



unidata



Satellite Data Access and Use: Issues, Challenges, and Opportunities for Academia

CLASS User Workshop
User Community panel
Boulder, CO
11 August 2005

Dr. Mohan Ramamurthy
Director, Unidata Program Center
UCAR Office of Programs
Boulder, CO



Unidata



■ Mission Statement:

Provide data, tools, and community leadership for enhanced Earth-system education and research.

At the Unidata Program Center, we

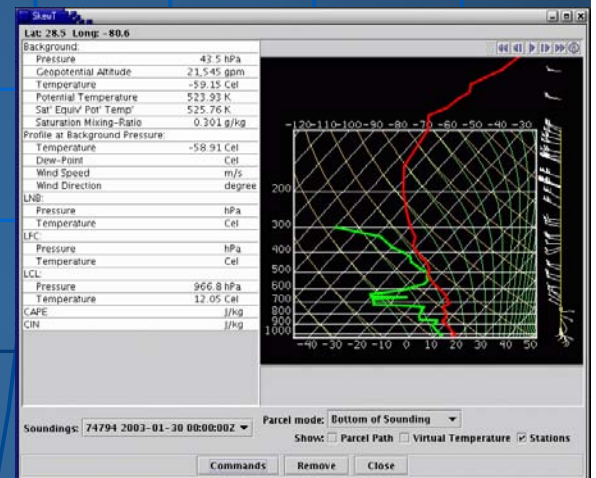
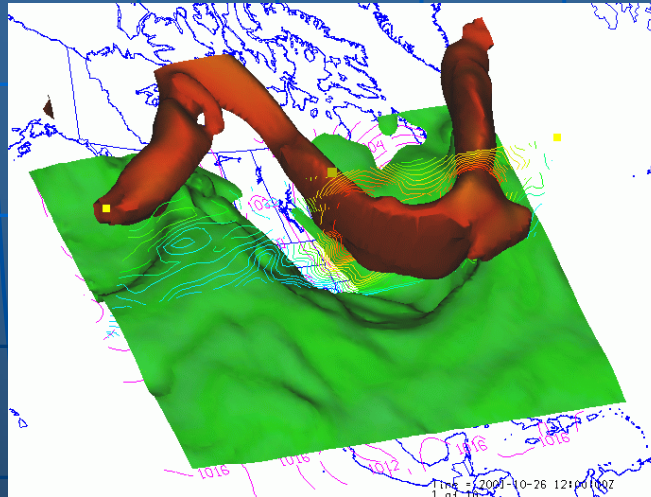
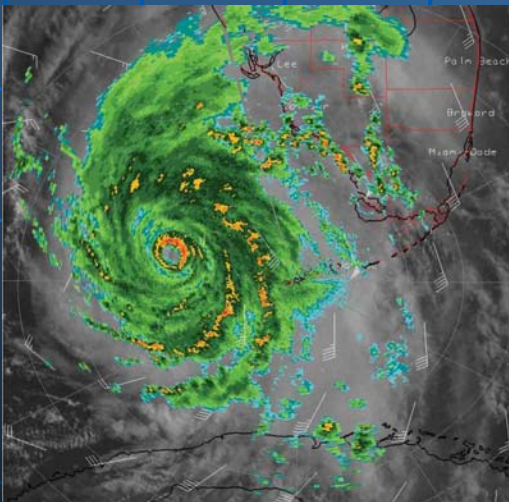
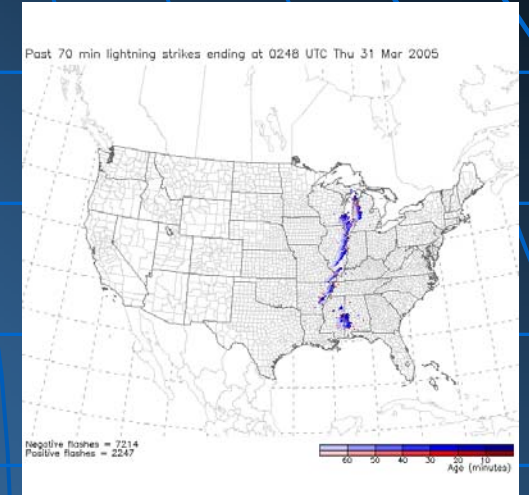
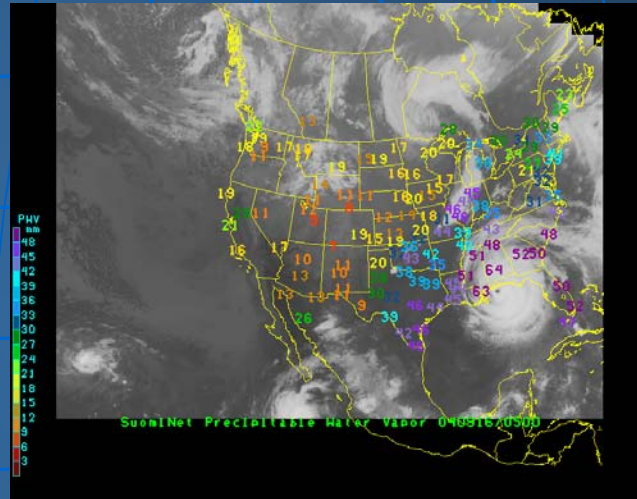
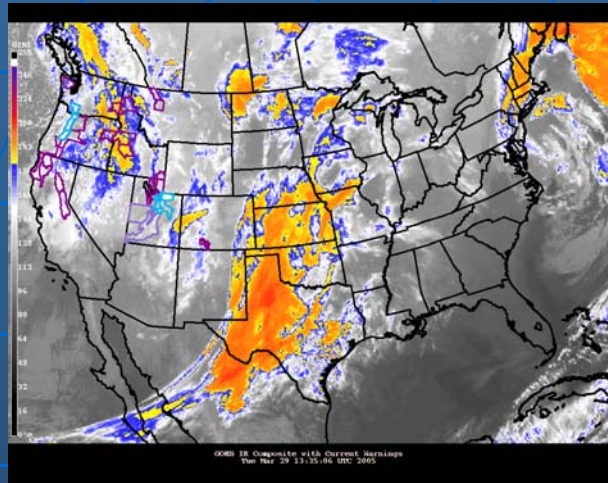
- *Facilitate [Real-time] Data Access*
- *Provide Tools (GEMPAK, McIDAS, IDV, LDM, netCDF)*
- *Support Faculty and Staff*
- *Build and Advocate for a Community*

