



unidata

# 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

---



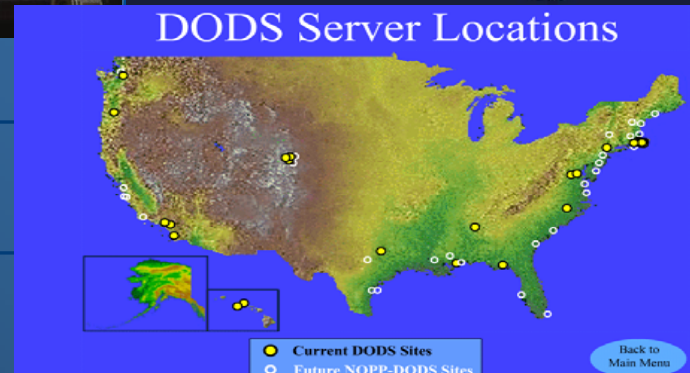
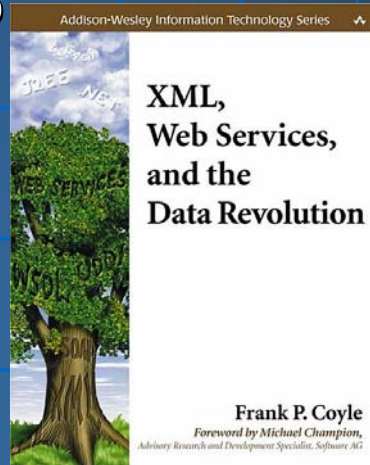
- To provide data to the user's application (analysis/visualization tool) in a transparent, consistent, readily useable form;
  - Deliver the right data in the right format at the right time for ready use
- Finding the right data is almost as important as providing the data
- Users don't care as much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

Source: Peter Cornillon (2003)



# Technology Evolution: Enabling a New Generation of Data Services

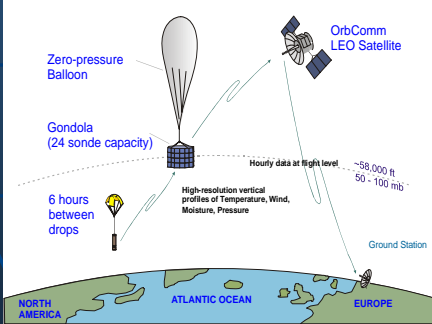
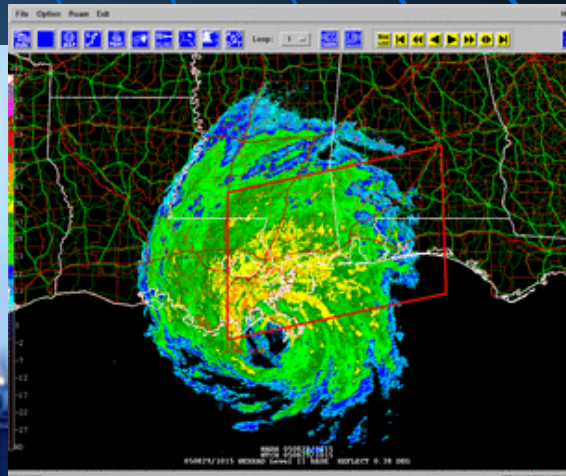
- Internet & the World Wide Web
- Commodity microprocessors
- Object-oriented programming
- Open standards
- Web services
- Extensible Markup Language (XML)
- Global, high-bandwidth and wireless networks
- Digital libraries
- Collaboratories
- Grid Computing/e-Science
- Data Portals and Federated, distributed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





