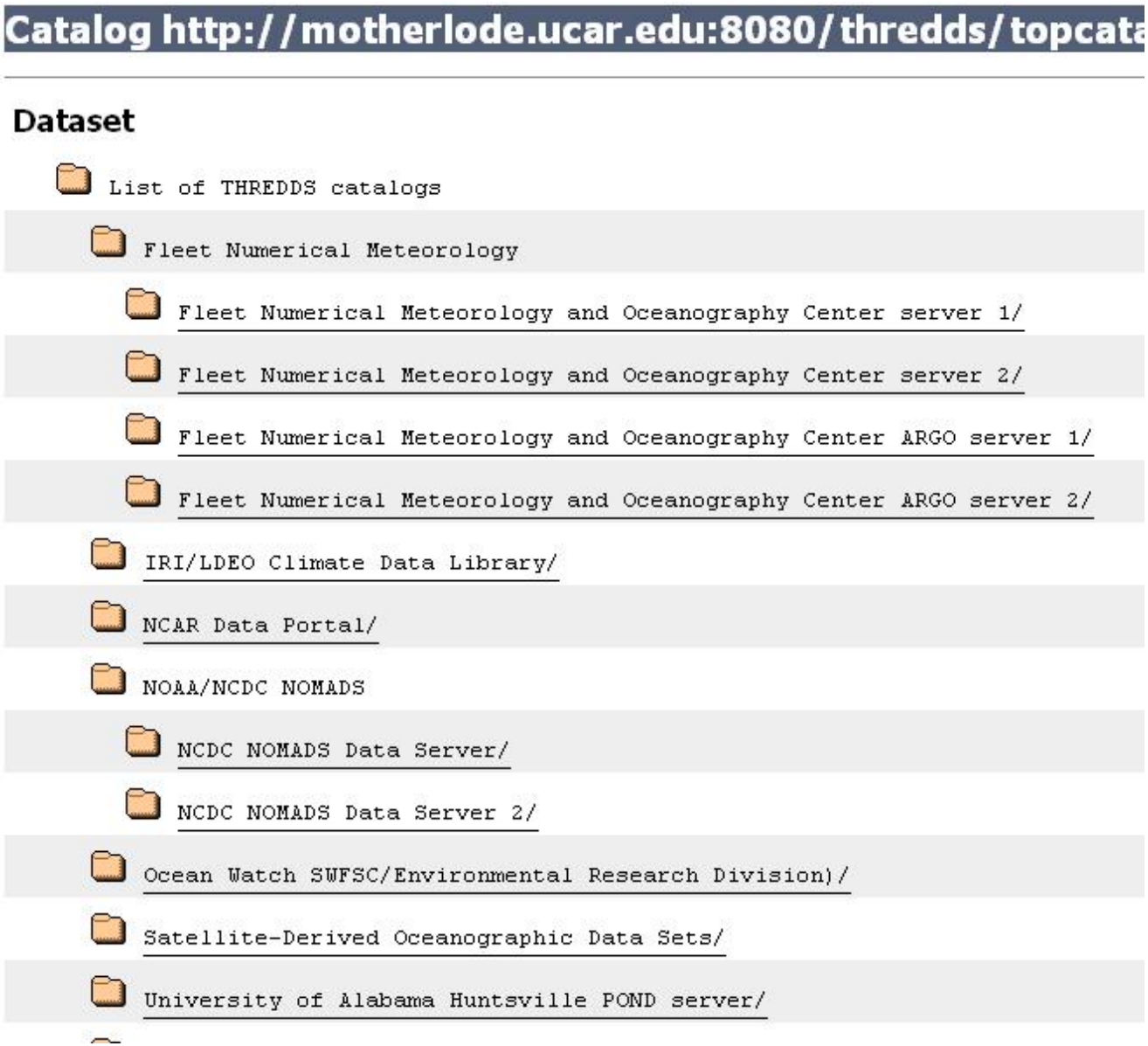
http://cdp.ucar.edu/

When this is actually rolling in a couple years then search systems will be more effective