Everything we provide (data, software and support) is free.



unidata

Real-time Data Examples





unidata

Typical Unidata Systems



- Typical Unidata systems at universities: \$5K - \$30K
- Each year, we give modest equipment awards to our community (\$5K-\$25K range) to participate in the Unidata system
 - This year, we received 24 proposals for about \$450K
 - We will be awarding a total of \$100K
- We normally recommend the use of commodity systems based on 64-bit X86 (Athlon, Itanium) processors running Unix/Linux OS
- Memory, Storage and network interface are usually more important than processor power for data access

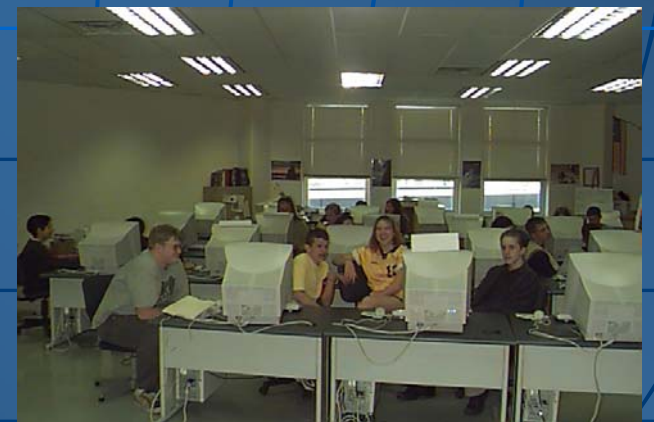




Impact of Unidata



- According to a 2001-02 survey:
 - 2/3rd of our sites have users outside of the atmospheric sciences.
 - Over 21,000 students per year use Unidata tools and data in classrooms and labs.
 - More than 1,800 faculty and research staff use Unidata products in teaching and research.
 - Unidata-connected university programs influence over 55,000 K-12 students annually.



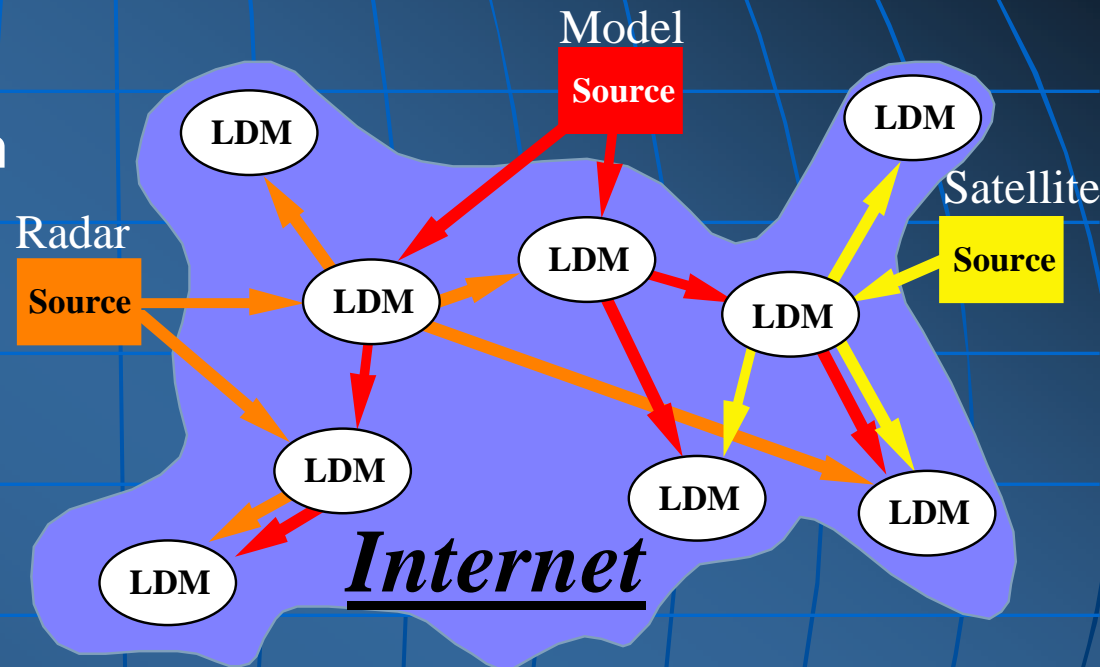


unidata

Internet Data Distribution



- About 150+ sites are participating in Unidata Internet Data Distribution (IDD) system
- Approximately 2 GB of data injected/hour *from distributed sources*;
- Unidata IDD/LDM uses more of the Internet2 than any other *advanced application*;
- Over 15 Terabytes of data transmitted each week – 4-5% of Internet2 traffic).

