Trends in Geoscience Data Services

CONDUIT-CRAFT Meeting
AMS Annual Meeting
Atlanta, GA
31 January 2006

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

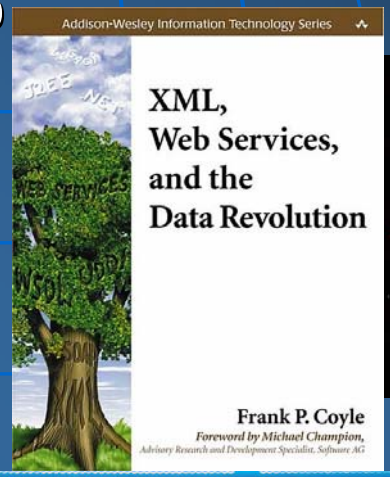
The Ultimate Objective of a Data System/Service

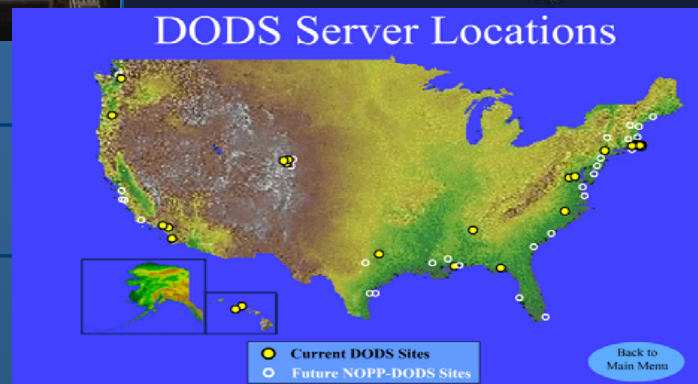
- To provide data to the user's application (analysis/visualization tool) in a transparent, consistent, readily useable form;
  - Deliver the right data in the right format at the right time for ready use
- Finding the right data is almost as important as providing the data
- Users don't care as much about technology as they do about transparency and usability;

Source: Peter Cornillon (2003)

Technology Evolution: Enabling a New Generation of Data Services

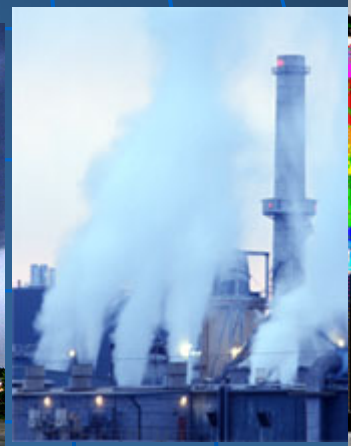
- Internet & the World Wide Web
- Commodity microprocessors
- Object-oriented programming
- Open standards
- Web services
- Extensible Markup Language (XML)
- Global, high-bandwidth and wireless networks
- Digital libraries
- Collaboratories
- Grid Computing/e-Science
- Data Portals and Federated, distributed Servers
- Geographic Information Systems
- Knowledge environments
- Ontologies and Semantic web
- Data mining and knowledge discovery

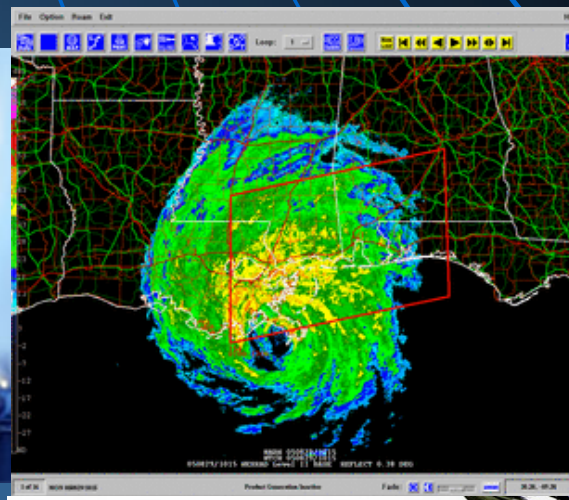


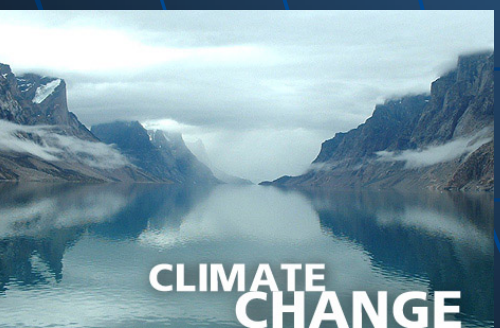


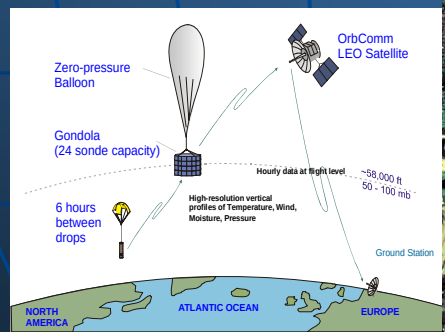
Opportunity to Solve Real-world Problems

unidata

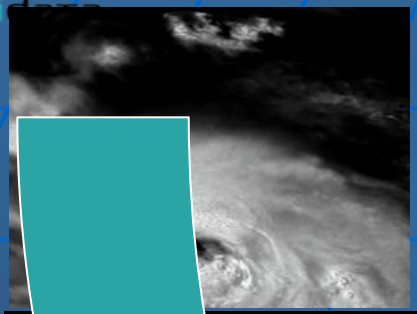


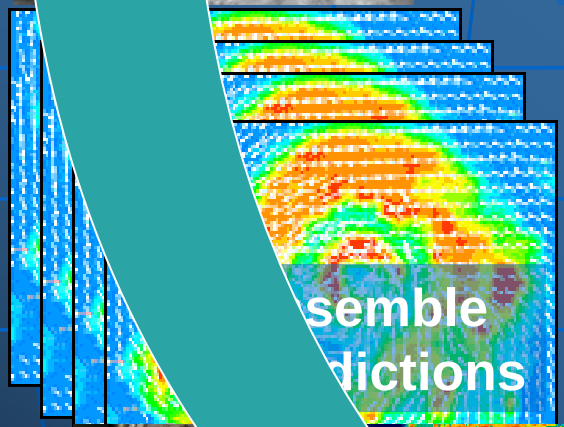


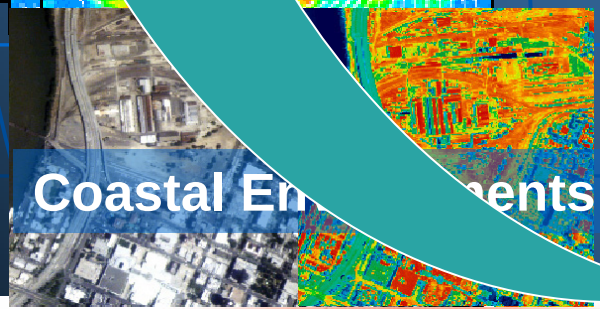


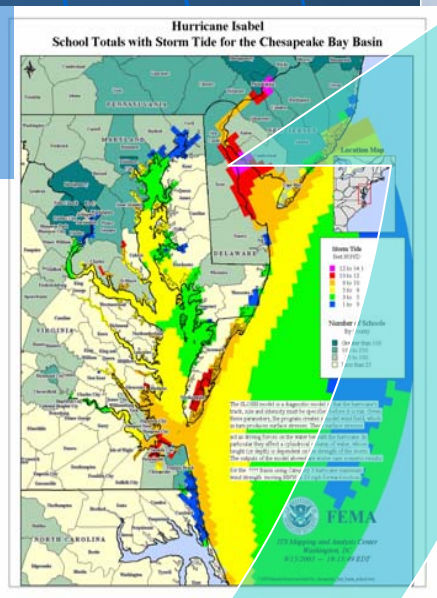


End-to-end Data Services









GIS Integration

semble
dictions

Coastal En ents

Decision Support

Need integrated services

Opportunities and Challenges

GOES-R (2012)