# Opportunity to Solve Real-world Problems

unidata

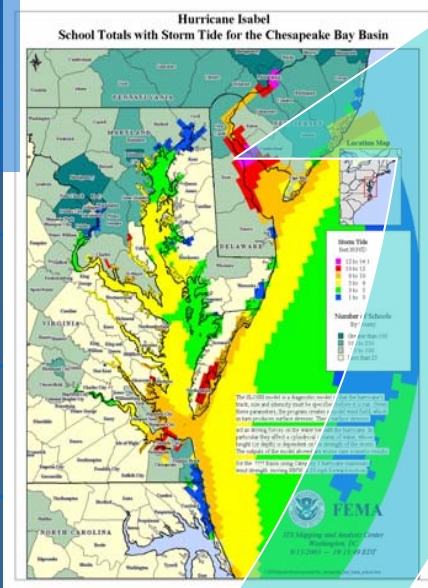


# End-to-end Data Services

**GIS  
Integration**

**ensemble  
predictions**

**Coastal Environments**



**Decision  
Support**

**Need integrated services**







# Data Services: An Evolution

- An evolution from proprietary data systems towards more open standards-based web services.
- Data services must address core business needs; This is not a distant vision but a reality for e-Commerce.
  - KNOW YOUR USER!
- The transition to web services creates incredible opportunities but presents many challenges.

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



## Understanding Web Services

XML, WSDL, SOAP, and UDDI

Eric Newcomer





# Google Maps & Google Earth

## Personal Weather Stations Google Map

Display Type:  Temperature / Wind  Dew Point / Humidity  Precipitation

Map Satellite Hybrid

Current Forecast

**North Mission (Valencia & Market)**

Temperature: 45.3 °F  
 Dew Point: 35.4 °F  
 Humidity: 68%  
 Wind: Calm  
 Pressure: 30.09in  
 Precipitation: 0in/hr

[KCASANFR49](#)  
 lat: 37.77  
 lon: -122.42  
 Elevation: 0 ft

Temperature Dew Point

Map showing various weather stations with temperature and humidity data points.

Select a Location

Search:  Go

Place	Station	Dew Point
Marin County DPW Corporation Yard	KCANICAS2	35.5° F
Deer Park	KCAFAIRF7	36.4° F
Muir Beach	KCAMUIRB1	37.8° F
Red Hill	KCASANAN1	30.6° F
Near Library	KCAMILLV4	26.7° F
San Rafael	KCASANRA2	29.6° F
St. Marys	KCAMORAG2	36.3° F
Daly City	KCADALYC1	35.8° F
Vallemar District	KCAPACIF1	32.8° F
Stern Grove	KCASANFR44	38.8° F
Above Bridgeway	KCASAUSA3	36.3° F
Pacific Highlands		
Larkey Park		
Peacock Gap		
Tiburon		
Twin Peaks		
Westwood Park		
The Panhandle		
Noe Valley		

Member of the  
 Weather Underground  
 Data Exchange  
 26122  
 members

Google Earth

File Edit View Add Tools Help

Local Search directions

Places

- NOAA TOPHADOWATCH #336 Valid from 2:30 PM until 1:00 AM EDT for FLORIDA
- URGENT - IMMEDIATE BROADCAST REQUESTED
- WDSS - Example: Hurricane Wilm
- Reflectivity Colorbar
- Merged Reflectivity Composite
- Merged Reflectivity Composite
- Merged Reflectivity Composite
- Merged Reflectivity Composite

Layers

- Layers
- National Geographic Magazine
- Google Earth Community
- Community Showcase
- Google Earth Community (thread)
- Dining
- Lodging
- Bars/ATMs
- Bars/Chubs
- Coffee Shops
- Shopping Malls
- Major Retail
- Movie/DVD Rentals
- Grocery Stores

Tomado Watch #336

Point: 29°00'28.211" N 86°59'29.581" W Proj: 42056 Streaming: 100% Eye alt: 903.44 mi

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 science and software computers use link structure, the money that The term refers to systems loosely coupled. Thus, "service scientific research works of interest "e-Science," similar but broader