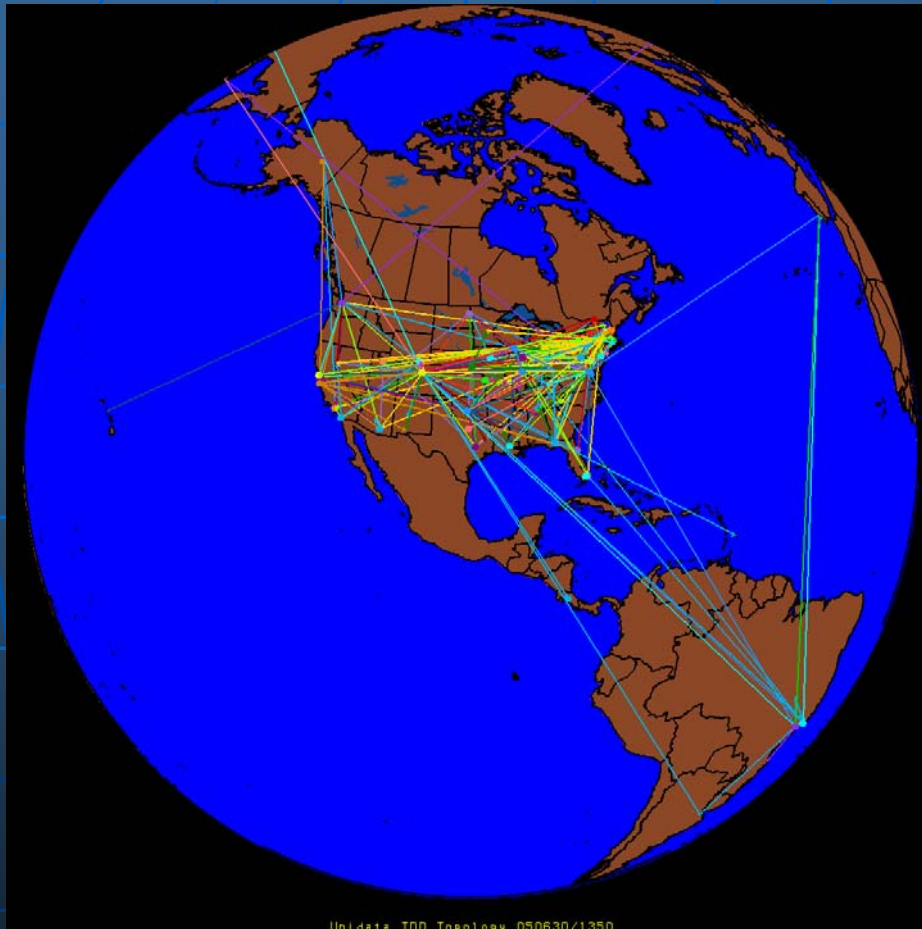


The LDM is now ranked #3 (behind HTTP, NNTP) in Internet 2 usage.

Last year, LDM surpassed FTP.



IDD: A Real-Time Snapshot



The Unidata community is diverse and global.

There are over 150 institutions in North and South America, Europe, and Asia receiving real-time data using the Unidata Internet Data Distribution system and Local Data Manager.

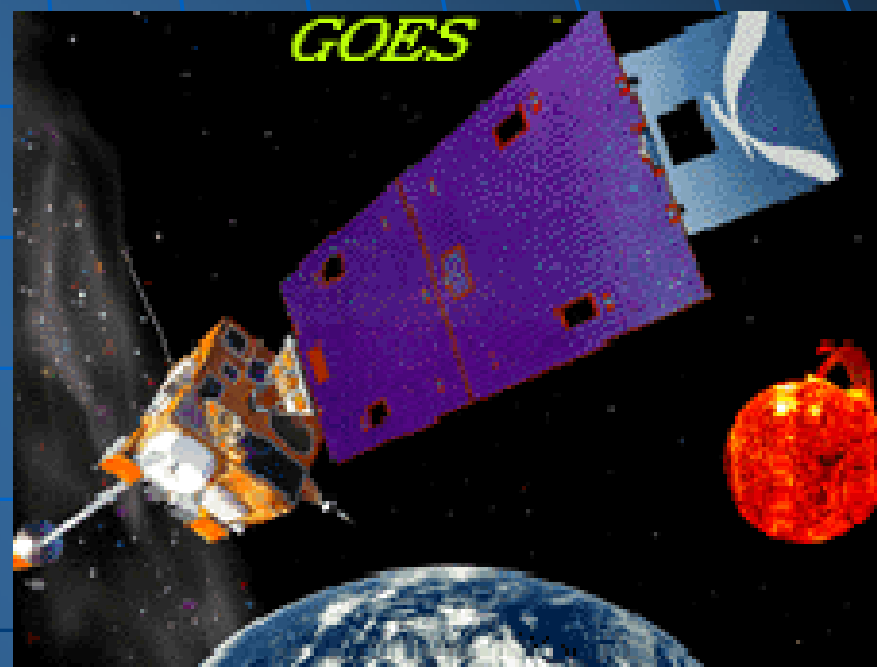
There are probably as many LDM sites behind firewalls and intranets whose performance we do not monitor.



GOES Subscribers in the Unidata IDD



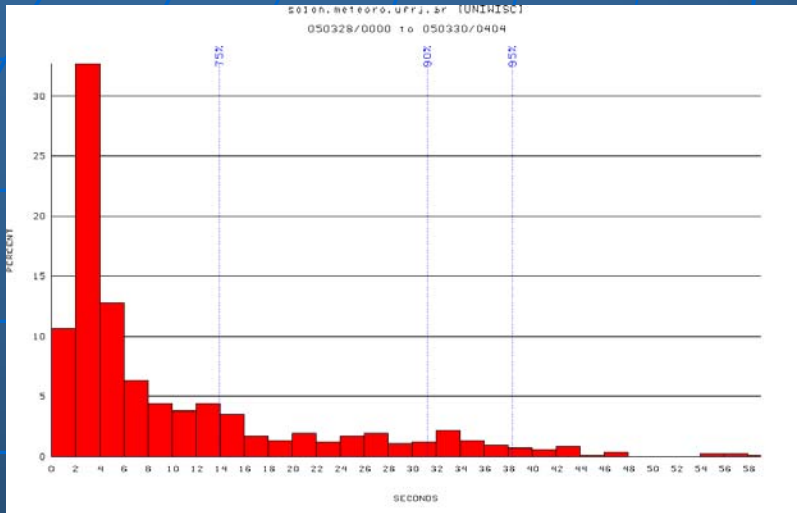
- 117 Hosts in 81 domains
- Very high participation rate
- Over 20 products are distributed, including many products derived by CIMSS/SSEC
- Data Rate: 20-30 MB/Hour
- GOES imager products are one of the most popular products for education



