Unidata and Community Broadening: One User at a Time

HDF and HDF-EOS Workshop
Aurora, CO

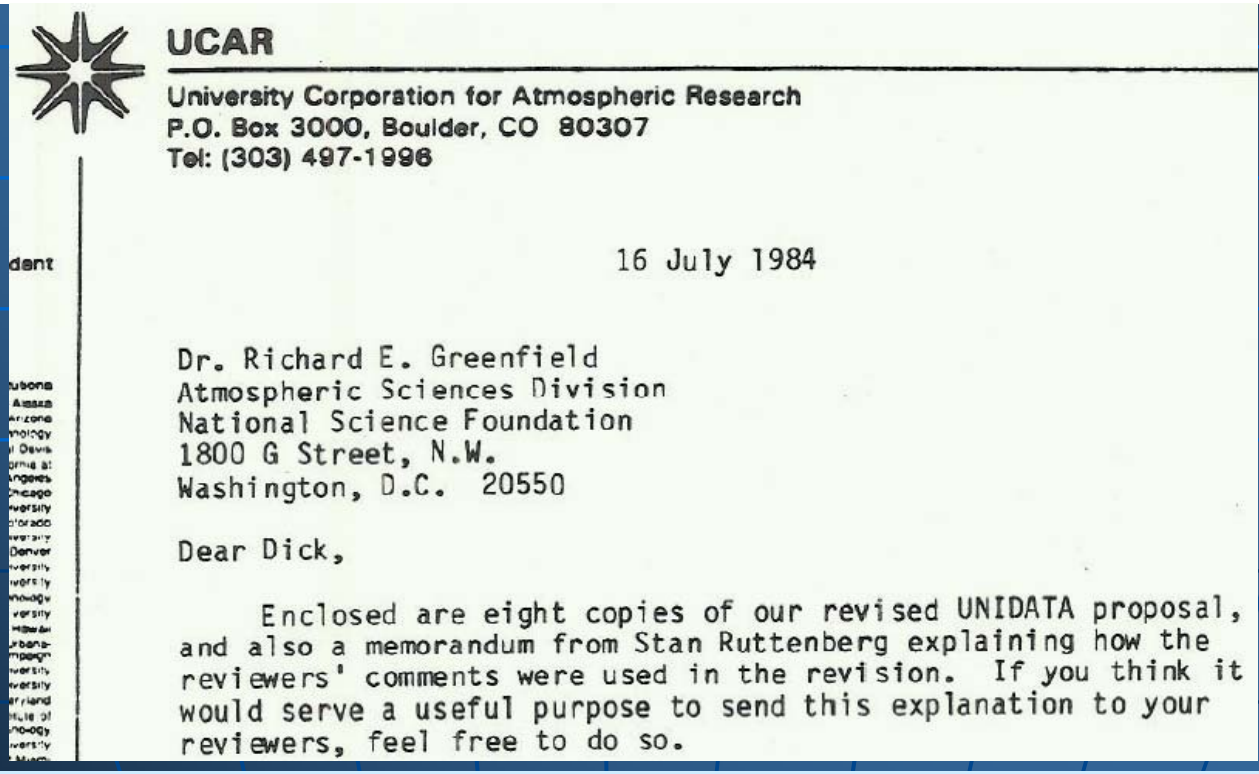
15 October 2008

Dr. Mohan Ramamurthy
Director, Unidata
University Corporation for Atmospheric Research
Boulder, CO

Unidata: Created Through A Grass-roots Effort

Conceived at a 1983 grass-roots workshop;

Funded primarily by NSF/ATM UCAR and Lower Atmospheric Facilities Section;



UCAR

University Corporation for Atmospheric Research
P.O. Box 3000, Boulder, CO 80307
Tel: (303) 497-1996

16 July 1984

Dr. Richard E. Greenfield
Atmospheric Sciences Division
National Science Foundation
1800 G Street, N.W.
Washington, D.C. 20550

Dear Dick,

Enclosed are eight copies of our revised UNIDATA proposal, and also a memorandum from Stan Ruttenberg explaining how the reviewers' comments were used in the revision. If you think it would serve a useful purpose to send this explanation to your reviewers, feel free to do so.

A Program of the community, by the community, and for the community

Unidata: A Benevolent For Academia

- We don’t do science, but we empower scientists
- We don’t teach students, but we facilitate education and learning
- And so on…

Principal Drivers Shaping Unidata's Work

- Science
- Education
- Technology
- Social & Organizational Evolution

What are the community needs w.r.t. data services?

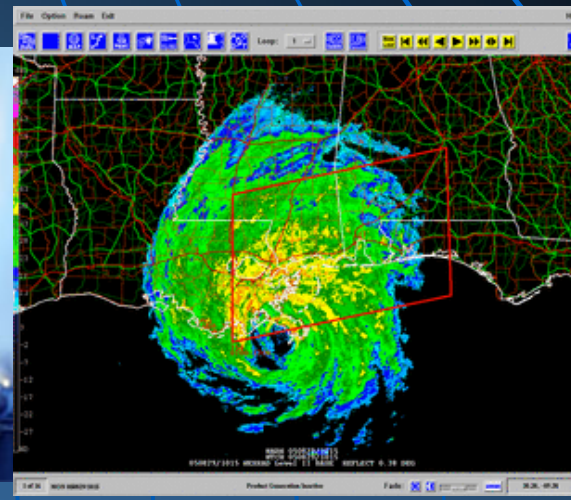
Focus on Unidata's niche and core competencies

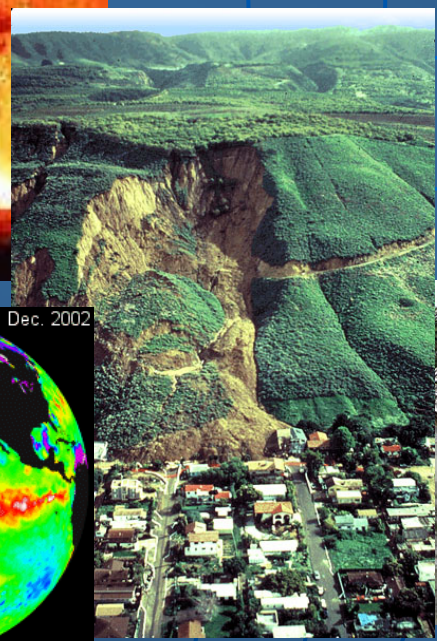
Science Drivers

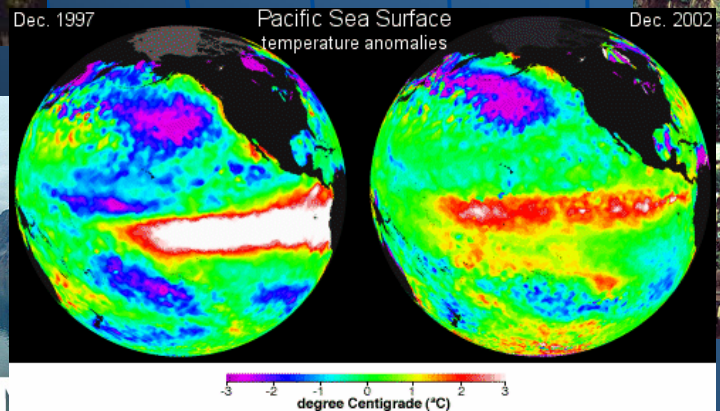
- Environmental problems like global change & water cycle transcend disciplinary as well as geographic boundaries, requiring multidisciplinary approaches and global teams for solving them;
- Rapid advances in observational technologies, especially in remote sensing;
- Increasing use of complex, coupled modeling systems;

Research studies on societal impact of hurricane-related flooding involve integrating data from atmospheric sciences, oceanography, hydrology, geology, geography, and social sciences.

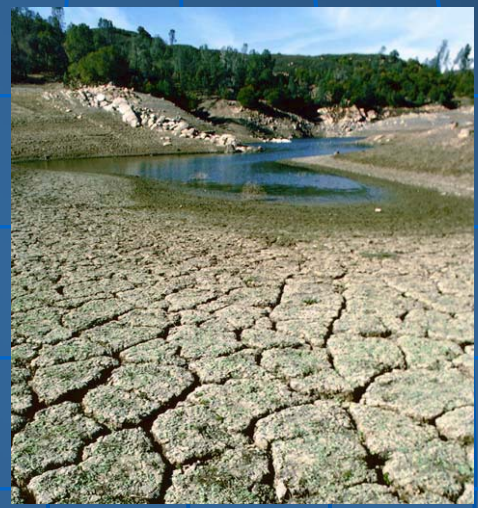
Science Drivers: Examples







Societal Impact

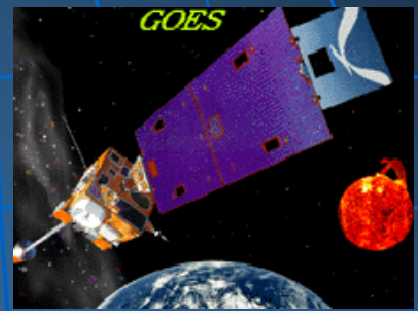


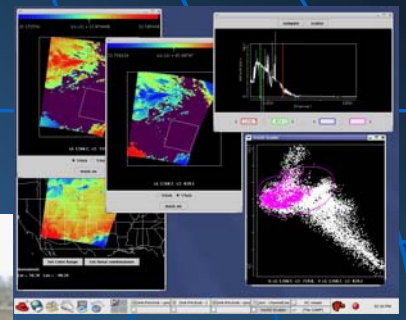
Some of the most challenging problems in the geosciences are at the interface of disciplines

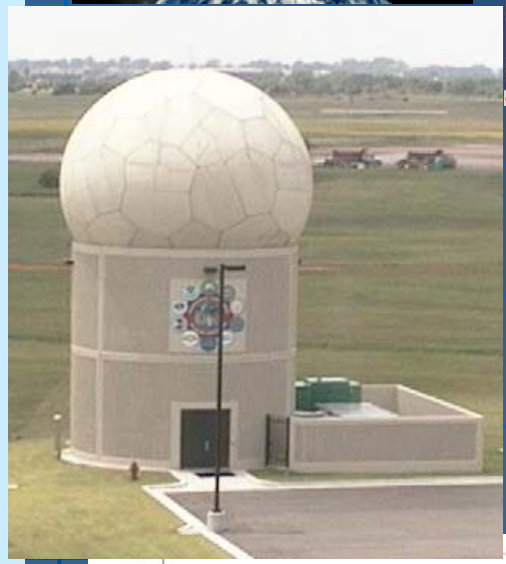
About 2/3rd of Unidata sites have users outside of the atmospheric sciences.