Direct Access to Examples

- http://motherlode.ucar.edu:8080/thredds/topcatalog.html
- http://nomads.ncdc.noaa.gov:8085/thredds/
- http://motherlode.ucar.edu:8080/thredds/catalog.html
- http://www.unidata.ucar.edu/projects/THREDDS/DataPublications/
- http://www.unidata.ucar.edu/projects/THREDDS/DataPublications/EarlyLEAD/EarlyLEAD.xml
- http://lead4.unidata.ucar.edu:8080/thredds/catalog/model/UCAR/UNIDATA/WRFNMM/catalog.xml
- http://localhost:8080/thredds/catalog.html
- http://noesis.itsc.uah.edu/
- http://gcmd.nasa.gov
- http://cdp.ucar.edu/
- http://zeus.pin.unifi.it/projects/**gi-go**/**gi-go**.jnlp

Motherlode Portal Catalog of Catalogs



unidata

NCDC Server

Catalog http://nomads.ncdc.noaa.gov:9091/dods/thredds

Dataset	Si
NCDC NOMADS GrADS Server 2	
GDAS-BUFR/	
GFDL-CM2.1-DATASETS/	
IGRA_MONTHLY_RAOBS/	
IGRA_STATION_RAOBS/	
NCEP_GFS/	
NCEP_GFS_ANALYSIS/	
NCEP_NAM/	
NCEP_NAM_ANALYSIS/	
NCEP_NARR-A_MONTHLY/	
NCEP_NARR-A_MONTHLY_3hr/	

nidata

NCEP NAM Individual Run

Dataset: File_Access/NAM_CONUS_40km_conduit_20070225_1800.grib1

- *Data format:* GRIB-1
- *Data size:* 455.0 Mbytes
- *Data type:* Grid
- *Naming Authority:* edu.ucar.unidata
- *ID:* fmrc/NCEP/NAM/CONUS_40km/conduit/files/NAM_CONUS_40km_conduit_20070225_1800.grib1

Documentation:

- **summary:** Individual data file, which comprise the Forecast Model Run Collection.
- **summary:** Model runs are made at 00Z, 06Z, 12Z, and 18Z and have analysis and forecasts every 3 hours out to 84 hours.
- **summary:** NCEP North American Model : AWIPS 212 (R) Regional - CONUS - Double Resolution. Horizontal = 185 by 129 points, resolution 40.63 km LambertConformal projection. Vertical = surface, 1000 to 50 hPa pressure levels, layers, and depth.
- **summary:** NCEP Nonhydrostatic Mesoscale Model (NMM) and Gridpoint Statistical Interpolation (GSI) analysis, running in the Weather Research and Forecasting (WRF) infrastructure.
- COMET MetEd (Meteorology Education and Training) documentation
- NCEP Model documentation
- **rights:** Freely available
- **processing_level:** Transmitted through Unidata Internet Data Distribution.
- **processing_level:** Read by CDM Forecast Model Run Collection.

Access:

1. **OPENDAP:** http://motherlode.ucar.edu:8080/thredds/dodsC/fmrc/NCEP/NAM/CONUS_40km/conduit/files/NAM_CONUS_40km_conduit_20070225_1800.grib1
2. **HTTPServer:** http://motherlode.ucar.edu:8080/thredds/fileServer/fmrc/NCEP/NAM/CONUS_40km/conduit/files/NAM_CONUS_40km_conduit_20070225_1800.grib1
3. **WCS:** http://motherlode.ucar.edu:8080/thredds/wcs/fmrc/NCEP/NAM/CONUS_40km/conduit/files/NAM_CONUS_40km_conduit_20070225_1800.grib1
4. **NetcdfServer:** http://motherlode.ucar.edu:8080/thredds/ncServer/fmrc/NCEP/NAM/CONUS_40km/conduit/files/NAM_CONUS_40km_conduit_20070225_1800.grib1

unidata

NetCDF Subset Server

NetCDF Grid Subset Server

Select Grids, optionally bounding box and time range. A NetCDF file using CF-1 Conventions is returned.