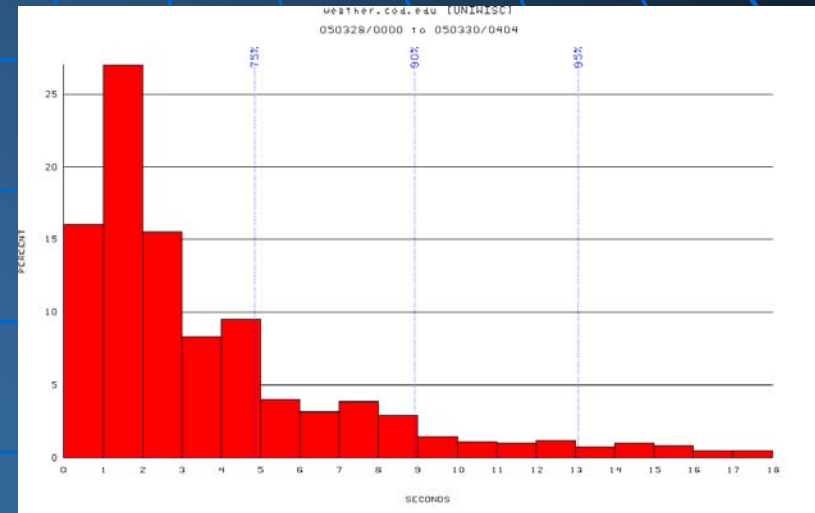


unidata

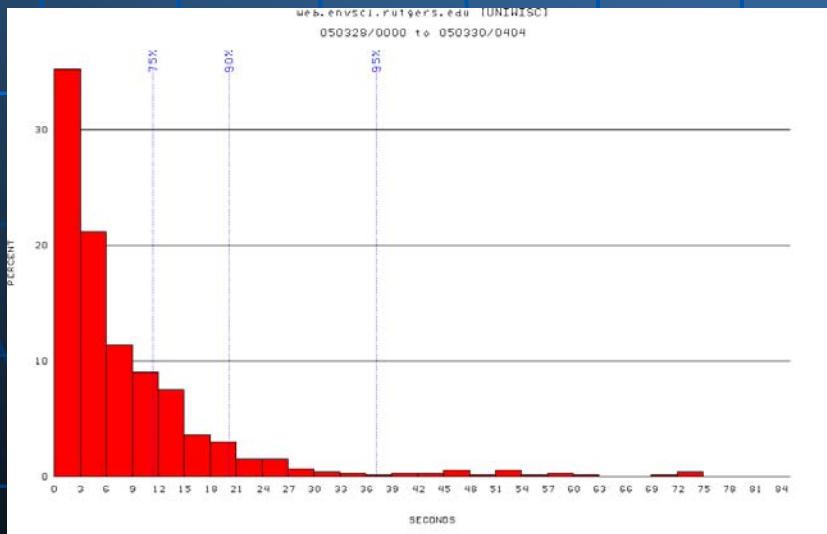
GOES Data Reception



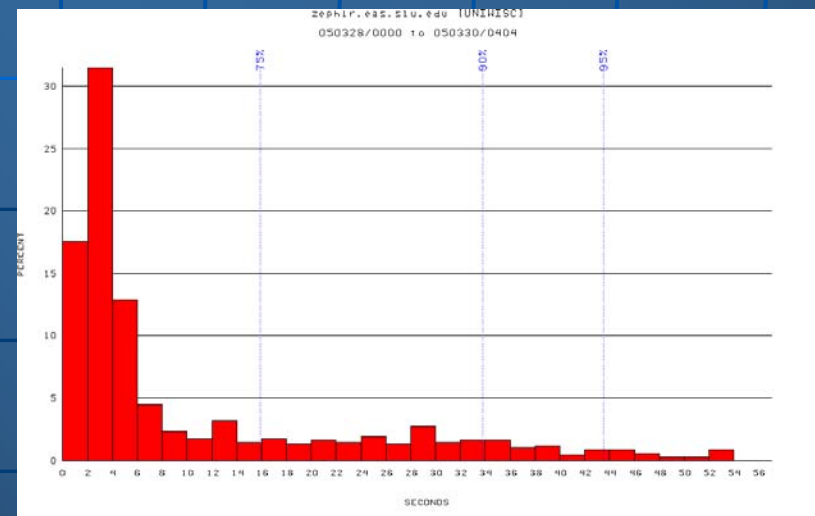
UFRJ, Brazil



College of DuPage



Rutgers University



St. Louis Univ.