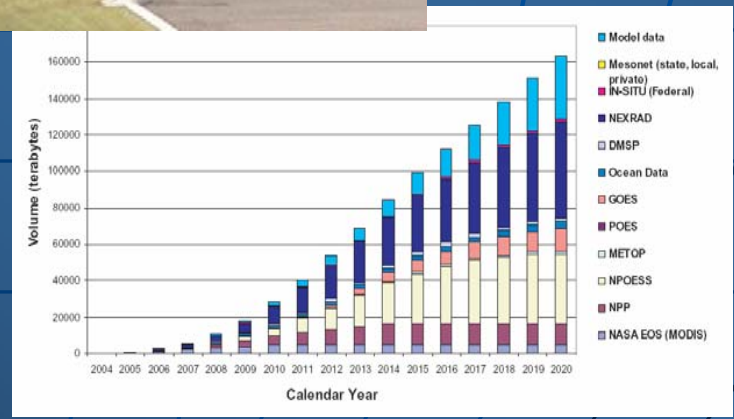
The Impending Data Deluge

- GOES-R (2014)
  - *Hyperspectral Environmental Suite (~1600 channels)*
- NPOESS (2012)
  - Raw data rate: 3 terabytes per day
- Both NPOESS and GOES-R will have data rates 30-60 times the current rates
- Global, coupled models at a grid spacing of 1-5 km, integrated for multi-decades
- Phased Array Radar, with 20 to 30-second volume scan rates, compared with 5-7 minutes with the WSR-88D
- NCAR is working on a Global WRF model for use in both Weather and Climate research

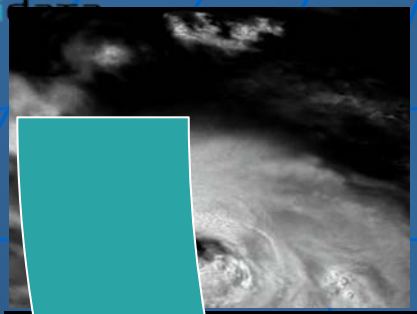


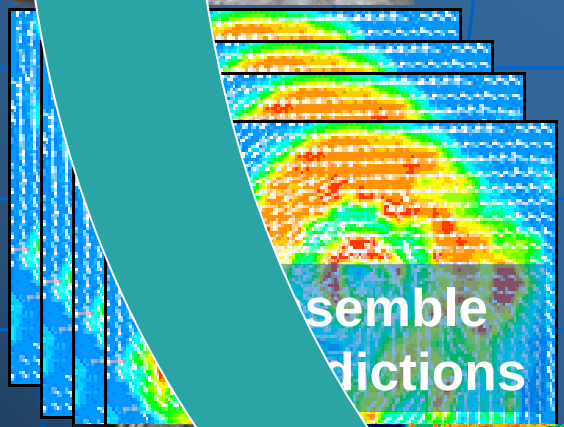


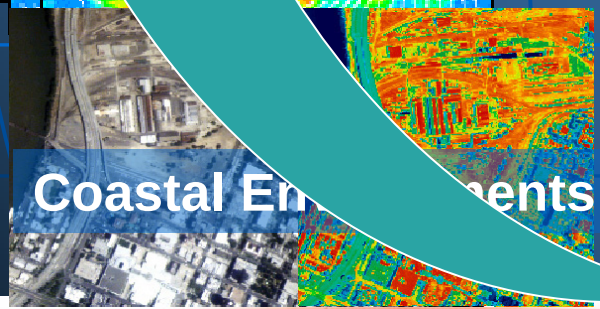


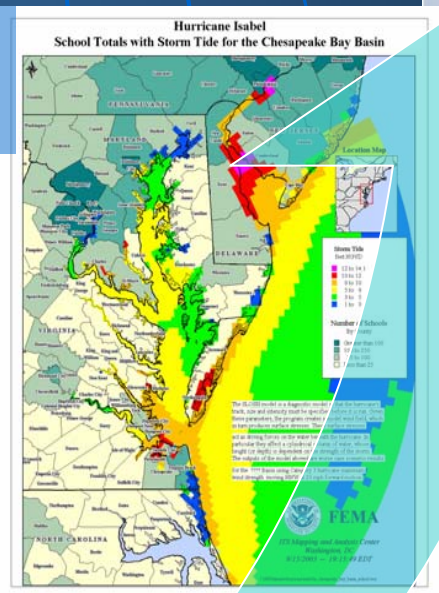


End to End Data and Information Services









GIS Integration

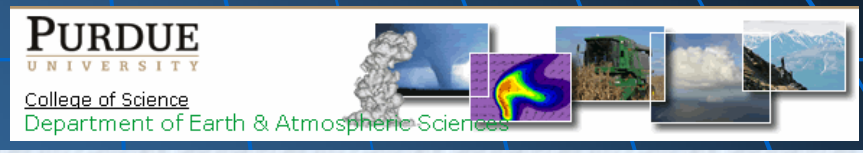
semble
dictions

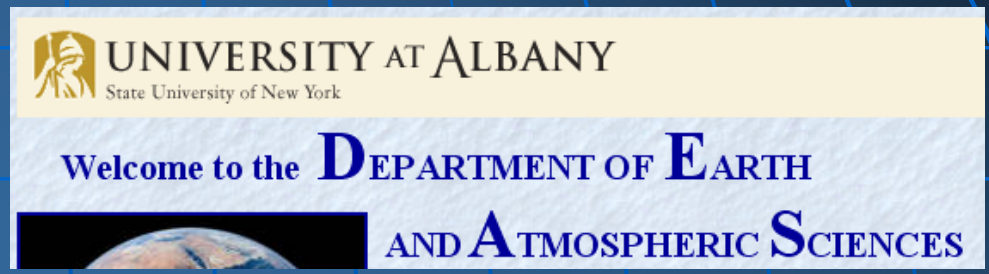
Coastal En ents

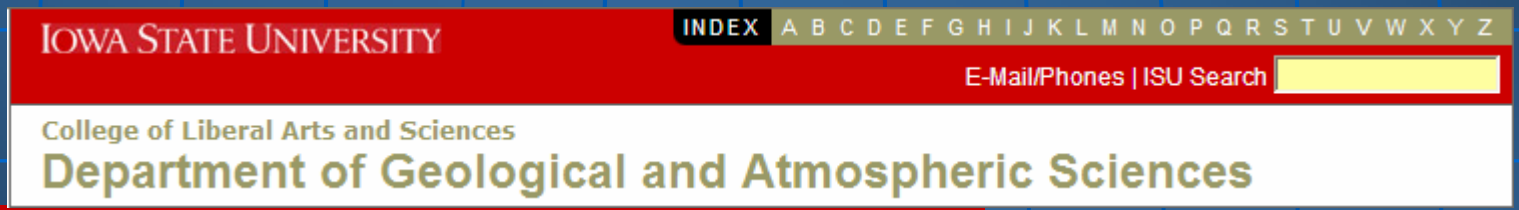
Emergency Response

Need integrated services

Evolution in Academia







These days, few traditional meteorology programs exist. **There are notable exceptions (OU, FSU, and Penn State).**

Mission

To provide the data services, tools, and cyberinfrastructure leadership that advance Earth system science, enhance educational opportunities, and broaden participation.

We

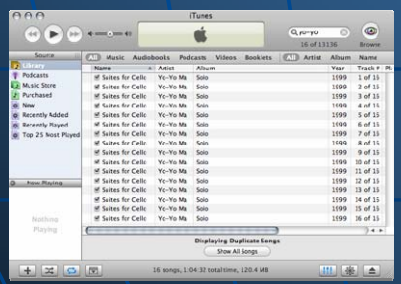
- ***Facilitate data access***
- ***Provide tools for data access, management, and analysis and visualization***
- ***Provide comprehensive support***
- ***Engage in community building and advocacy***

Vision

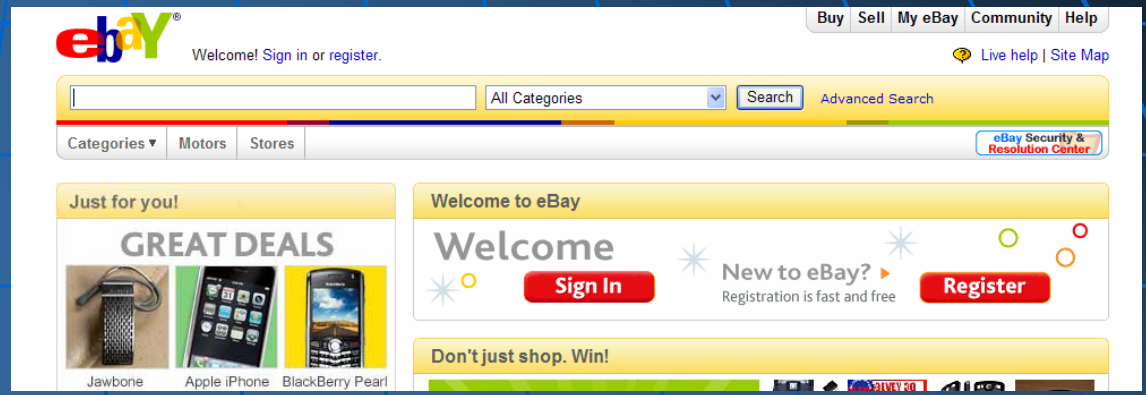
Unidata's vision calls for providing comprehensive, well-integrated, and end-to-end data services.