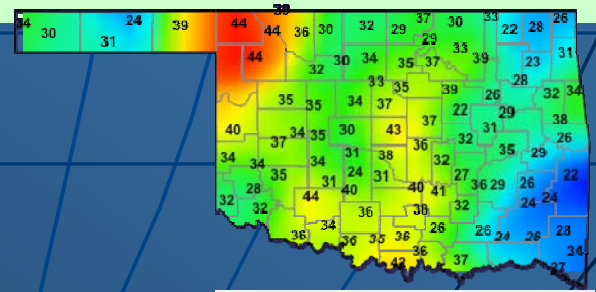
Hyperspectral Environmental Suite (~1600 channels)

NPOESS (201?)

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

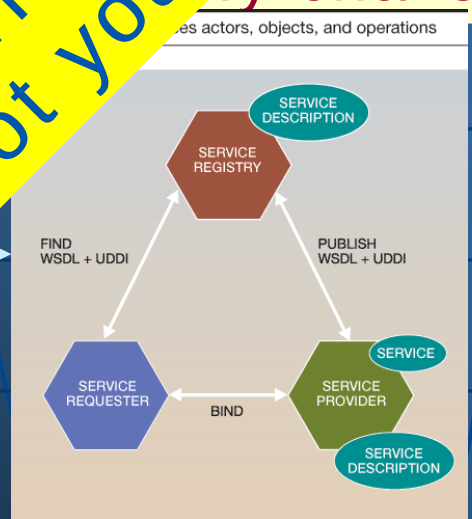
- Heterogeneity and complexity of distributed observing, modeling, data, and communication systems
- Nature of data coverage: diversity and multiple spatial and temporal scales
- Use of legacy and contemporary technologies
- Lack of standards and interoperability
- User community not monolithic
- Political, technological, and cultural and regulatory barriers
- Integration with GIS and Decision Support Systems

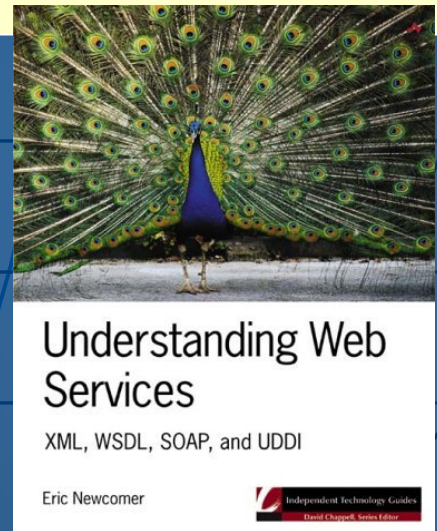


Data Services: An Evolution

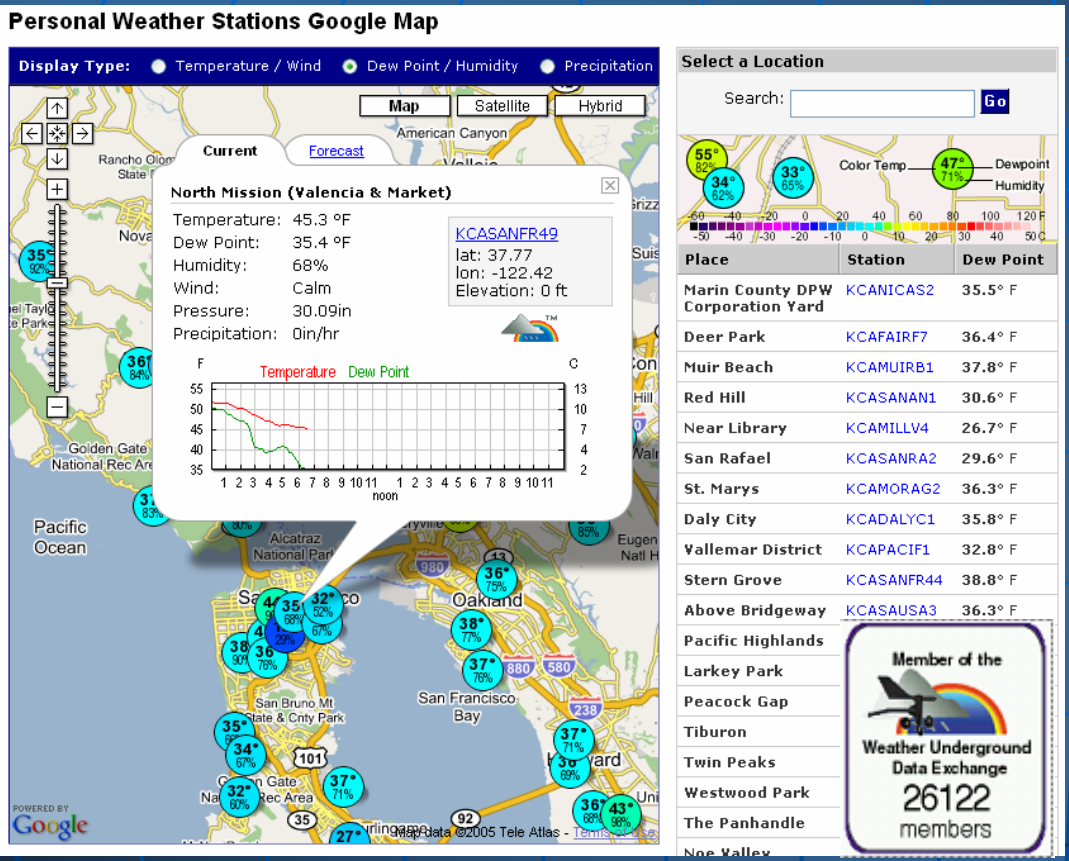
- An evolution from proprietary data sy wards more open standards-based web ser
- Data services must address co needs; This is not a distant vision but a reality t eCommerce.
  - KNOW YOUR USER!
- The transition to we s creates incredible opportunities but po y challenges.