Dataset: NAM_CONUS_40km_conduit_20070225_1800.grib1

Base Time: 2007-02-25T18:00:00Z

Available Forecast Hours= (0.0 3.0 6.0 9.0 12.0 15.0 18.0 21.0 24.0 27.0 30.0 33.0 36.0 39.0 42.0 45.0 48.0 51.0 54.0 57.0 60.0 63.0 66.0 69.0 72.0 75.0 78.0 81.0 84.0)

Select Grid(s):

- ☐ Absolute_vorticity
- ☐ Accumulated_snow
- ☐ Albedo
- ☐ Baseflow-groundwater_runoff
- ☐ Best_4-layer_lifted_index
- ☐ Blackadars_mixing_length_scale
- ☐ Brunt-Vaisala_frequency2
- ☐ Categorical_freezing_rain
- ☐ Categorical_freezing_rain_surface
- ☐ Categorical_ice_pellets
- ☐ Categorical_rain
- ☐ Categorical_rain_surface
- ☐ Categorical_snow
- ☐ Clear_sky_upward_solar_flux
- ☐ Cloud_Ice
- ☐ Cloud_Ice_hybrid
- ☐ Cloud_water
- ☐ Cloud_water_hybrid

Bounding Box (decimal degrees):

West Longitude: -153.212
East Longitude: -49.029
North Latitude: 57.381
South Latitude: 11.968

Forecast Hours:

Starting:
Ending:

Add Lat/Lon variables if needed

☐

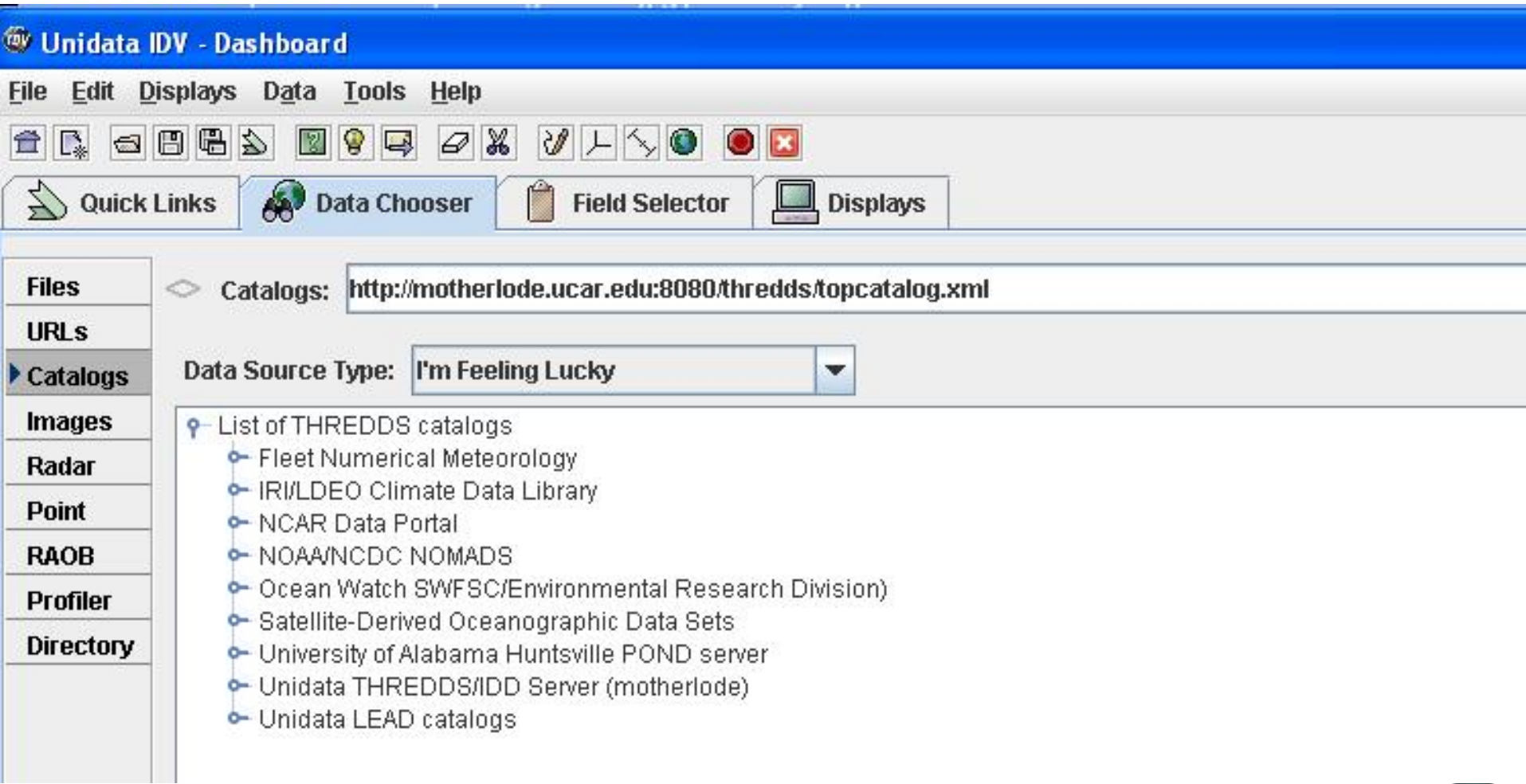
Submit Reset

unidata

Catalog of catalogs in XML (programmatic interface)

```
- <catalog name="THREDDS Top Catalog, points to other THREDDS catalogs" version="1.0.1">
  - <dataset name="List of THREDDS catalogs">
    - <dataset name="Fleet Numerical Meteorology">
        <catalogRef xlink:href="http://usgodae1.usgodae.org/catalog.xml" xlink:title="Fleet Numerical Meteorology and Oceanography Ce
        name=""/>
        <catalogRef xlink:href="http://usgodae2.usgodae.org/catalog.xml" xlink:title="Fleet Numerical Meteorology and Oceanography Ce
        name=""/>
        <catalogRef xlink:href="http://usgodae1.usgodae.org/argo_catalog.xml" xlink:title="Fleet Numerical Meteorology and Oceanograp
        ARGO server 1" name=""/>
        <catalogRef xlink:href="http://usgodae2.usgodae.org/argo_catalog.xml" xlink:title="Fleet Numerical Meteorology and Oceanograp
        ARGO server 2" name=""/>
      </dataset>
      <catalogRef xlink:href="http://iridl.ldeo.columbia.edu/SOURCES/thredds.xml" xlink:title="IRI/LDEO Climate Data Library" name="
      <catalogRef xlink:href="http://dataportal.ucar.edu/metadata/ucar.thredds" xlink:title="NCAR Data Portal" name=""/>
    - <dataset name="NOAA/NCDC NOMADS">
        <catalogRef xlink:href="http://nomads.ncdc.noaa.gov:9090/dods/thredds" xlink:title="NCDC NOMADS Data Server" name=""/>
        <catalogRef xlink:href="http://nomads.ncdc.noaa.gov:9091/dods/thredds" xlink:title="NCDC NOMADS Data Server 2" name="
      </dataset>
      <catalogRef xlink:href="http://oceanwatch.pfeg.noaa.gov/thredds/catalog.xml" xlink:title="Ocean Watch SWFSC/Environmental Res
```

Catalog of catalogs in IDV (Catalog from within a Client)



unidata

TDS-related Search Systems

- George Mason/NASA Catalog Services for the Web collaborative project
- NCAR Community Data Portal
- Global Change Master Directory
- UAH NOESIS Earth system search system

NCAR CDP Data Search

New CDP data portal release: the newest 4.0 release of CDP is based on a completely renovated software infrastructure. Please notify any problems to CDP support