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



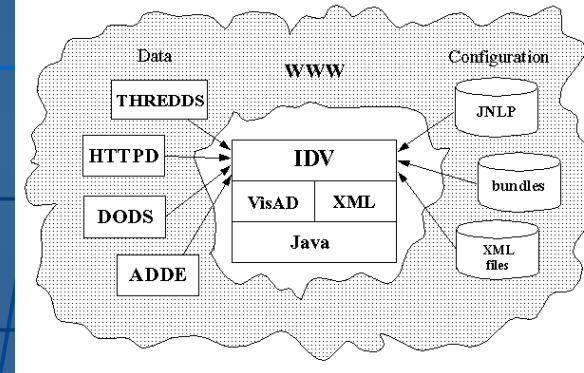
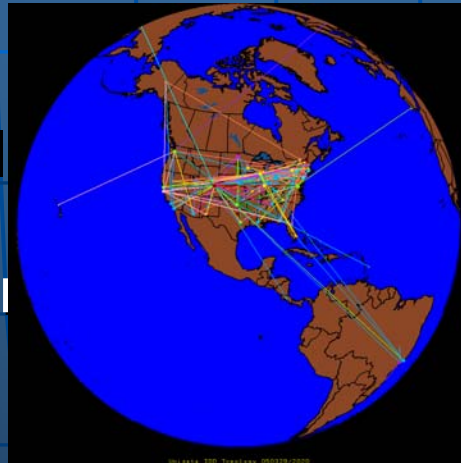
# Broad Data Categories

- 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

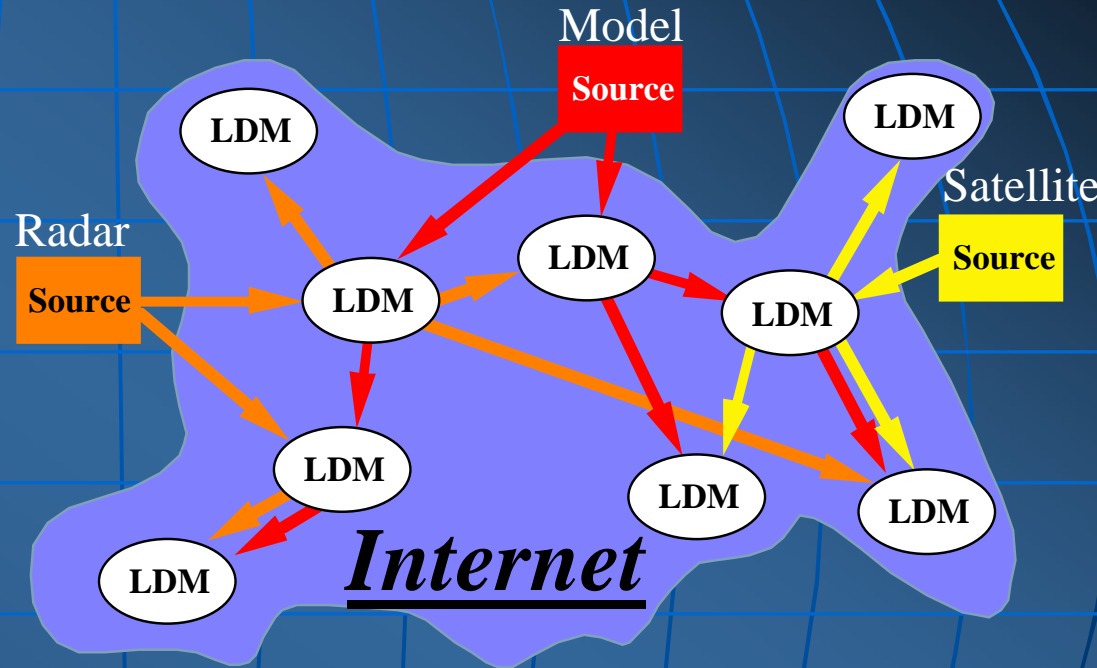
- In the future, the community will need a hybrid solution that couples a satellite-based reception system with a terrestrial, Internet-based data access system
- Both local and remote data access mechanisms will be required (push and pull systems)
- Notification services for data availability will play 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).