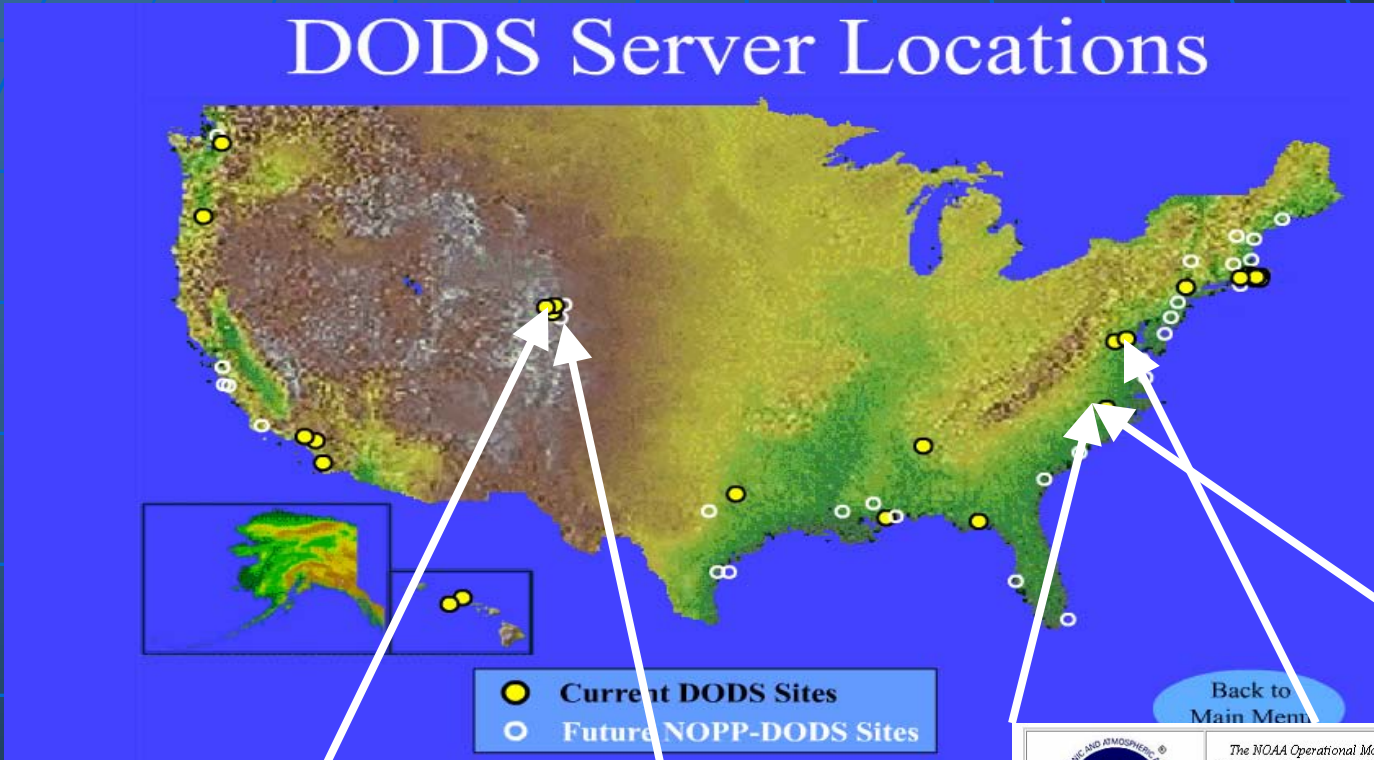
unidata

OPeNDAP/DODS



- Open Source Project for a Network Data Access Protocol
- OPeNDAP is the discipline independent core infrastructure for remote data access
 - One can design a system of distributed data servers using the OPeNDAP protocol
 - OPeNDAP servers deal with data population and provision.
 - This has been done for real-time and retrospective operational model data (NOMADS, THREDDS, Community Data Portal)
 - Thin and thick client applications can access and use data on distributed OPeNDAP servers
 - The data can be cataloged using THREDDS and published via digital libraries



OPeNDAP/DODS Servers



Community Data Portal
a gateway to data for Atmospheric Sciences
 hosted by the National Center for Atmospheric Research

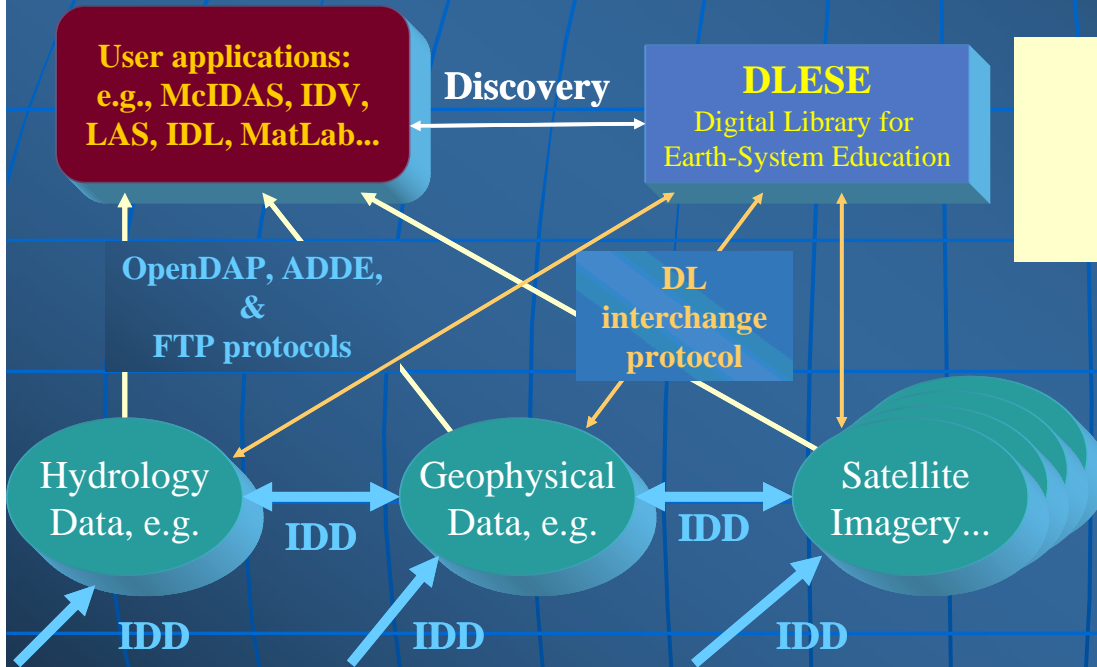
BROWSE	SEARCH	ANALYZE
Hierarchical access to the UCAR data catalogs by Division <input type="text"/> by Date <input type="text"/> by Geographic Area <input type="text"/> by Project <input type="text"/> by Topic <input type="text"/>	Query the UCAR metadata holdings Keywords : <input type="text"/> optional qualifiers: Data Type <input type="checkbox"/> Observed data <input type="checkbox"/> Modeled data Category <input type="text"/> Atmospheric	Data analysis and visualization tools DODS aggregation server Live Access Server GrADS E-viewers

UCAR data catalogs conform to the *THREDDS* specification. These catalogs were written by data managers and specialists from ATD, CcD, GLESE, SCD and UNIDATA as an R&D effort promoted by the NCAR Data Management Working Group.

	<p style="font-size: x-small;">The NOAA Operational Model Archive and Distribution System (NOMADS) is a pilot project designed to provide real-time and retrospective format independent access to climate and weather model input and output data.</p> <p style="text-align: center; font-size: x-small;">About NOMADS FAQ</p>	
<p style="font-size: x-small;">Status Reports</p>	<p style="font-size: large; font-weight: bold; margin: 0;">NOMADS</p> <p style="font-size: x-small; margin: 0;">The NOAA Operational Model Archive and Distribution System</p>	<p style="font-size: x-small;">Program Plan and Data Management Vision</p>
<p style="font-size: x-small;">Using NOMADS</p>	<p style="font-size: x-small;">NOMADS Data Portals</p> <p style="font-size: x-small;">NOMADS Web Interface</p>	<p style="font-size: x-small;">Participants</p>

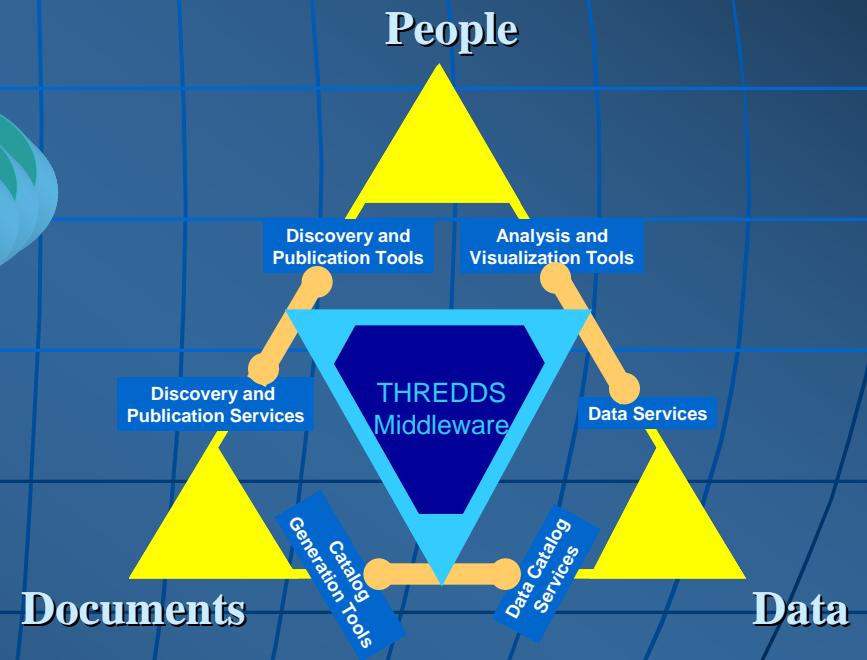


Thematic Real-time Environmental Distributed Data Servers (THREDDS)