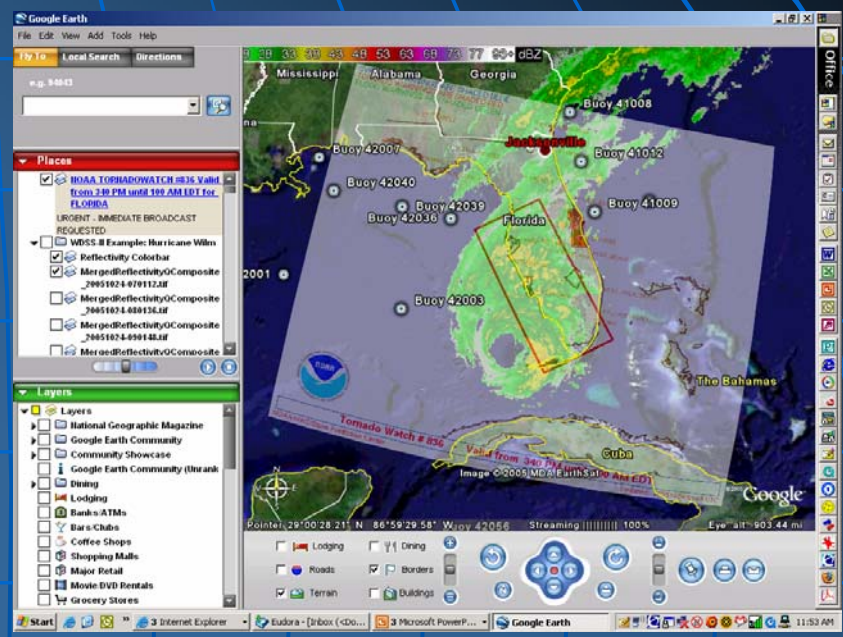
To borrow a line from GM: This is not your father's data service





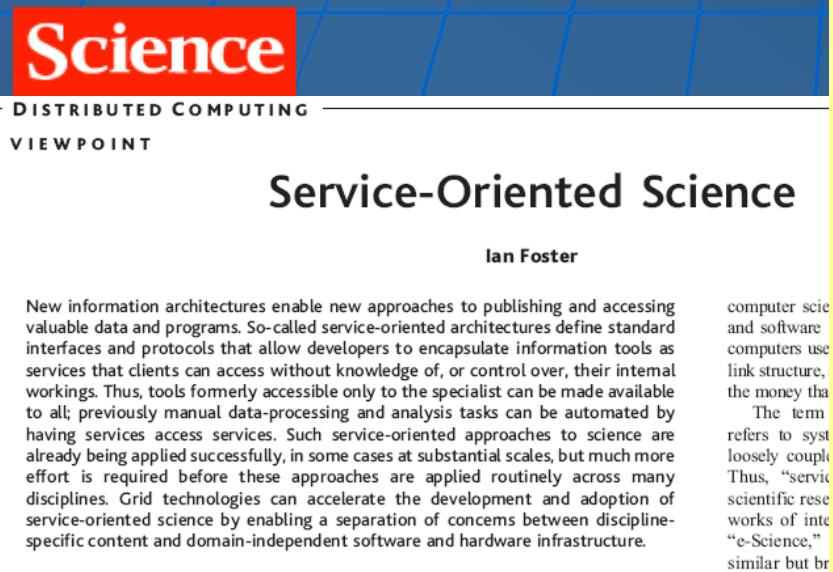
Google Maps & Google Earth





Challenge: How do we integrate these technologies with traditional data systems?

Service-Oriented Science



Science

DISTRIBUTED COMPUTING

VIEWPOINT

Service-Oriented Science

Ian Foster

New information architectures enable new approaches to publishing and accessing valuable data and programs. So-called service-oriented architectures define standard interfaces and protocols that allow developers to encapsulate information tools as services that clients can access without knowledge of, or control over, their internal workings. Thus, tools formerly accessible only to the specialist can be made available to all; previously manual data-processing and analysis tasks can be automated by having services access services. Such service-oriented approaches to science are already being applied successfully, in some cases at substantial scales, but much more effort is required before these approaches are applied routinely across many disciplines. Grid technologies can accelerate the development and adoption of service-oriented science by enabling a separation of concerns between discipline-specific content and domain-independent software and hardware infrastructure.

computer scie
and software
computers use
link structure,
the money tha
The term
refers to syst
loosely couple
Thus, "servic
scientific rese
works of inte
"e-Science,"
similar but br

Source: Ian Foster, Science, 6 May 2005

Web Services are self-contained, self-describing, modular applications that can be published, located, and invoked across the Web.

XML based Web Services are emerging as tools for creating next generation distributed systems that facilitate program-to-program interaction without the user-to-program interaction.

Besides recognizing the heterogeneity as a fundamental ingredient, these web services, independent of platform and environment, can be packaged and published on the internet as they can communicate with other systems using the common protocols.

Emerging web services standards are enabling much easier system-to-system integration.

Data Service Characteristics

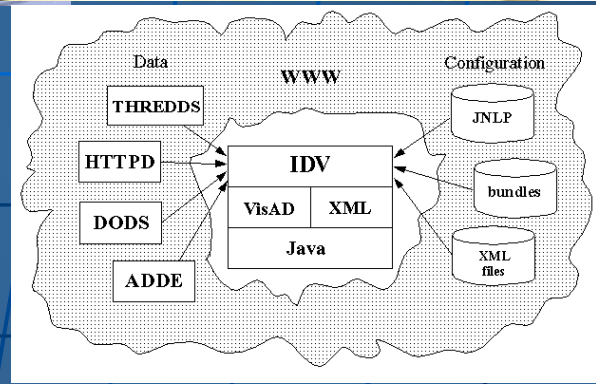
- User-friendly interface (e.g., portal)
- Transparency (format, protocol,...)
- Customization
- Server-side operations & analysis: subsetting, subsampling, mathematical operators, etc
- Aggregation of fields/variables, datasets, etc.
- Rich Metadata
- Integration across data types, formats, and protocols
- Intelligent client-server approaches
- Interoperability across services
- Flexibility and Scalability
- Ability to chain services
- Support an array of tools

Broad Data Categories

- Ideal Data Systems must provide a seamless, end-to-end services for accessing, utilizing and integrating data across the following data types:
  - Real-time data
  - Archived data
  - Field and Demonstration Project and Regional Campaign data
  - Episodic (Case Study)
  - Data from related disciplines (hydrology, oceanography, cryosphere, chemical and biosphere - soil, vegetation, canopy, evapotranspiration)
  - GIS databases