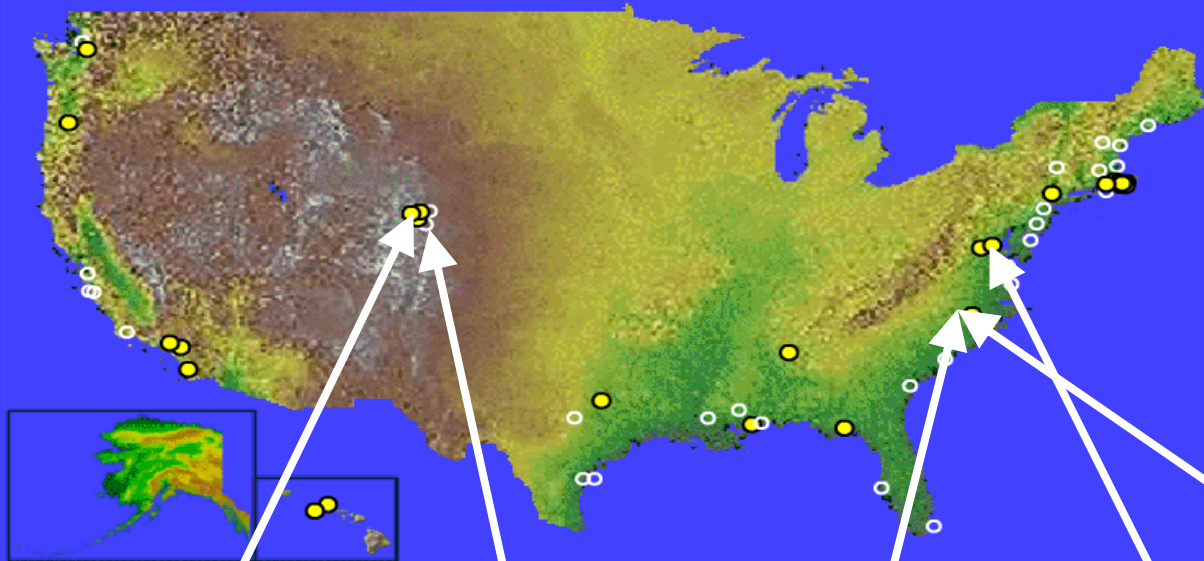
The LDM is now ranked #3 (behind HTTP and NNTP) in Internet 2 usage.

Last year, LDM surpassed FTP.



# OPeNDAP/THREDDS Servers

## DODS Server Locations



 **Current DODS Sites**  
 **Future NOPP-DODS Sites**

[Back to Main Menu](#)

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

**BROWSE**

Hierarchical access to the UCAR data catalogs

- by Division
- by Date
- by Geographic Area
- by Project
- by Topic

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

**SEARCH**

Query the UCAR metadata holdings

Keywords :

Optional qualifiers:



Data Type  Observed data  Modeled data

Category

**ANALYZE**

Data analysis and visualization tools

- DODS aggregation server
- Live Access Server
- GrADS
- E-viewers

	<p>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><a href="#">About NOMADS</a>   <a href="#">FAQ</a></p>	
<p><a href="#">Status Reports</a></p>	<p><b>NOMADS</b></p> <p>The NOAA Operational Model Archive and Distribution System</p>	<p><a href="#">Program Plan and Data Management Vision</a></p>
<p><a href="#">Using NOMADS</a></p>	<p><a href="#">NOMADS Data Portals</a></p> <p><a href="#">NOMADS Web Interface</a></p>	<p><a href="#">Participants</a></p>