These include an array of functions for collecting, finding, and accessing data; data/content management tools for generating, cataloging, and exchanging metadata; and submitting/publishing, sharing, analyzing, visualizing, and integrating data.

When this vision is realized, users — no matter where they are, how they are connected to the Internet, or what computing device they use — will be able to find and access a plethora of geosciences data, experience how all of the aforementioned services work together, and use Unidata-provided tools and services both productively and creatively in their research, education, and outreach activities.



E-Commerce Analogs



Facilitate data providers as well as users; empower users to share/publish their data

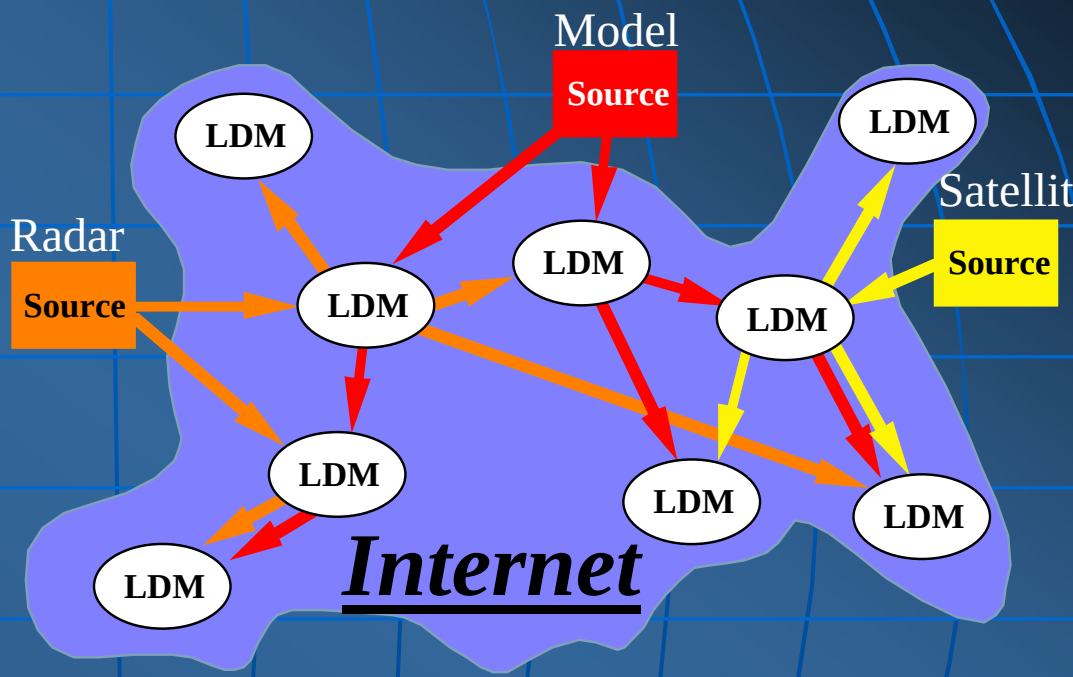
Unidata in a Nutshell

The principal focus of our activities is on real-time weather data provision. In addition, Unidata develops tools, middleware, and services that contribute to broader cyberinfrastructure needs of the geosciences community.

Our most widely used software, netCDF, is used in over 70 countries and it has been incorporated into 50 open source software packages and 15 commercial packages.

Internet Data Distribution

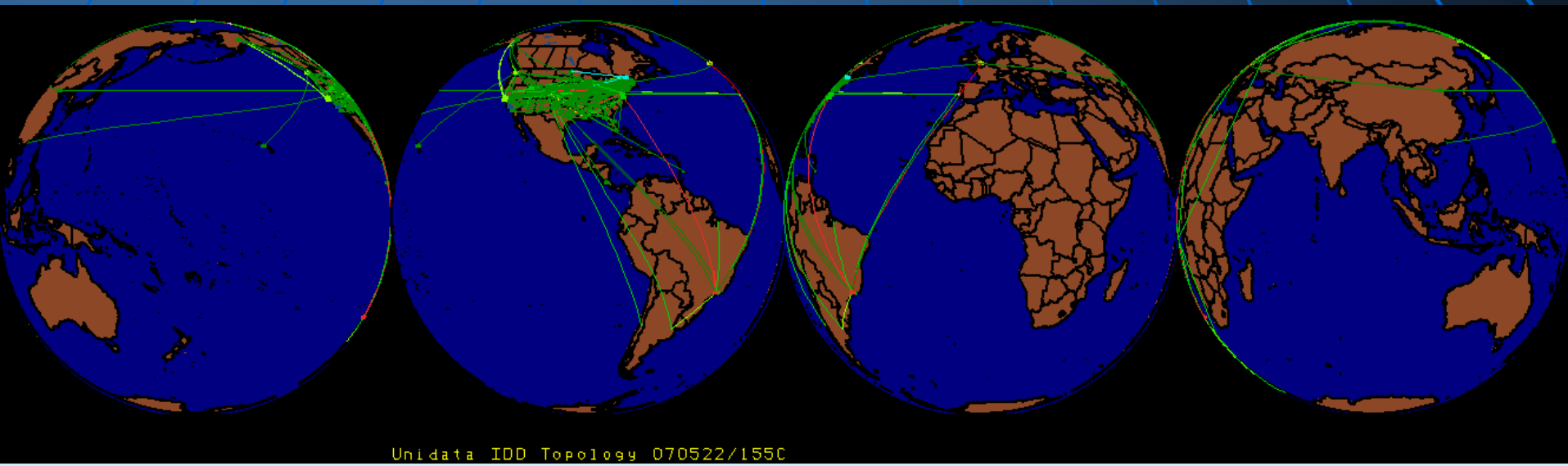
- Over 260 sites worldwide are participating in Unidata Internet Data Distribution (IDD) system
- Data ingested is about 12-13 GB/hour, but in bursts it is up to 23 GB/hour
- The LDM uses more of the Internet2 bandwidth than any other *advanced application (18-20 TB/week)*



The LDM is now ranked #3 (behind HTTP and SSH) in Internet 2 usage.

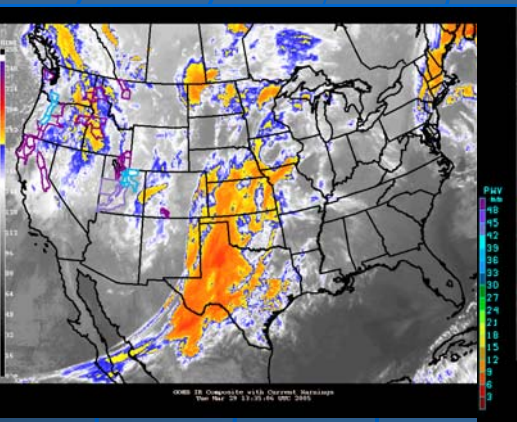
For most people, Unidata is like a utility! Data flows 24x7 and students and faculty use it without knowing much about Unidata.

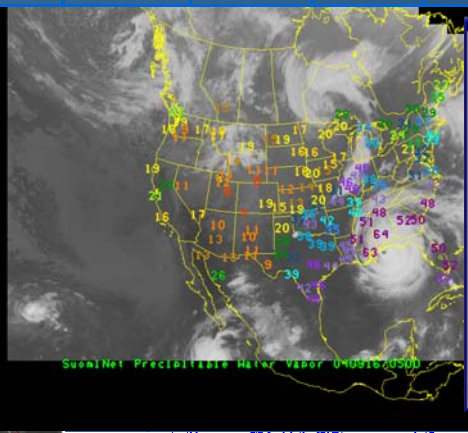
Internet Data Distribution Network

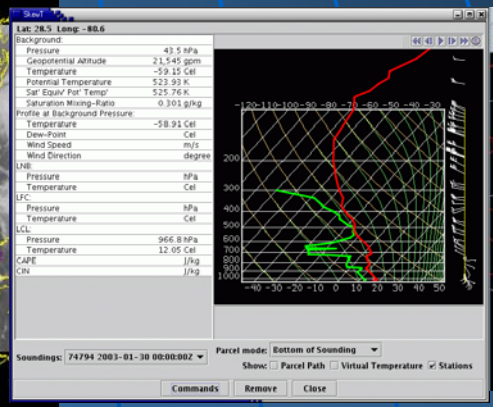


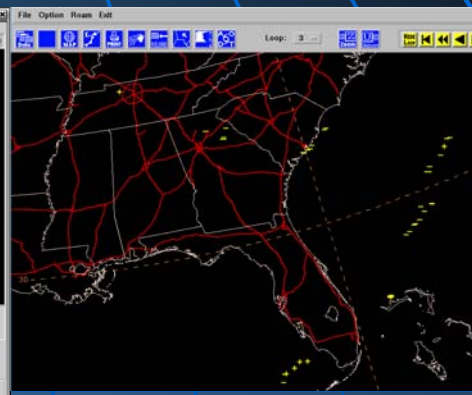
- The IDD is a global data sharing network. Currently, we have nearly 500 host machines in over 260 unique network domains running the latest version of our Local Data Manager software and reporting real time statistics.
- The IDD supports over 30 different data streams, with new data streams being added on a regular basis.
- The volume of data currently flowing to the users through the IDD is about five gigabytes per hour. The volume is expected to increase as more data become available.