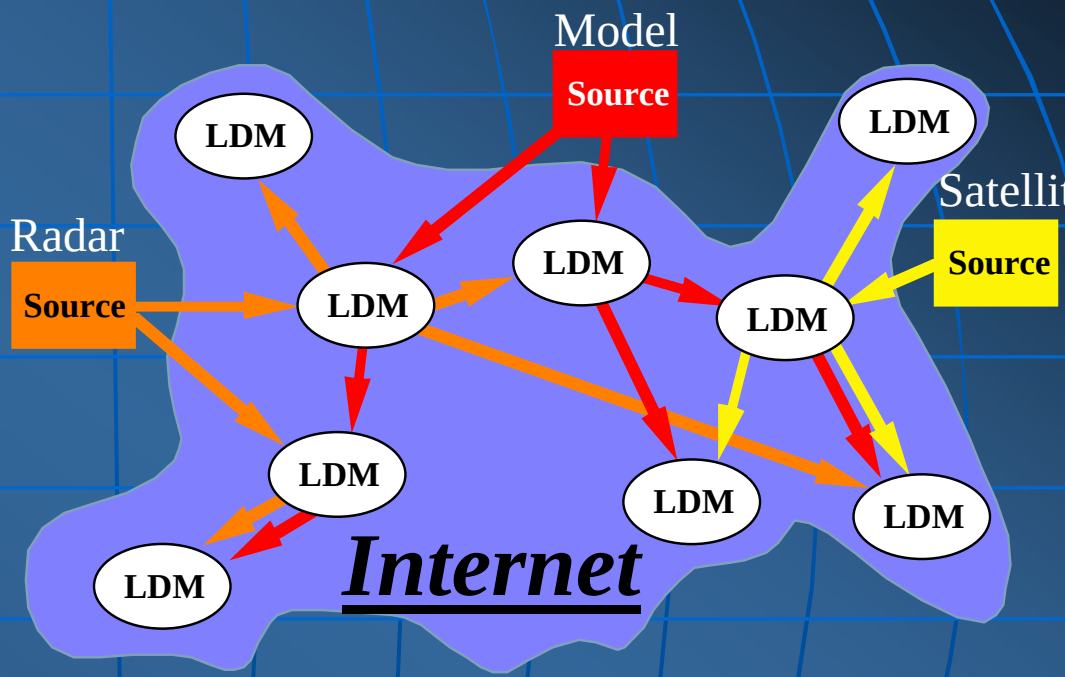
Ideal Data Services Will Need to use Hybrid Access Methods

- In the future, the 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 (push and pull systems)
- Notification services for data availability will play increasingly important role



Unidata IDD

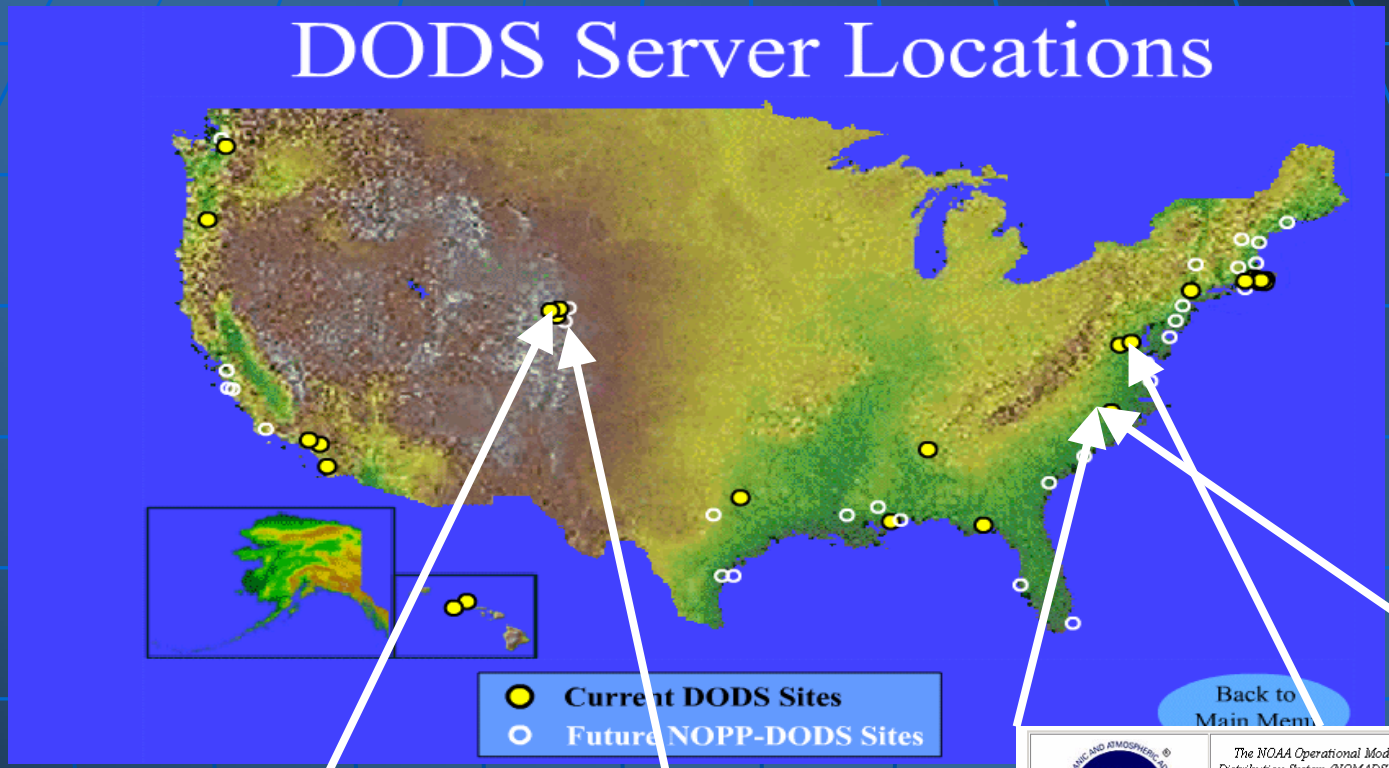
- Over 160+ sites are participating in Unidata Internet Data Distribution (IDD) system
- Approximately 3-4 GB of data injected/hour *from distributed sources*;
- Unidata IDD/LDM uses more of the Internet2 than any other *advanced application;*
- Approx. 15 Terabytes of data are transmitted via Internet 2 each week - 3-4% of their traffic).

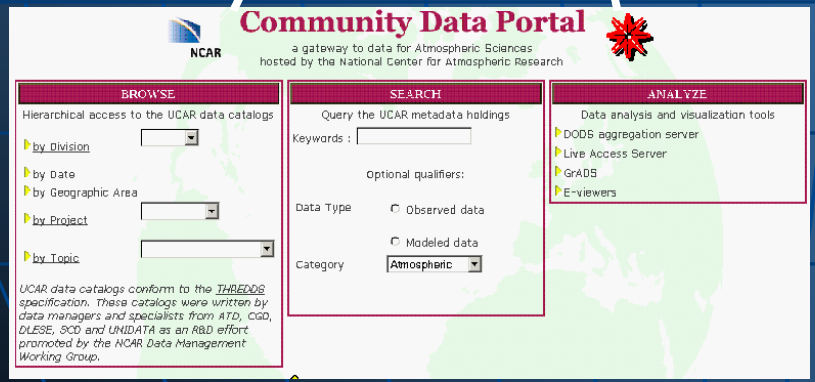


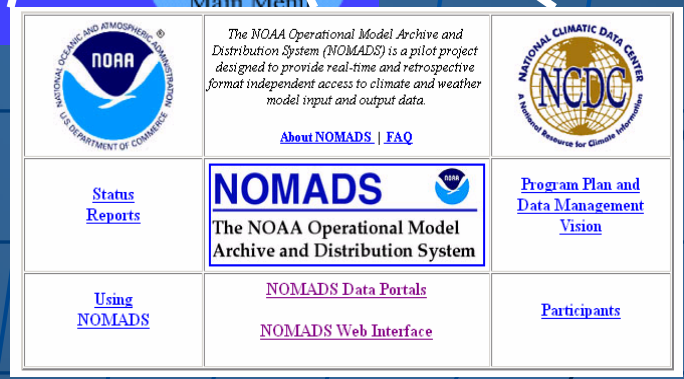
The LDM is now ranked #3 (behind HTTP and NNTP) in Internet 2 usage.