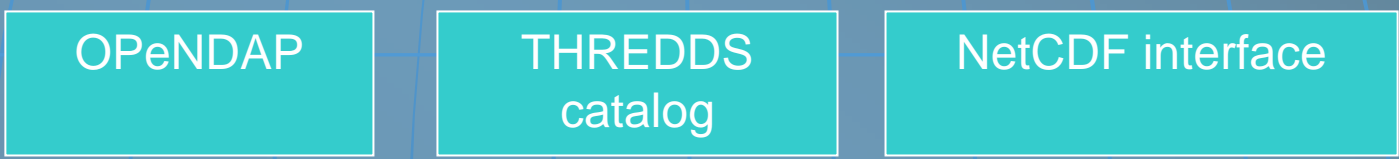
# TDS: A Collection of Services



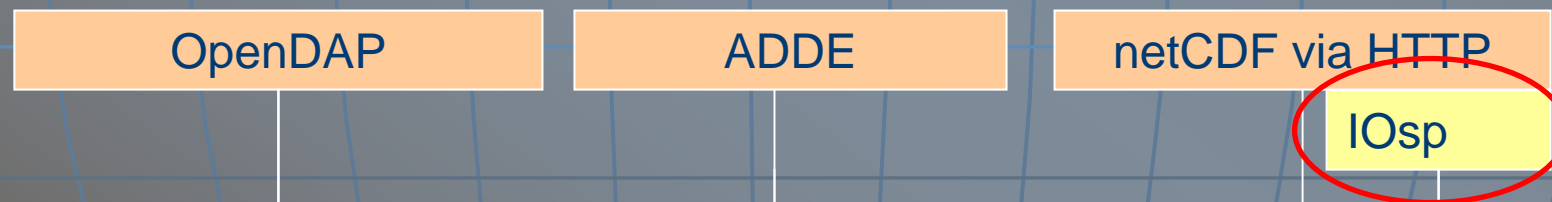
Primary Interfaces



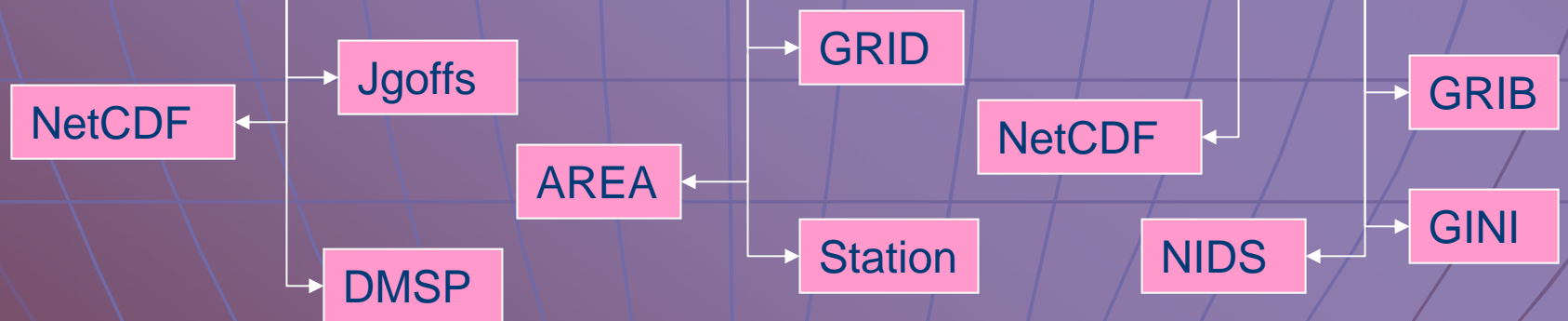
Underlying Interfaces



Local/Remote Services

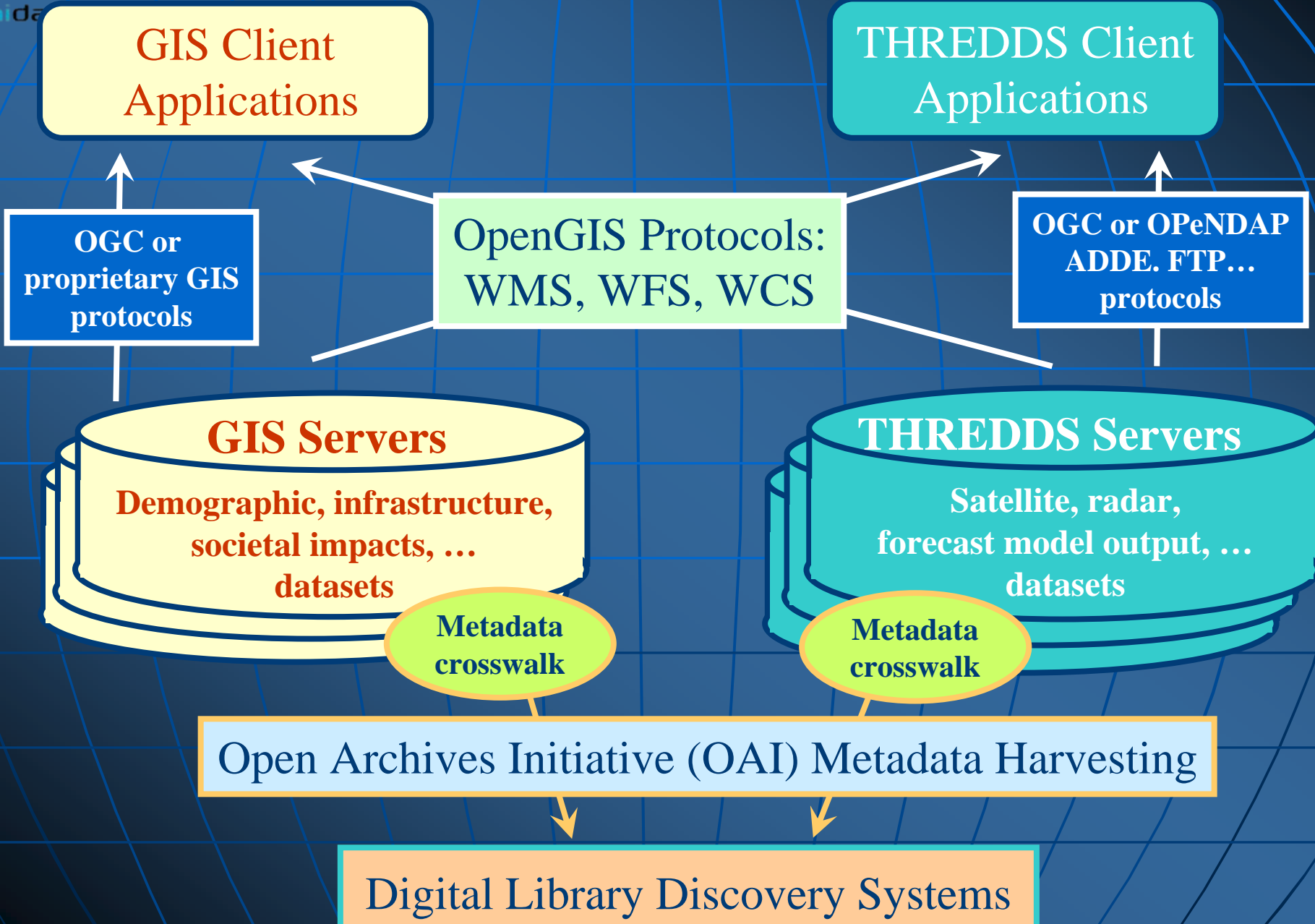