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



NOAAPort & DVB-S



- NOAAPort is transitioning to using Digital Video Broadcast – Satellite (DVB-S) technology to deliver high volumes of data to the field;
- New DVB-S technology is scalable to meet data requirements using a single demodulator (up to 43 Mbits/s per demodulator);
- Unidata is working with the community to transition sites to DVB-S; Our solution allows sites to integrate data received via DVB-S reception with IDD data.



The Ultimate Objective of a Data System/Service



- To provide requested data to the user's application (analysis/visualization tool) in a transparent, consistent, readily useable form;
- Users don't care as much about technology as they do about transparency and usability;
- Solution? Interoperability of components, systems and services, achieved using standards.

Source: Peter Cornillon (2003)

Need for a Hybrid System

- Given the very high data rates from future satellite systems, the university community will need a hybrid solution that couples a satellite-based reception system with a terrestrial, Internet-based data access system
- Both local and remote data access mechanisms will be required
- Solutions using remote data access protocols such as ADDE and OPeNDAP already exist
- Need to create seamless data access regardless of mechanism and data will need to work with the same set of tools



Typical Use Cases



- Teaching: undergraduate, graduate and survey courses
 - Lab work, Projects, Term papers, etc.
- Research (thesis/dissertation, case studies, assimilation for local modeling, etc.)
- Disciplinary and multidisciplinary studies - ability to integrate data from different sources is key
- Outreach: real-time weather products, collaborations with NWS, etc.

Community requires seamless access to real-time, archived, and case study data sets



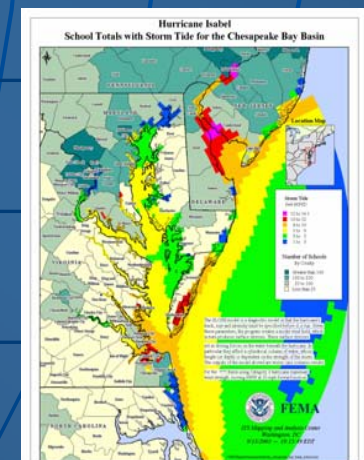
unidata

Unidata 2008: Shaping the Future of Data Use in the Geosciences



We are moving from an era of data provision towards one in which data- and related web-services are important;

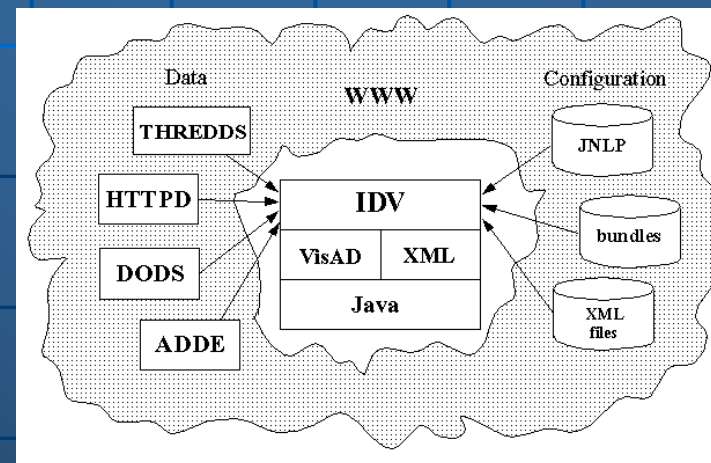
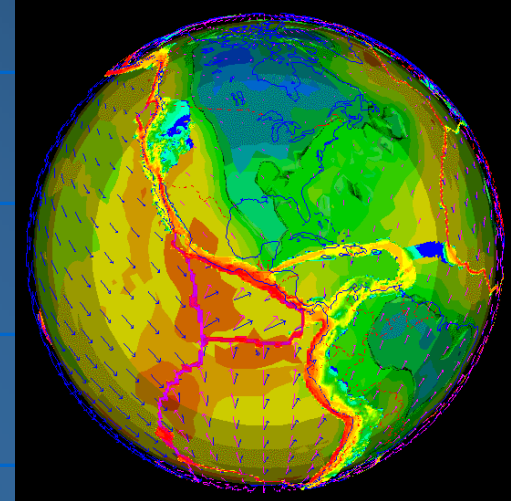
Multidisciplinary integration and synthesis are emphasized.



Integrated Data Viewer



- Unidata's newest scientific analysis and visualization tool
- Built on VisAD library
- Java™ framework and reference application
- Provides 2- and 3-D displays of geoscientific data
- Stand-alone or networked application
- Support for data sharing and real-time collaboration among distributed users