Last year, LDM surpassed FTP.

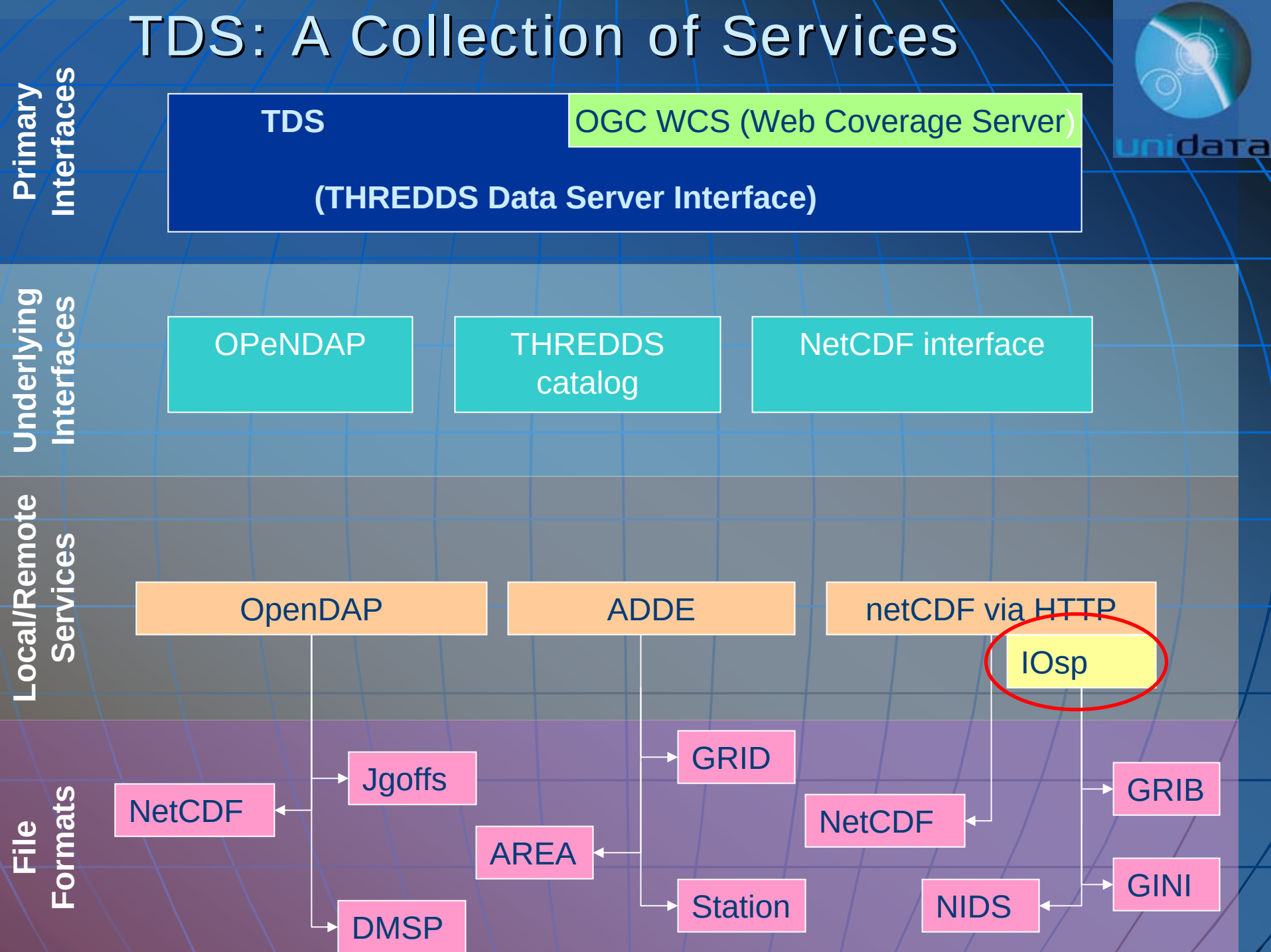
OPeNDAP/THREDDS Servers

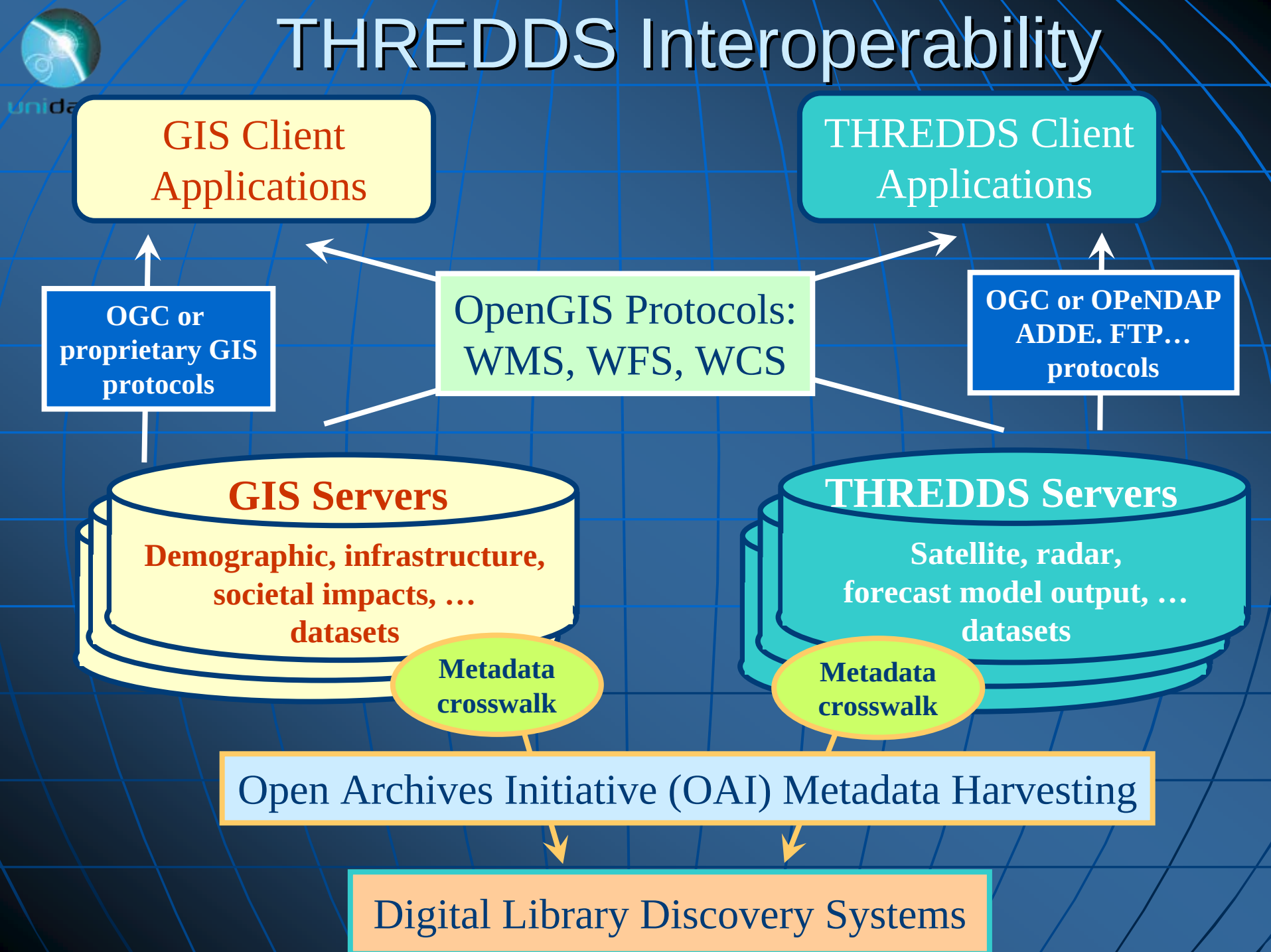






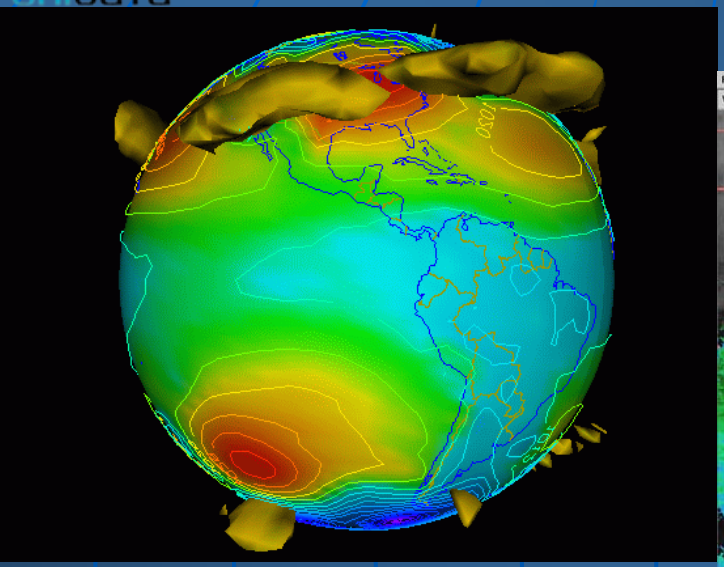
TDS: A Collection of Services




THREDDS Interoperability
GIS Client Applications
THREDDS Client Applications
OGC or proprietary GIS protocols
OpenGIS Protocols: WMS, WFS, WCS
OGC or OPeNDAP ADDE. FTP… protocols
GIS Servers
Demographic, infrastructure, societal impacts, … datasets
THREDDS Servers
Satellite, radar, forecast model output, … datasets
Metadata crosswalk
Metadata crosswalk
Open Archives Initiative (OAI) Metadata Harvesting
Digital Library Discovery Systems