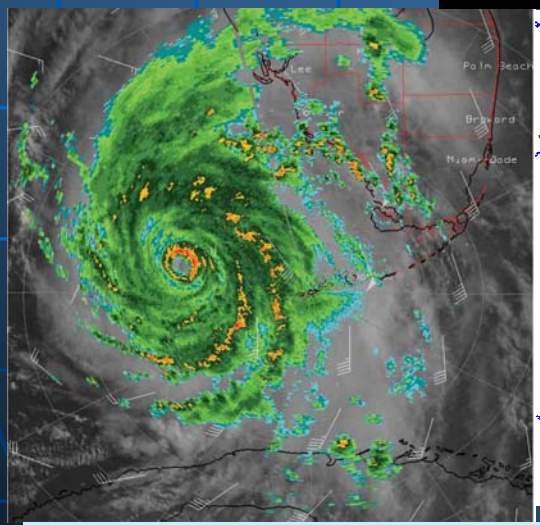
Real-time Data Examples

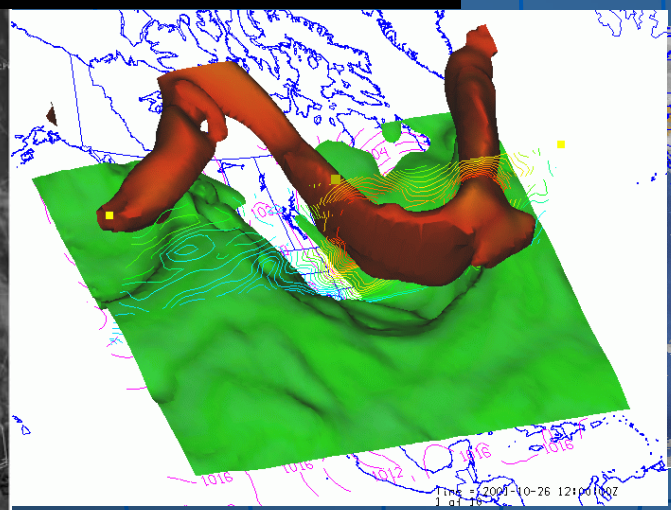


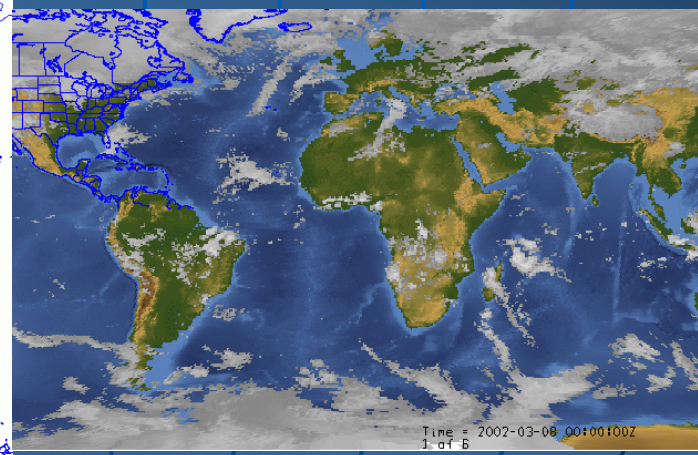












We provide over 30 data streams (surface & upper-air obs., radar, satellite imagery, model output, lightning data, ACARS, NWS bulletins, etc.)

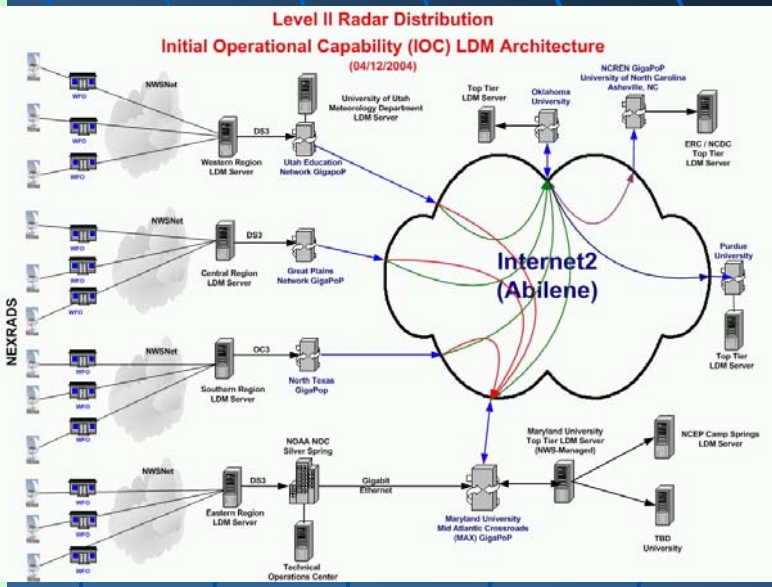
Keys to Community Broadening

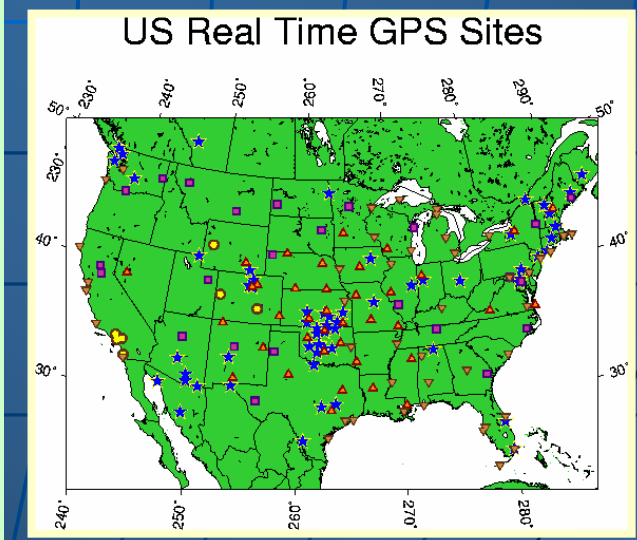
- Provide products and services useful to a broad range of users
- Adopt open source approaches
- Make everything free!!!
- Enable platform independence
- Develop/use common frameworks
- Develop interoperable systems and services
- Build and engage the community
- Foster a culture of sharing
- “Create” evangelists

Community broadening has to be organic!

Unidata & Cyberinfrastructure

- Unidata's niche is in developing tools, middleware and services for data access;
- We facilitate data access and visualization on low-cost computers, lowering the barrier for entry;
- Focus is on developing hardened solutions that work;
- Core belief in the advantages of open source development;
- Software development grounded on open standards;





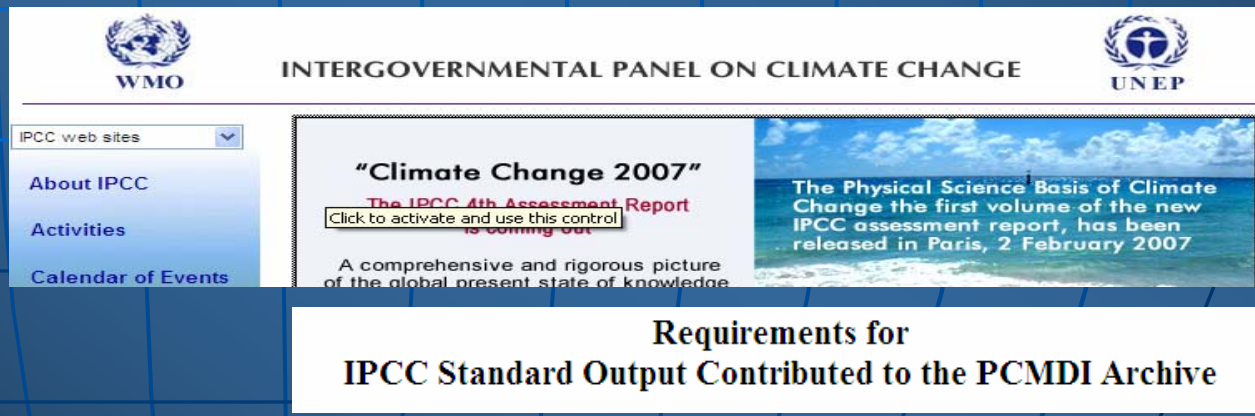
NetCDF: Network Common Data Format

- A Data Model for scientific data: variables, dimensions, attributes, coordinates
- API for data access in C, Fortran, Java, C++, Perl, Python, Ruby, …
- A Self-describing format for portable binary data