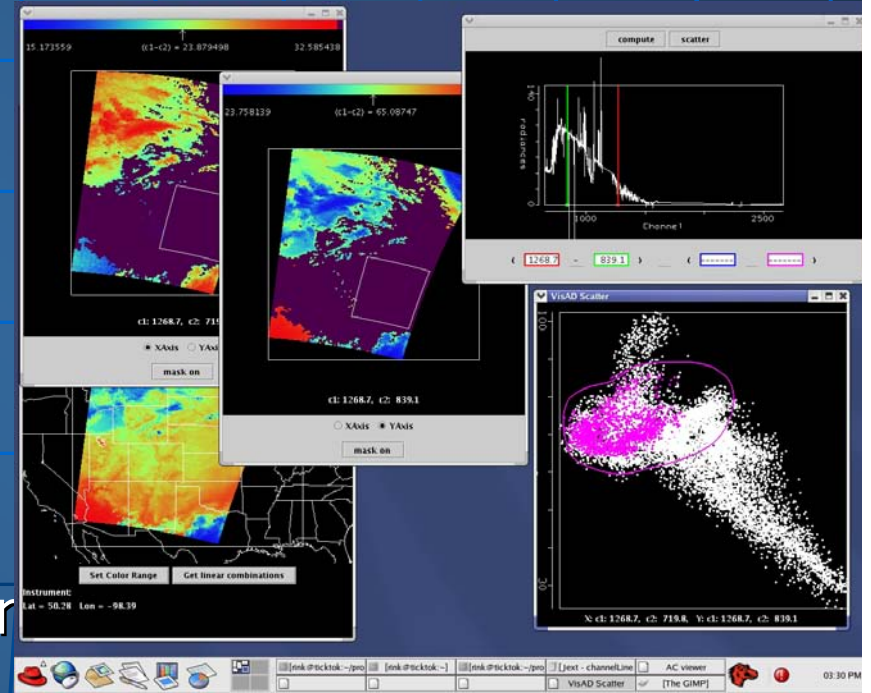


unidata

McIDAS-V



- Space Science and Engineering Center/University of Wisconsin-Madison will transition McIDAS into a package that is based on the Unidata IDV and VisAD
- See "The next generation of McIDAS: A look toward the future" by T H Achtor, T M Whittaker, D A Santek, AGU Fall Meeting, 2004.
- An important consideration for McIDAS-V: Visualization of hyperspectral data, which will be important with GOES-R





unidata



Application

Scientific Datatypes
Grid Station Image

NetcdfDataset

NetcdfFile

THREDDS

OpenDAP

ADDE

Catalog.xml

HDF5

NetCDF-3

I/O service provider

NetCDF-4

NIDS

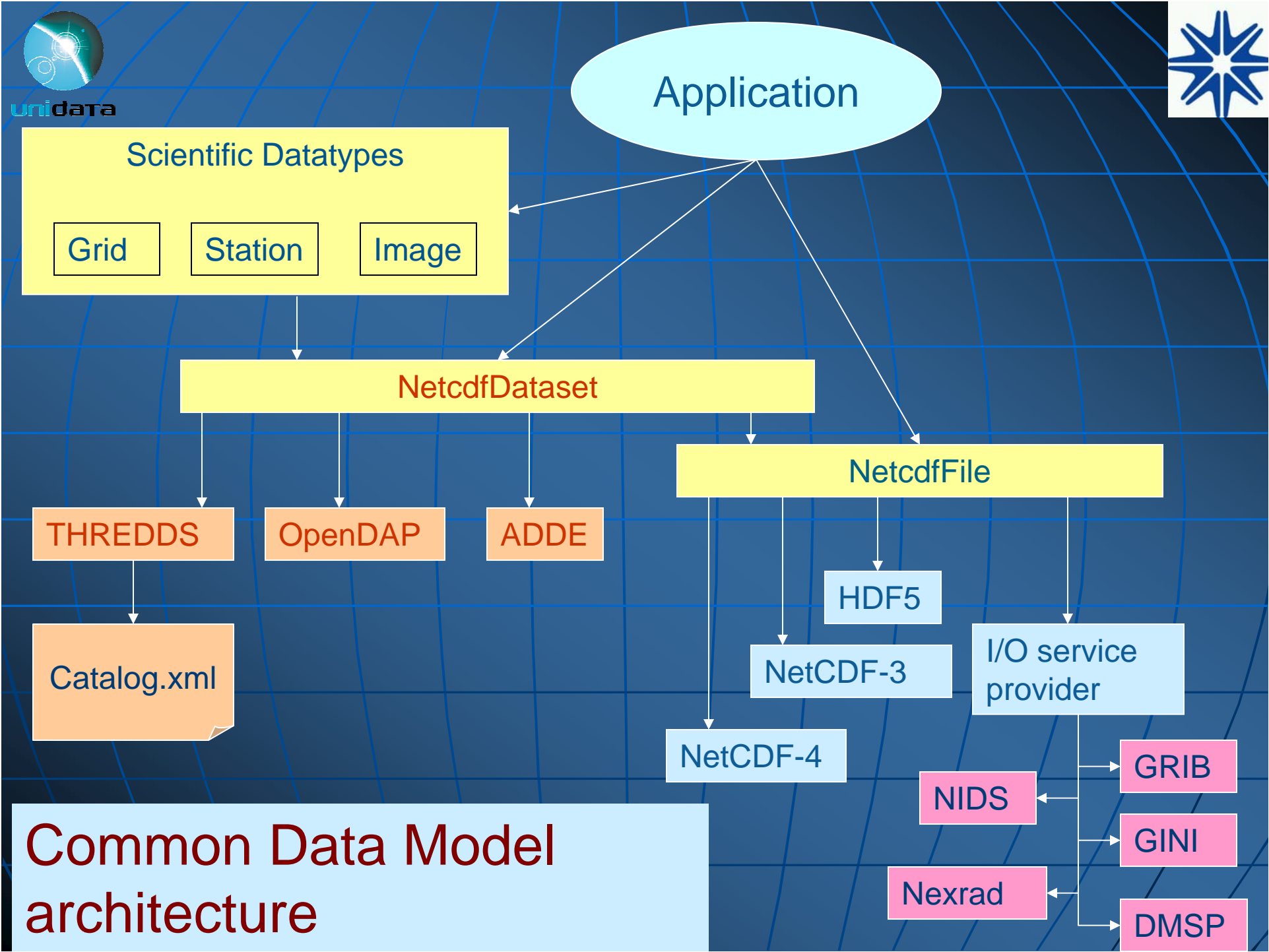
GRIB

GINI

Nexrad

DMSP

Common Data Model architecture





Unidata Users Committee



- Conducts periodic surveys to gather community input
- Provides input to Unidata on data stream selection and other issues
- Conducts Users Workshops every three years