The Community Data Portal (CDP) is a collection of earth science datasets from NCAR, UCAR, UOP, and participating organizations in the following research

easy search | advanced search

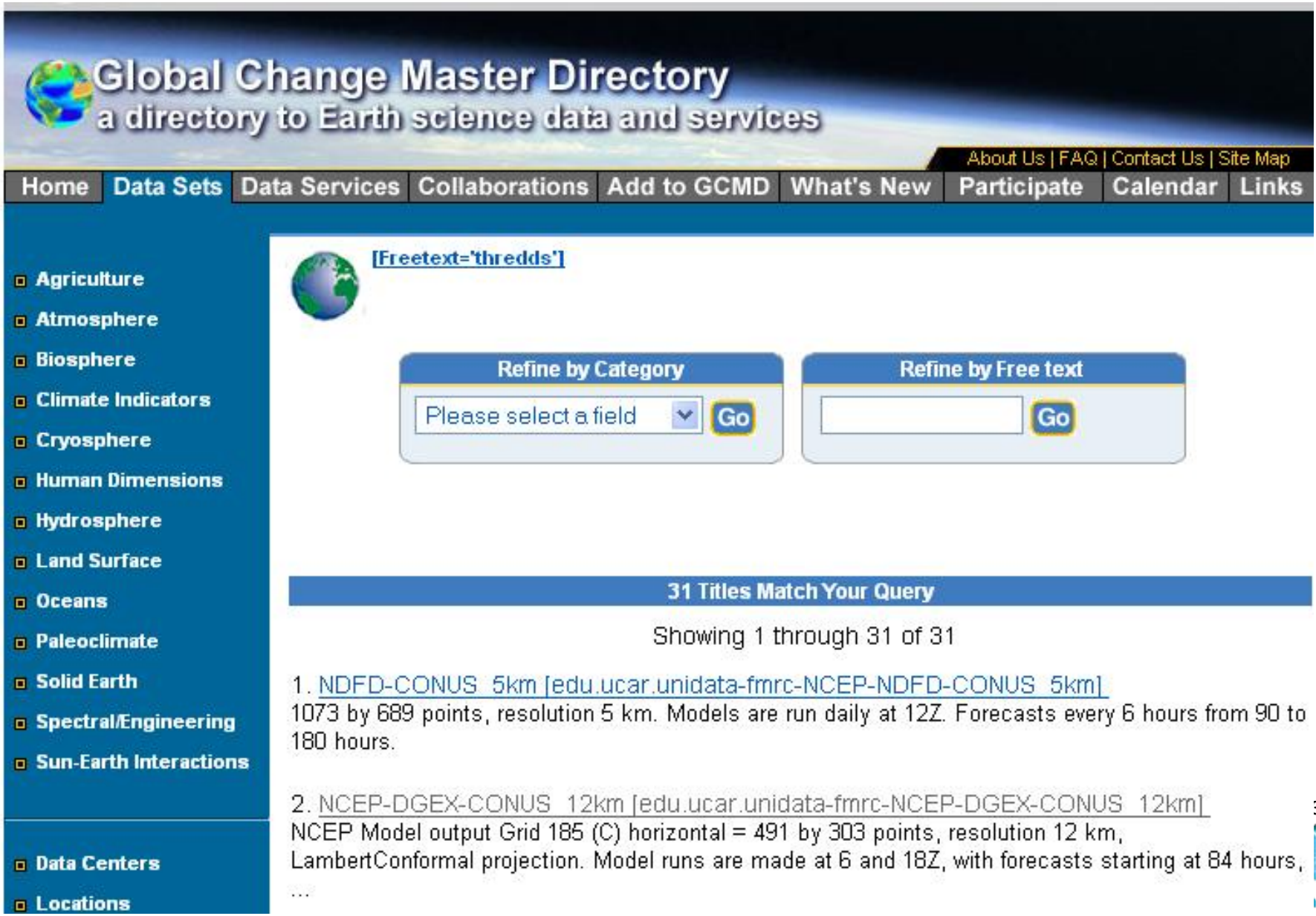
Search for Earth Science datasets by metadata keyword:

Search

browse

- ACD : Atmospheric Chemistry Models, Data Sets and Visualization Tools
  - MILAGRO : Milagro field campaign
    - Requires CDP login and Milagro membership (Click here to request access)
- CGD : Climate and Global Dynamics Division
- CME : Carbon in the Mountains Experiment
- COLA : Center for Ocean-Land-Atmosphere studies
- CU/CIRES/ENLIL : Heliospheric Model
- ECMWF : ERA40 Datasets
  - Requires ECMWF membership

GCMD Search for “thredds”



UAH NOESIS Earth System Search

unidata