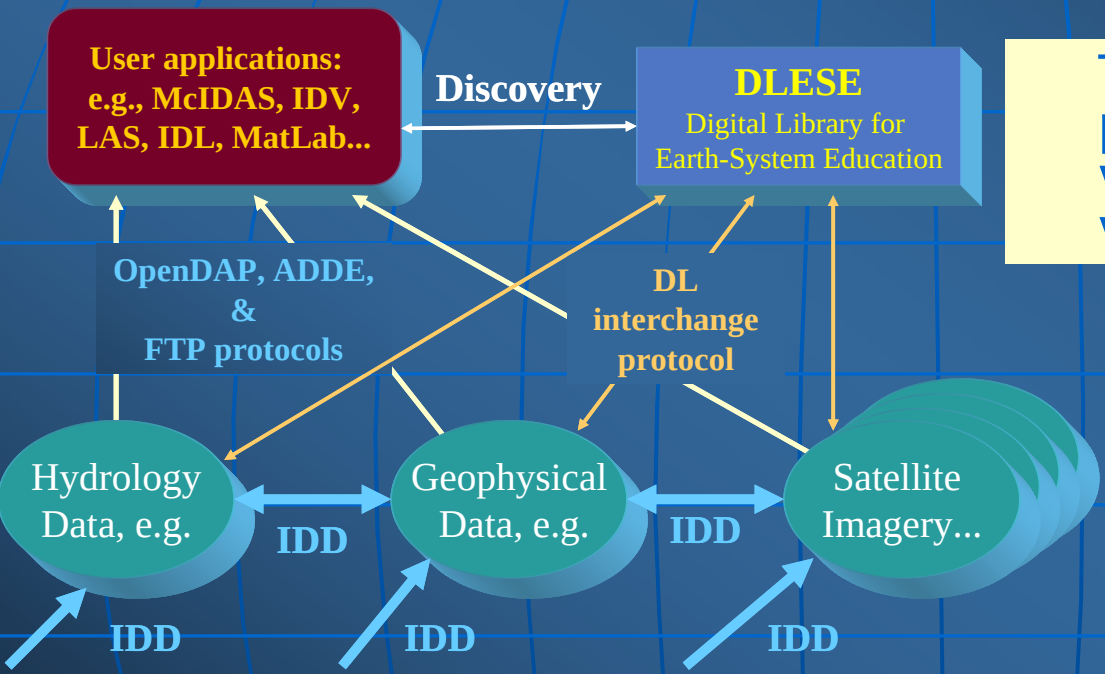
Users need not know anything about the format

NetCDF has become a de facto standard for geosciences data

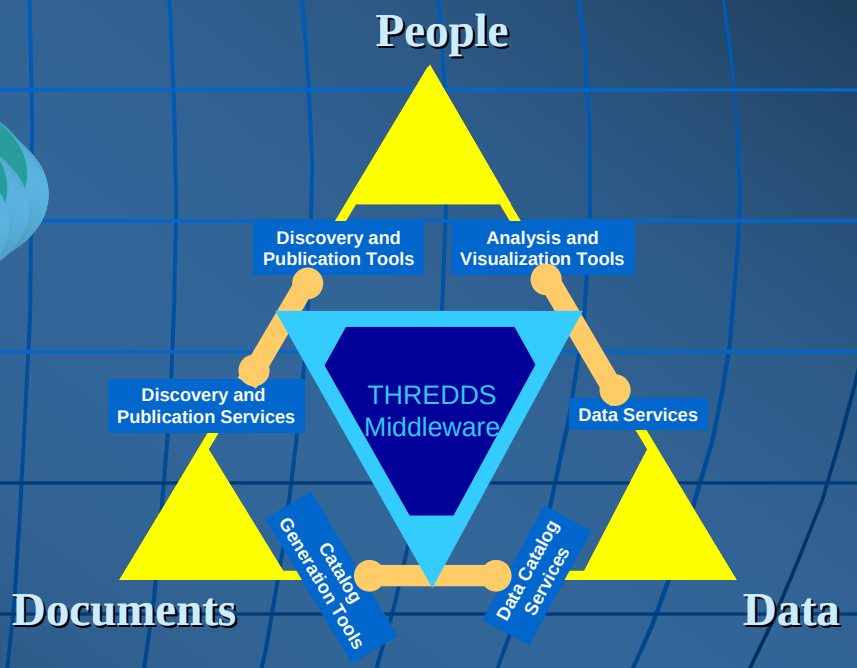
NetCDF & HDF-5 have been integrated to take advantage of the strengths of the two applications



Thematic Real-time Environmental Distributed Data Servers (THREDDS)



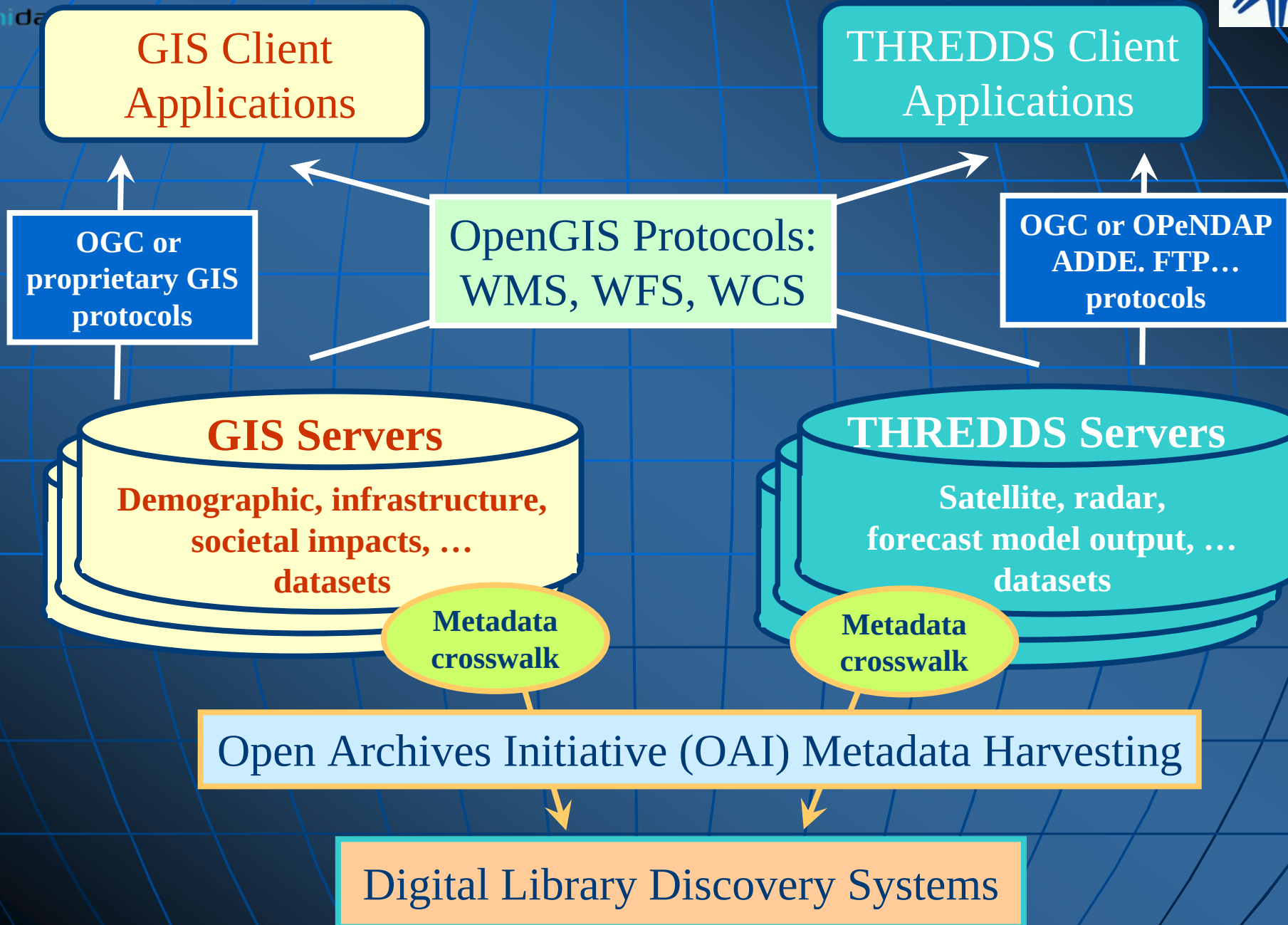
To make it possible to publish, locate, analyze, visualize, and integrate a variety of environmental data



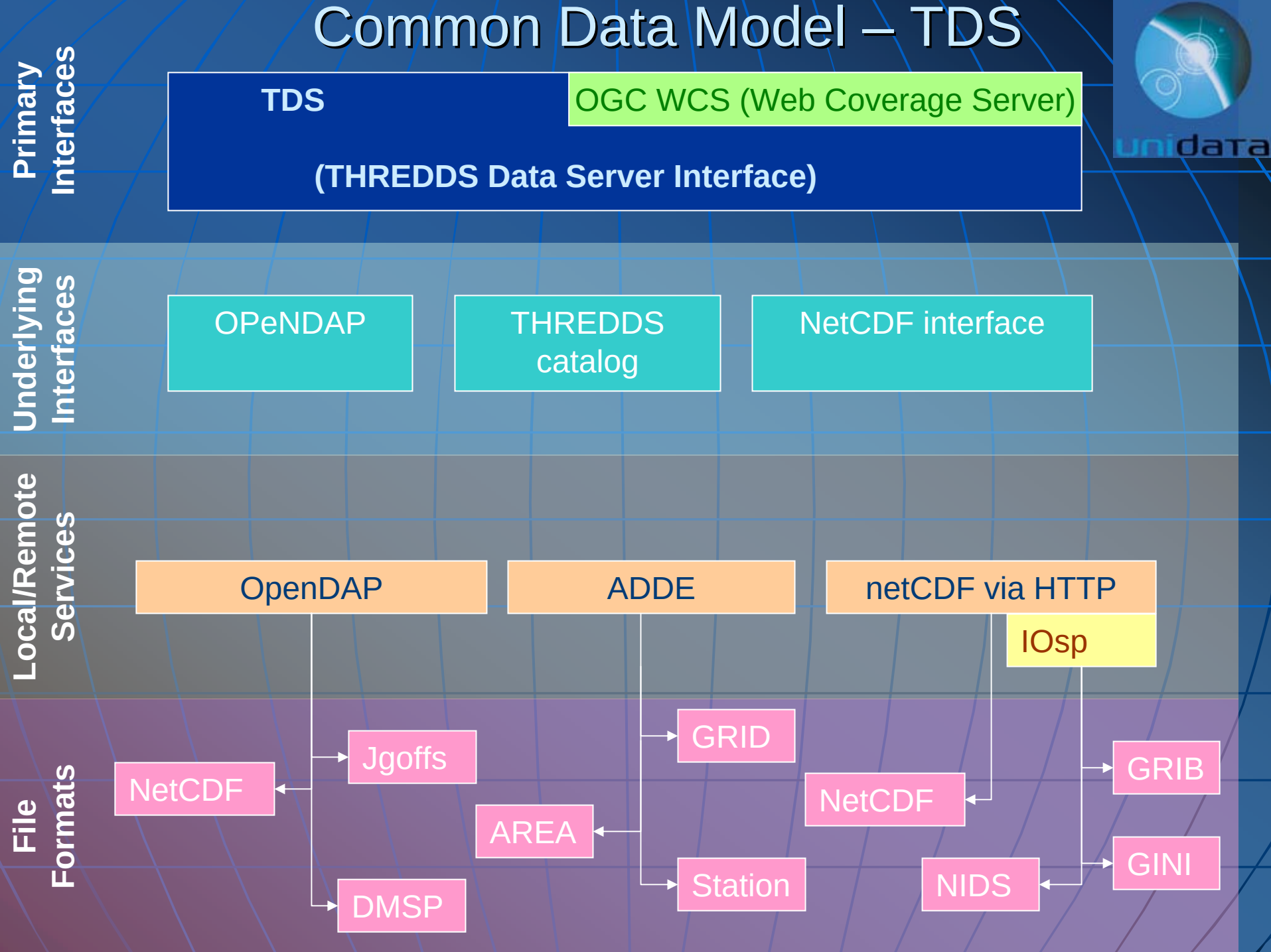
- Combines IDD "push" with several forms of "pull" and DL discovery
- About 25 data providers are partners in THREDDS