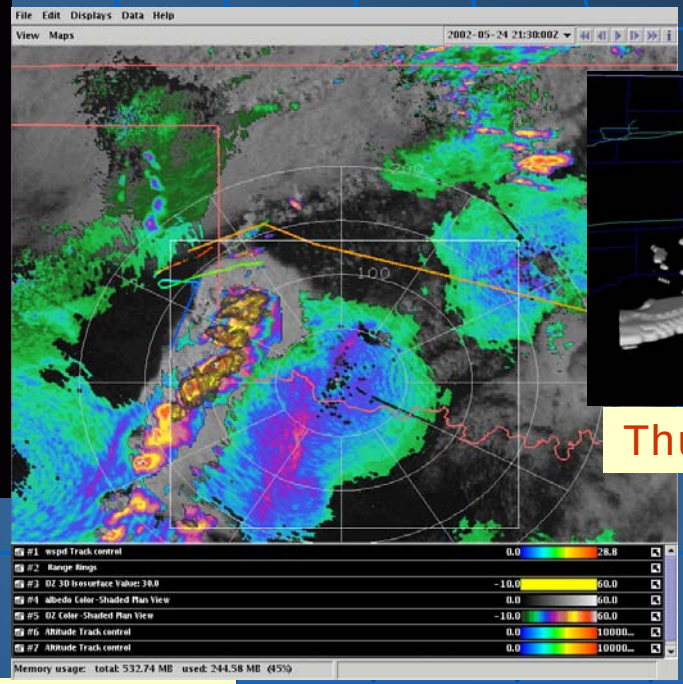
Remote Visualizations Using the IDV

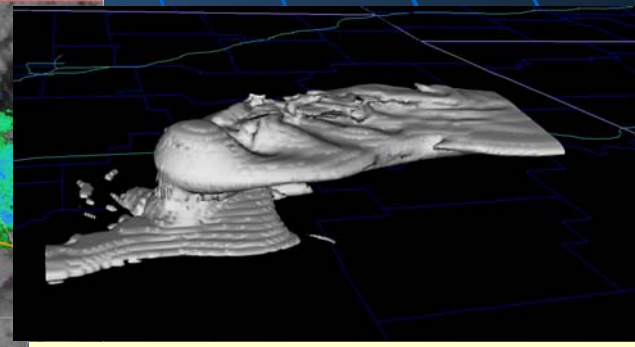
unidata



Sea-level Pressure and Upper-level Jet

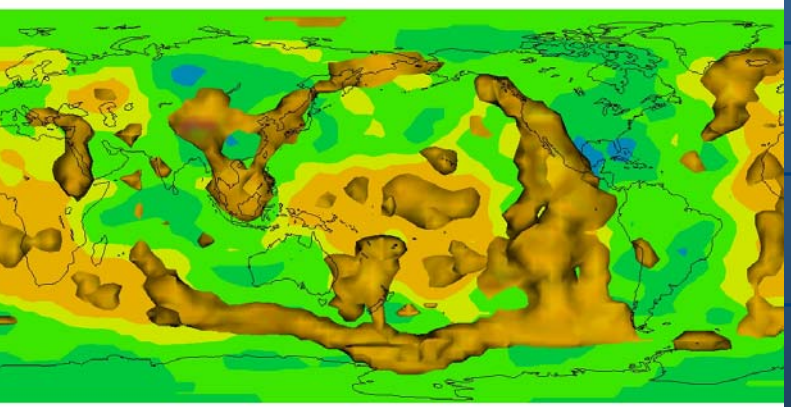
IDV in IHOP

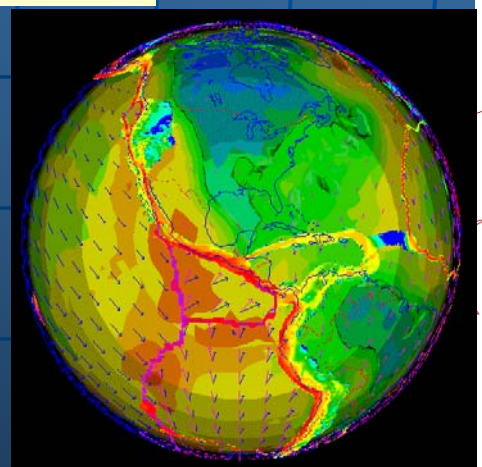




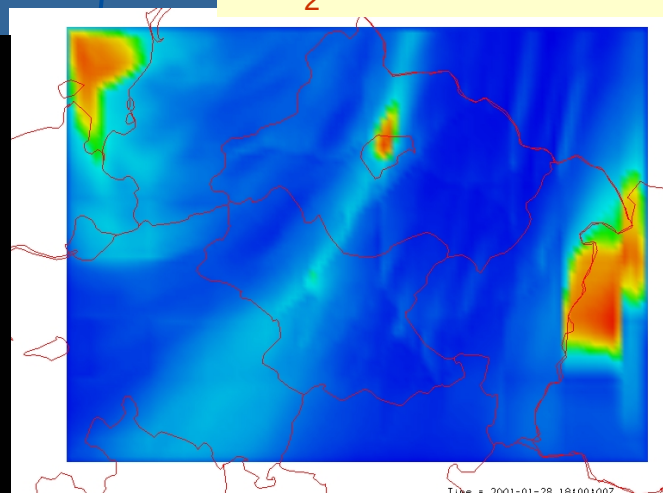
Thunderstorm Simulation

Mantle Tomography





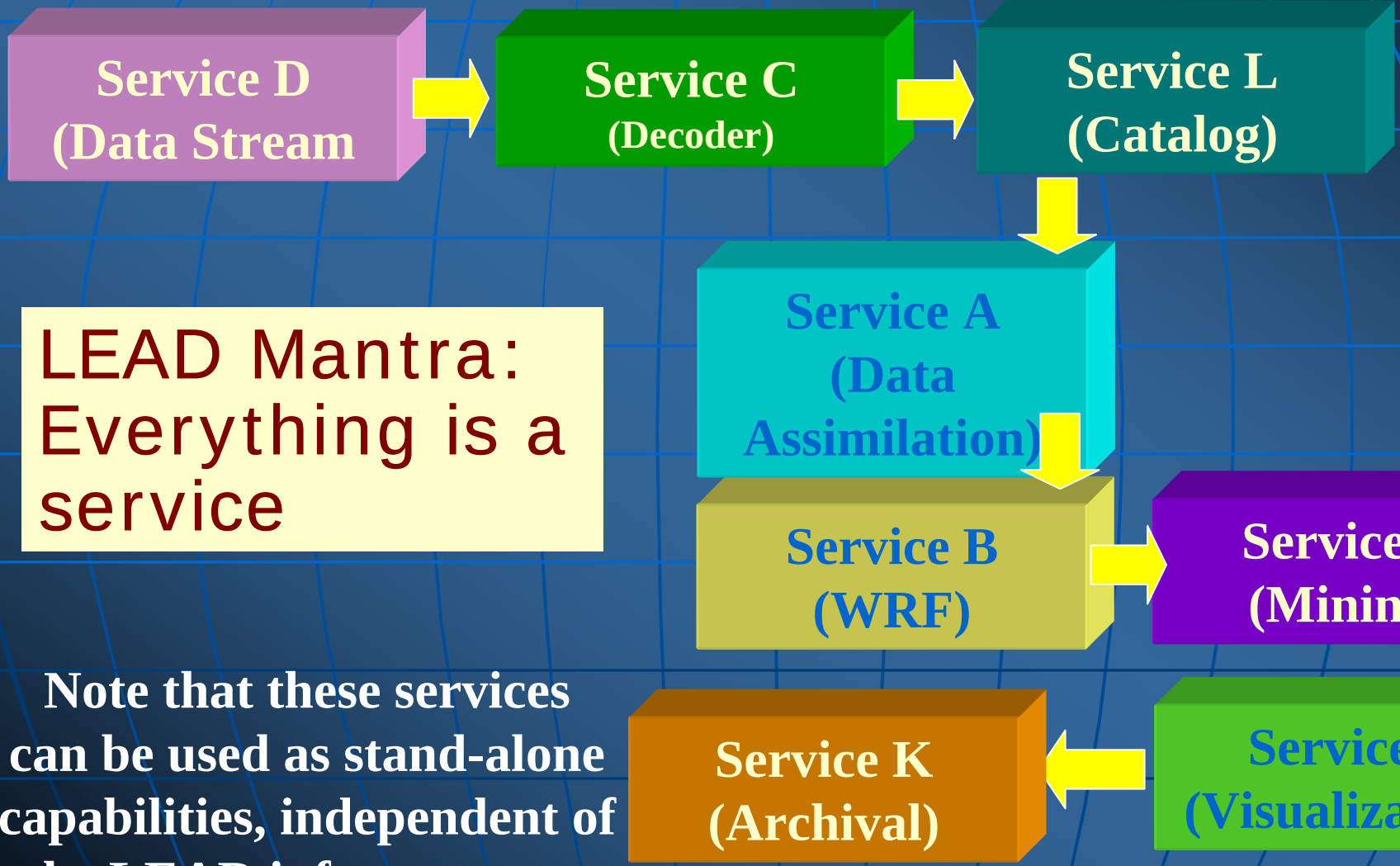
NO_2 concentration



Grid Computing

- *Refers to an infrastructure that enables the integrated, collaborative use of computers, networks, databases, and scientific instruments owned and managed by distributed organizations.*
- The terminology originates from a crude analogy to the electrical power grid; most users do not care about the details of power generation, distribution, etc, but your appliances work when you plug them into the socket.
- *Grid applications often involve large amounts of data and/or computing and require secure resource sharing across organizational boundaries.*
- Grid services are essentially web services running in a Grid framework.