File Formats





# THREDDS Interoperability



GIS Client Applications

THREDDS Client Applications

OpenGIS Protocols: WMS, WFS, WCS

OGC or proprietary GIS protocols

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

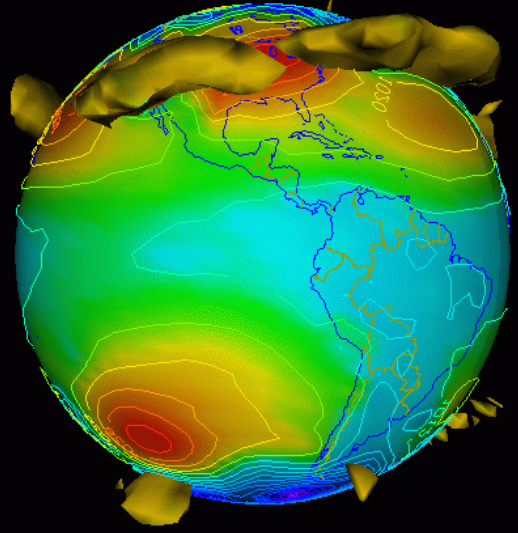
Digital Library Discovery Systems



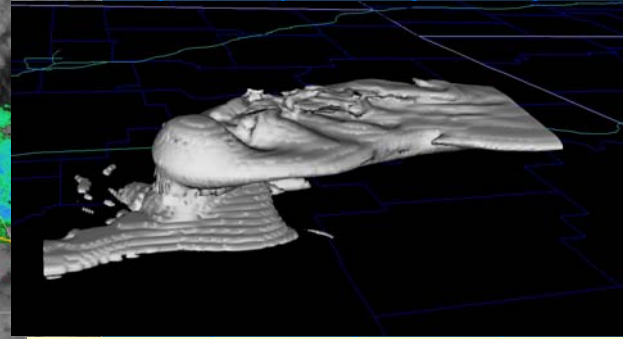
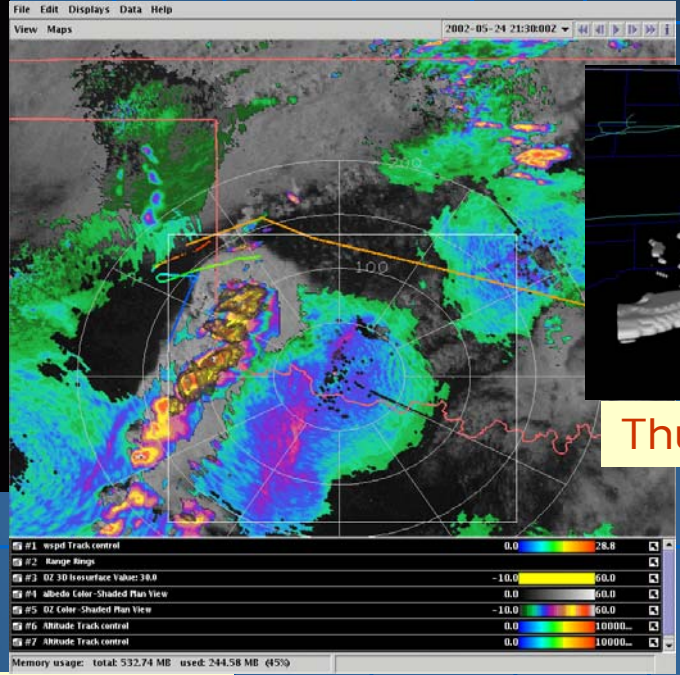
# Remote Visualizations Using the IDV



## IDV in IHOP

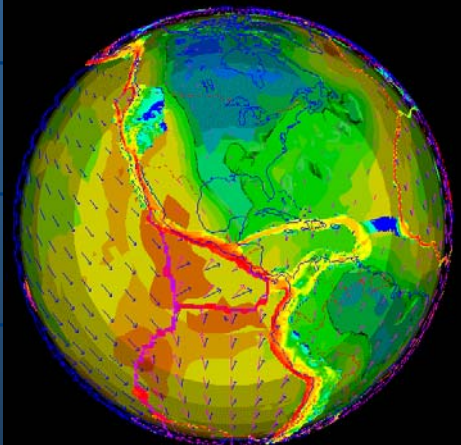
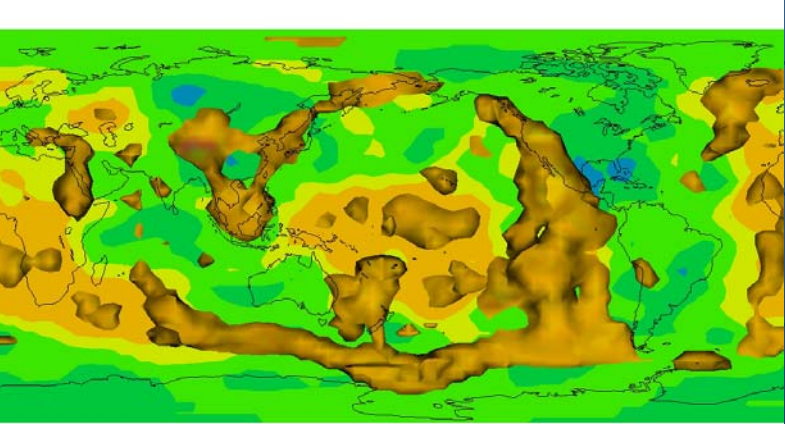