Connecting People with Documents and Data

THREDDS Interoperability

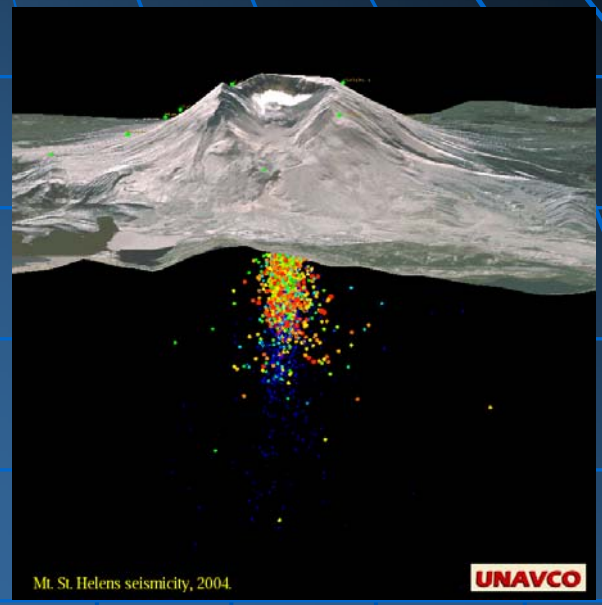


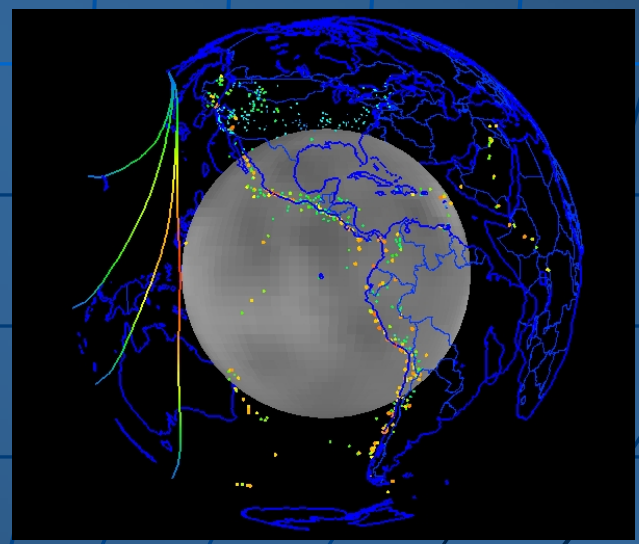
Common Data Model – TDS



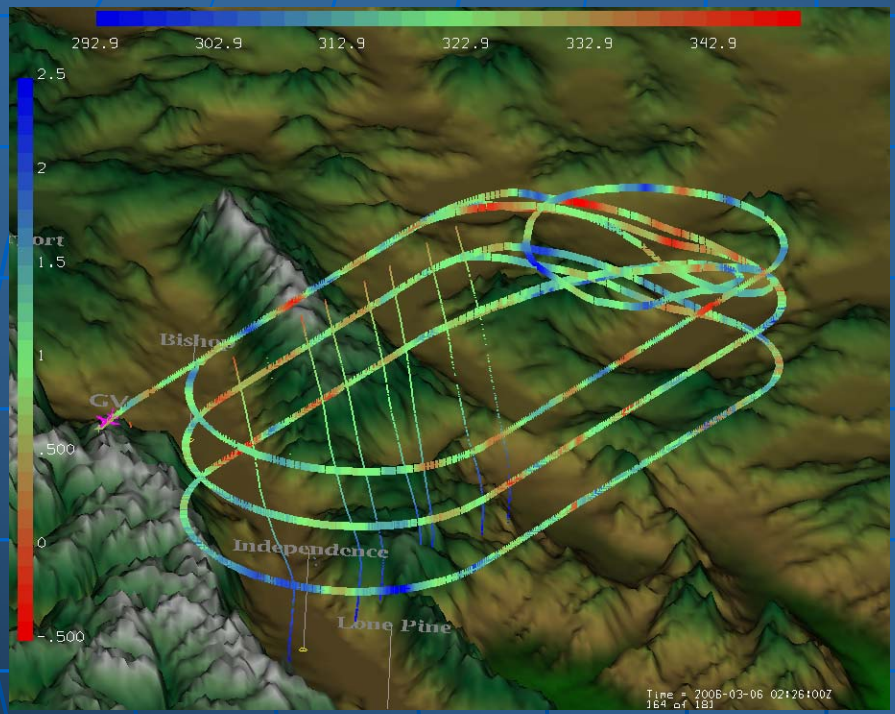
GEON-IDV

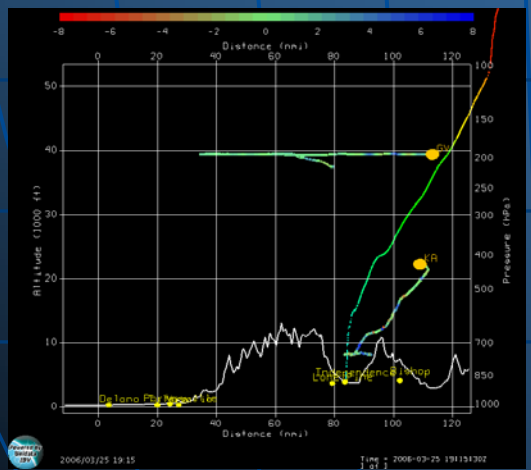
- GEON, a large NSF ITR project in solid Earth Sciences, is leveraging the IDV (and THREDDS) in a major way
- GEON-IDV, their visualization tool developed by Unavco, is an extension of the Unidata IDV for that community.
- The GEON-IDV visualizations are providing unique insights into processes associated with mantle convection, tomography and seismology and have led to new discoveries
- This is a great example of the broader impact of Unidata-developed cyberinfrastructure on the geosciences