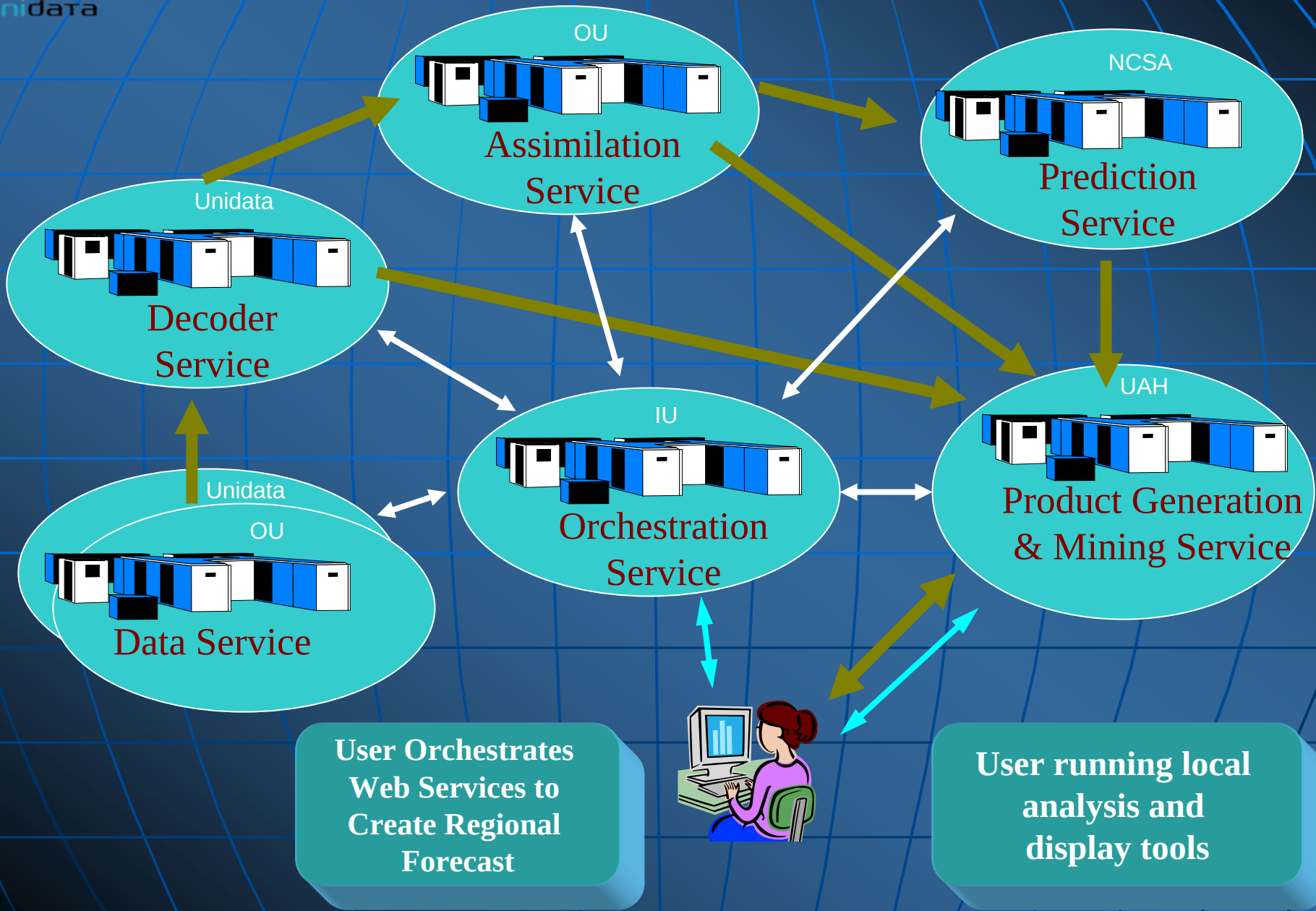
Solve Problems by Linking Services Together in Workflows



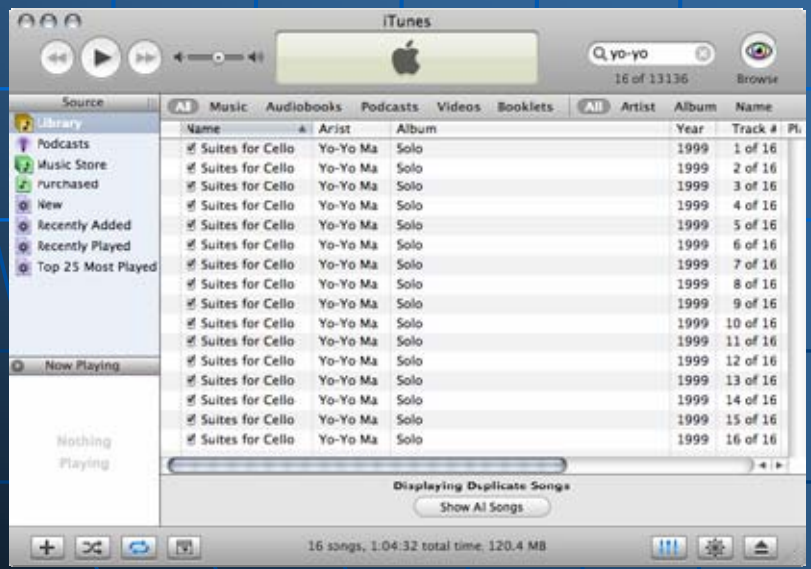
LEAD Mantra: Everything is a service

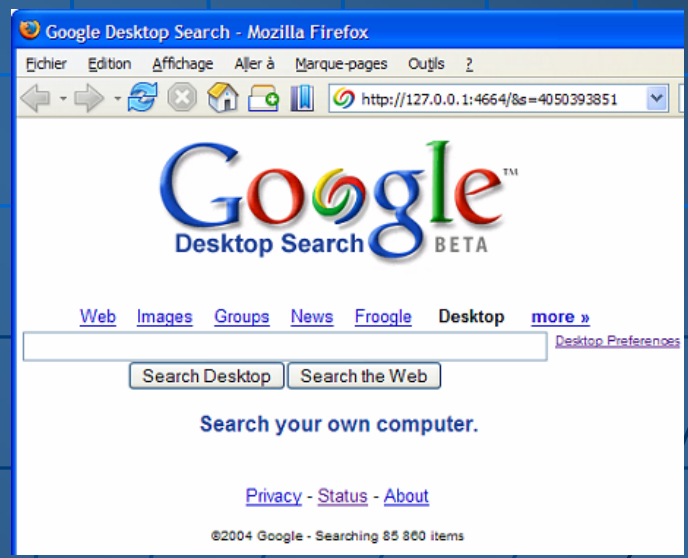
Note that these services can be used as stand-alone capabilities, independent of the LEAD infrastructure (e.g., portal)

LEAD: Data Services for NWP



Seamless access to and sharing of data anywhere
- From device, desktop, local and institutional libraries to archives anywhere on the Internet



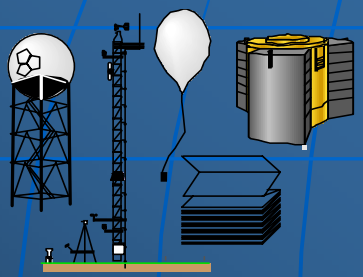


We can learn a lot from these examples

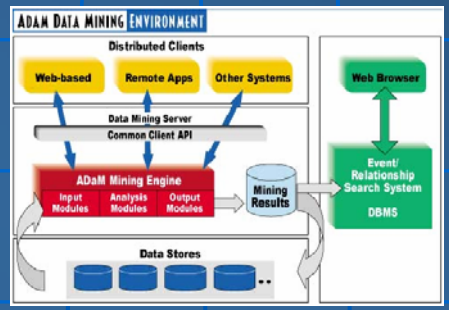
Web Services and E-Commerce

Tremendous strides have been made in the commercial space by companies like Amazon, EBay, Yahoo & Google to integrate diverse content with data

Dynamically Adaptive Data Services: Next Frontier



Streaming Observations



Data Mining

Storms Forming

Forecast Model

On-Demand Grid Computing

Monitoring and notification services play a crucial role