We envision seeking the advice of the Users Committee as we prepare for NPOESS & GOES-R data distribution



unidata

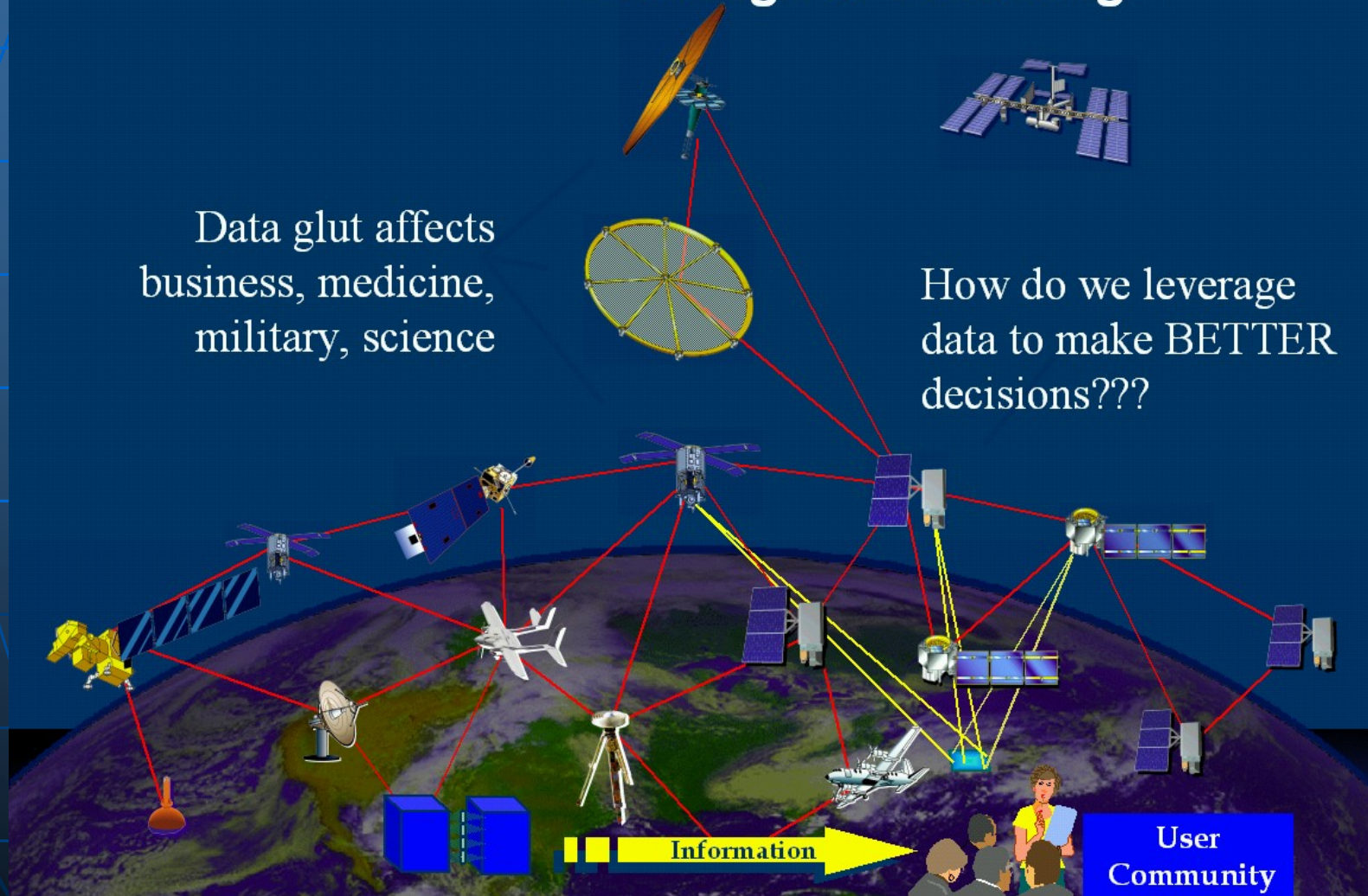


Data Glut

“...drowning in data but starving for knowledge”

Data glut affects
business, medicine,
military, science

How do we leverage
data to make BETTER
decisions???

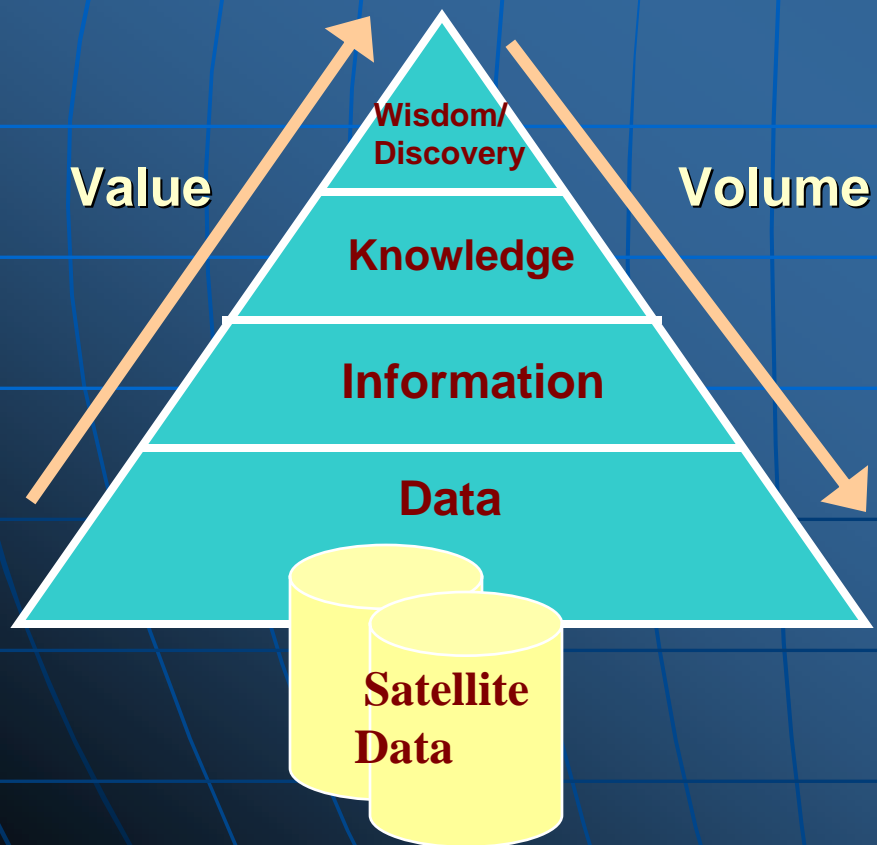


Source: Sara Graves, 2002

Data Mining and Knowledge Discovery



End Users (Decision Makers, Students, Scientists)



- Algorithm Development and Mining (ADaM) System, UAH-Huntsville
- ADAM provides knowledge discovery, feature detection and content-based searching for data values, as well as for metadata.

Source: Sara Graves, 2002



Thank You!

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