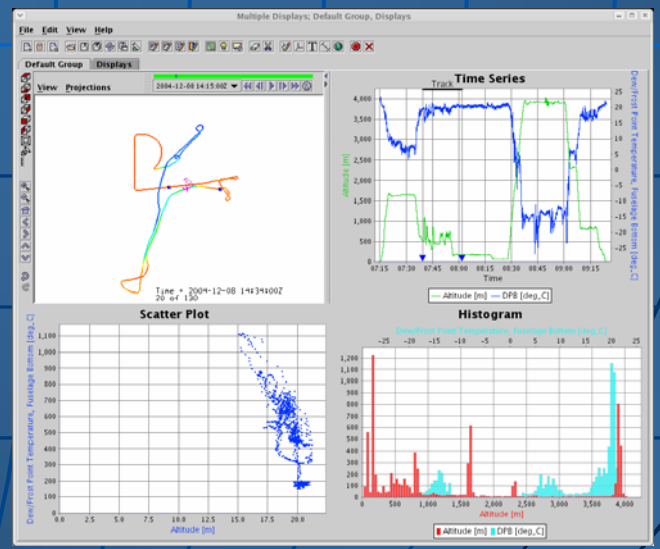




Field Projects

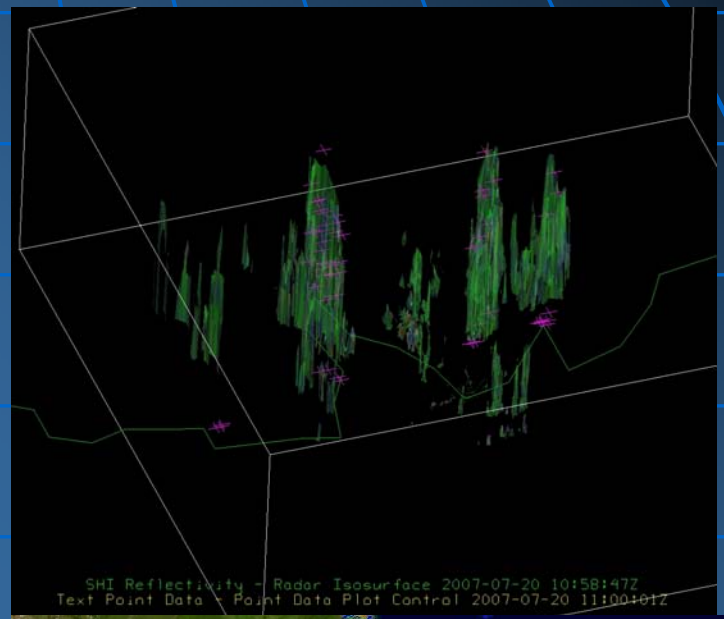


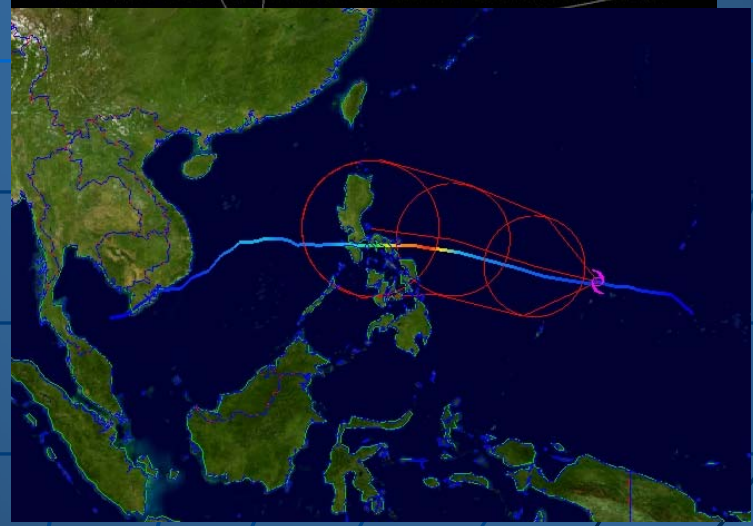




Collaborations with Shanghai Typhoon Institute

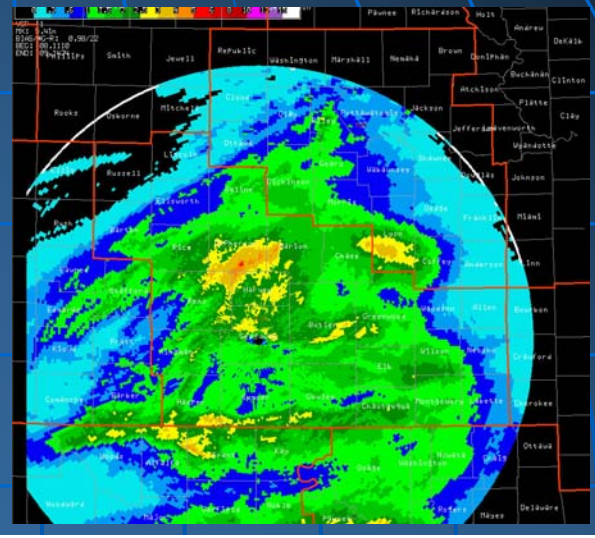
- STI and Unidata will collaborate to design and develop a Tropical Cyclone version of the IDV (TC-IDV)
  - TC-IDV will enhance capabilities of the IDV (diagnostic and visualization functions)
  - Provide an integrated system for TC research and operations
  - Unidata will provide support, training, modifications to IDV framework
  - IDV provided to STI under LPGL

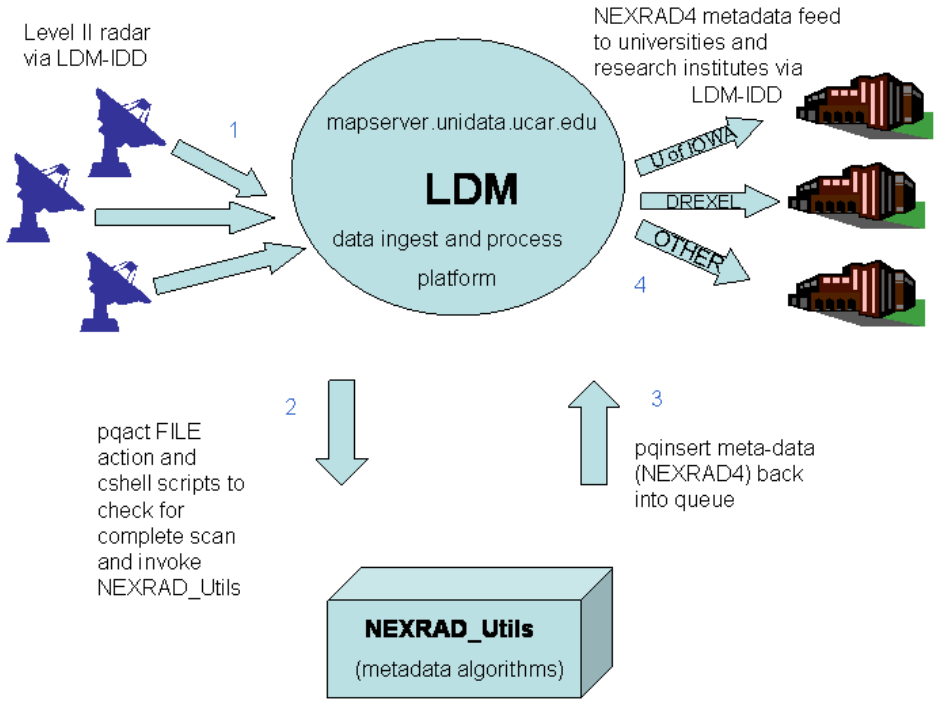




Unidata and Hydrology

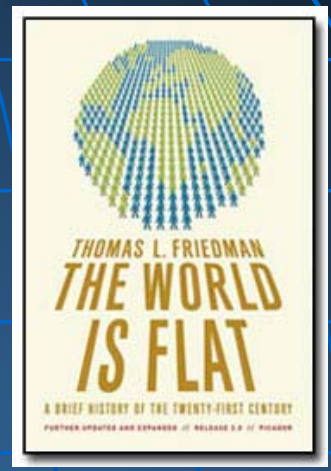
- The CRAFT Project opened new opportunities for collaboration with the hydrology community
- We are working closely with the hydrology community (CUAHSI) to bring Unidata systems and tools to that community
- HYDRO Feed – An Innovative, Value-added Data Stream that provides not just radar data, but value added metadata (e.g., #of pixels over 50 dbZ, precipitation rate in excess of 1"/hour, etc.)





Globalization and Networked Science

- Opportunities for individual empowerment
- Lightning swift advances in technology and communications, putting people all over the globe in touch with one another
- Distributed knowledge communities working collaboratively in virtual organizations
- Networked science tackling problems never possible before and creating new knowledge (e.g. IPCC assessments of climate change – the gold standard)



THE CHRONICLE OF HIGHER EDUCATION

Information Technology

Cyberinfrastructure: the Second Revolution

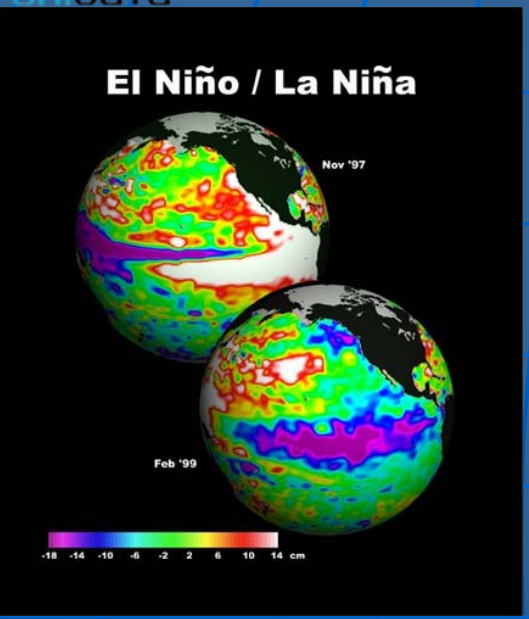
By ARDEN L. BEMENT

The Chronicle Review

The Dawn of Networked Science

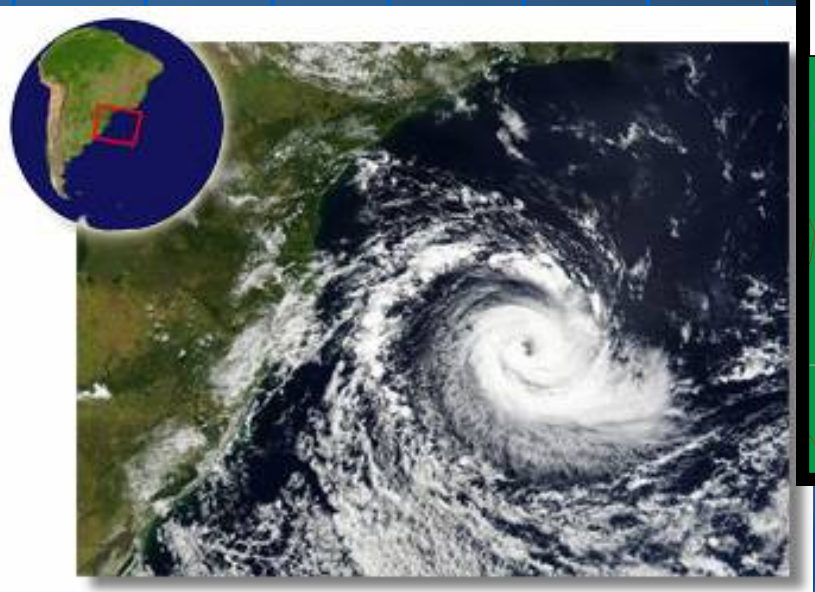
By DIANA RHOTEN

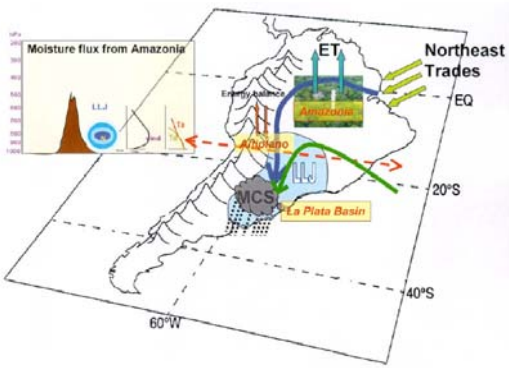
Importance of International Partnerships



SALLJEX, 2003

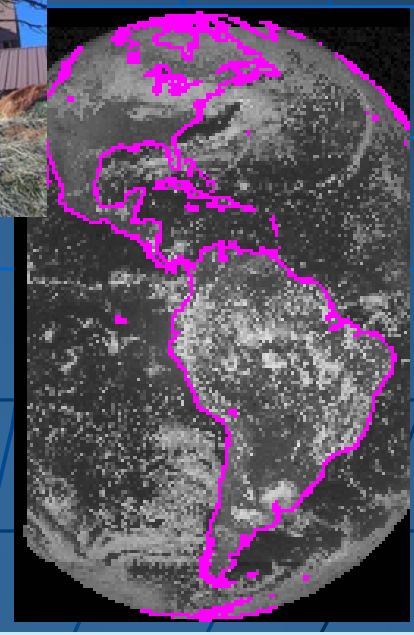
Hurricane Catarina, 2004





Data Sharing and Community Building

- Data Sharing and Community broadening are important goals for Unidata. Specific objectives include:
  - Democratizing access to and use of data that describe the dynamic earth system
  - Building capacity and empowering geoscientists and educators worldwide
  - Strengthening international science partnerships for exchanging knowledge and expertise
  - Effectuating sustainable cultural changes that recognize the benefits of data sharing, and
  - Building regional and global communities around specific geoscientific themes.



Automated Surface Sta

INPE, Brazil

University Corporation for Atmospheric Research

L I N K E D
ENVIRONMENTS
FOR ATMOSPHERIC
D I S C O V E R Y

UNC CHAPEL HILL

Colorado State University

UAH The University of Alabama in Huntsville

HOWARD UNIVERSITY

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

INDIANA UNIVERSITY
Quality Education. Lifetime Opportunities.

the university of OKLAHOMA

MILLERSVILLE UNIVERSITY

A large ITR project to develop a facility for on-demand, real-time data assimilation, weather prediction, and data mining.

The democratization of access to the above capabilities has brought a whole new community of users.

Cooperative Arctic Data and Information Service

- An IPY effort to support Arctic science studies, funded by NSF/OPP
- **Goal: To provide data services for the Arctic Observatories and other IPY-related projects**
- A collaborative project involving Unidata, NCAR/EOL, NCAR/CISL and NSIDC
- Several Unidata technologies will be applied, including the LDM, THREDDS Data Server, and the IDV
- Users in the polar sciences community are now starting to use Unidata systems and tools.

Integrating Data with Educational Materials

LEAD-TO-LEARN MODULES

(Modules created by Millersville LEAD undergraduate students)

Exploring The Polar Jet Stream

Students interact with numerical model output from the North American Mesoscale (NAM) model to explore the components of the polar jet stream.

Exploring Lake Effect Snow

Students interact with numerical model output to explore the ingredients for generating lake effect snow. Students use a case study that covers the event that occurred in the Oswego, NY area on January 28-30, 2004.

From Observations to Models

Students learn about the different data sources used to initialize numerical weather prediction (NWP) models as well as complexity of the data assimilation process used in most models.

Exploring Land/Sea Breeze Circulations

Students interact with numerical model output from the Global Forecast System (GFS) to explore the land/sea breeze circulation. Students use a case study that covers the event that occurred in Florida on September 1, 2005.

Future journal publications will not only have the article, but the associated data, animations, and links to community blogs where people are discussing the paper.

Our IDV and THREDDS work is facilitating this approach to data-publication integration.

Latin America Data Workshop

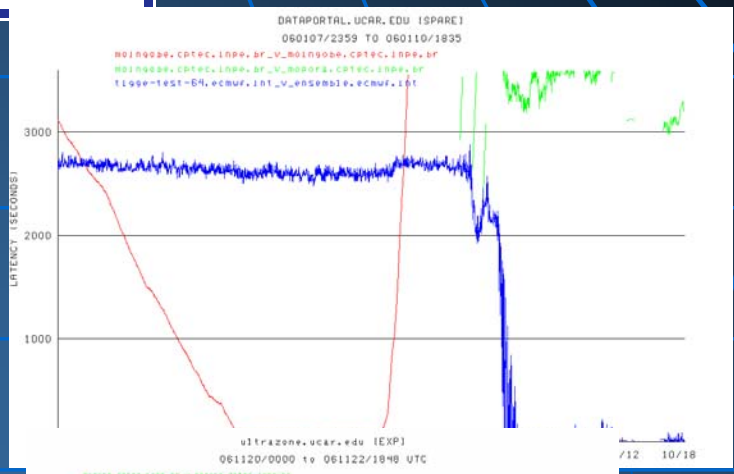
- Unidata, with funding from UOP's JOSS program and in conjunction with the Universidade de São Paulo's (USP's) Instituto de Astronomia, Geofísica e Ciências Atmosféricas (IAG), hosted a Latin American Data Workshop on August 21-23, 2008 in the IAG facilities on the USP campus in São Paulo, Brazil.
- The goals of the workshop were to:
  - Foster scientific partnerships for exchanging knowledge and expertise among U.S. and Latin American educators and researchers
  - Promote greater Latin American participation in free-and-open sharing of Earth System data
  - Inform workshop participants of the wide variety of data available through Unidata
  - Inform participants of the suite of freely-available analysis and display applications available through Unidata
- The 45 workshop participants came from 18 organizations in 6 countries in South, Central, and North America.
- The workshop was a resounding

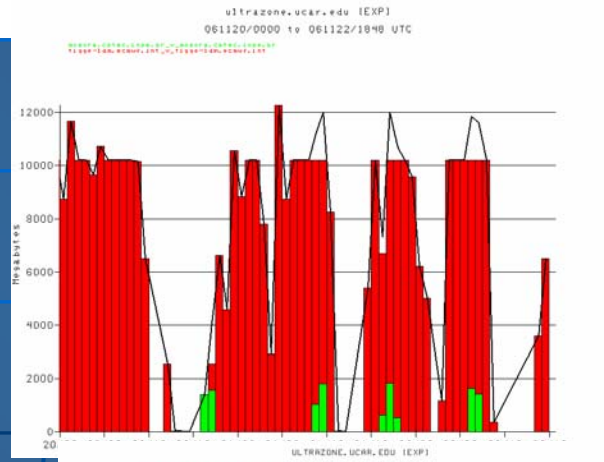
THORPEX

A Global Atmospheric Research Programme

TIGGE

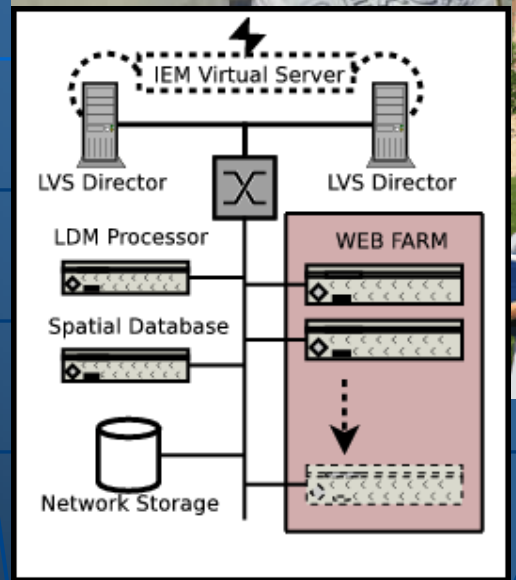
- Unidata technology is being used in the THORPEX Interactive Grand Global Ensemble (TIGGE) project
- There are three TIGGE Archive Centers (NCAR, ECMWF, and CMA, Beijing)
- Each center hosts data from operational global models at NCEP, FNMOC, ECMWF, UKMO, BMRC, JMA, CMA, MS Canada, CPTEC, Meteo-France, and KMA in GRIB2 Format on native grids.
- TIGGE is regarded as a WMO Information System pilot project;

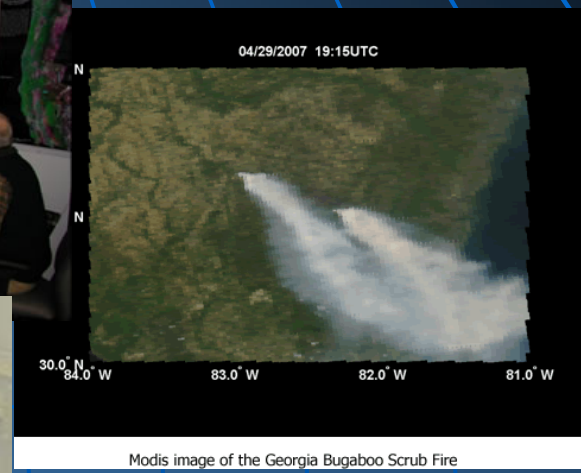




Community Equipment Awards

- The primary purpose is to encourage new academic members from diverse disciplinary backgrounds in the geosciences to join the Unidata community, and for existing members to continue their active participation
- A total of $100K is allocated for this program each year, targeting one or two themes
- Over the past 5 years, we have made awards to 40 universities, ranging from $5K to $25K.





Modis image of the Georgia Bugaboo Scrub Fire

Concluding Remarks

- We live in an exciting era in which the confluence of science, technology and societal behavior is reshaping the conduct of research and education.
- A new generation of data and information services are enabling new discoveries and the use of innovative education strategies.
- Scientific partnerships, data sharing, and community-building activities are crucial for this transformation.

Thank You!

- Questions?
- Contact information: mohan@ucar.edu
- http://www.unidata.ucar.edu/