Sea-level Pressure and Upper-level Jet

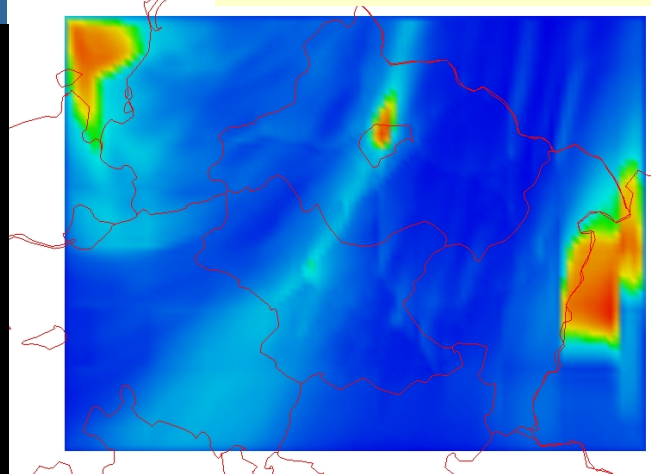


Thunderstorm Simulation

## Mantle Tomography



## NO<sub>2</sub> concentration





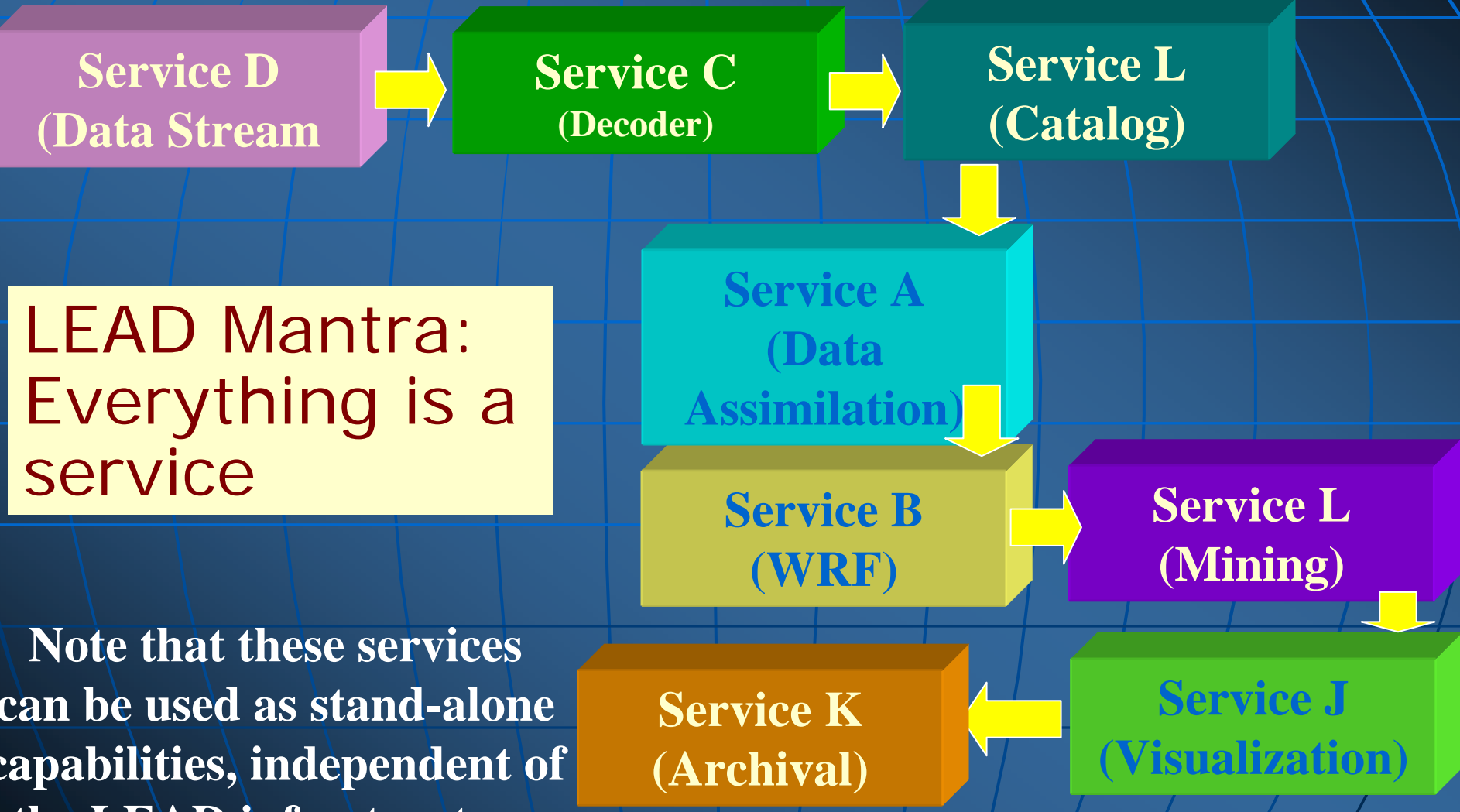
# Grid Computing

- *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

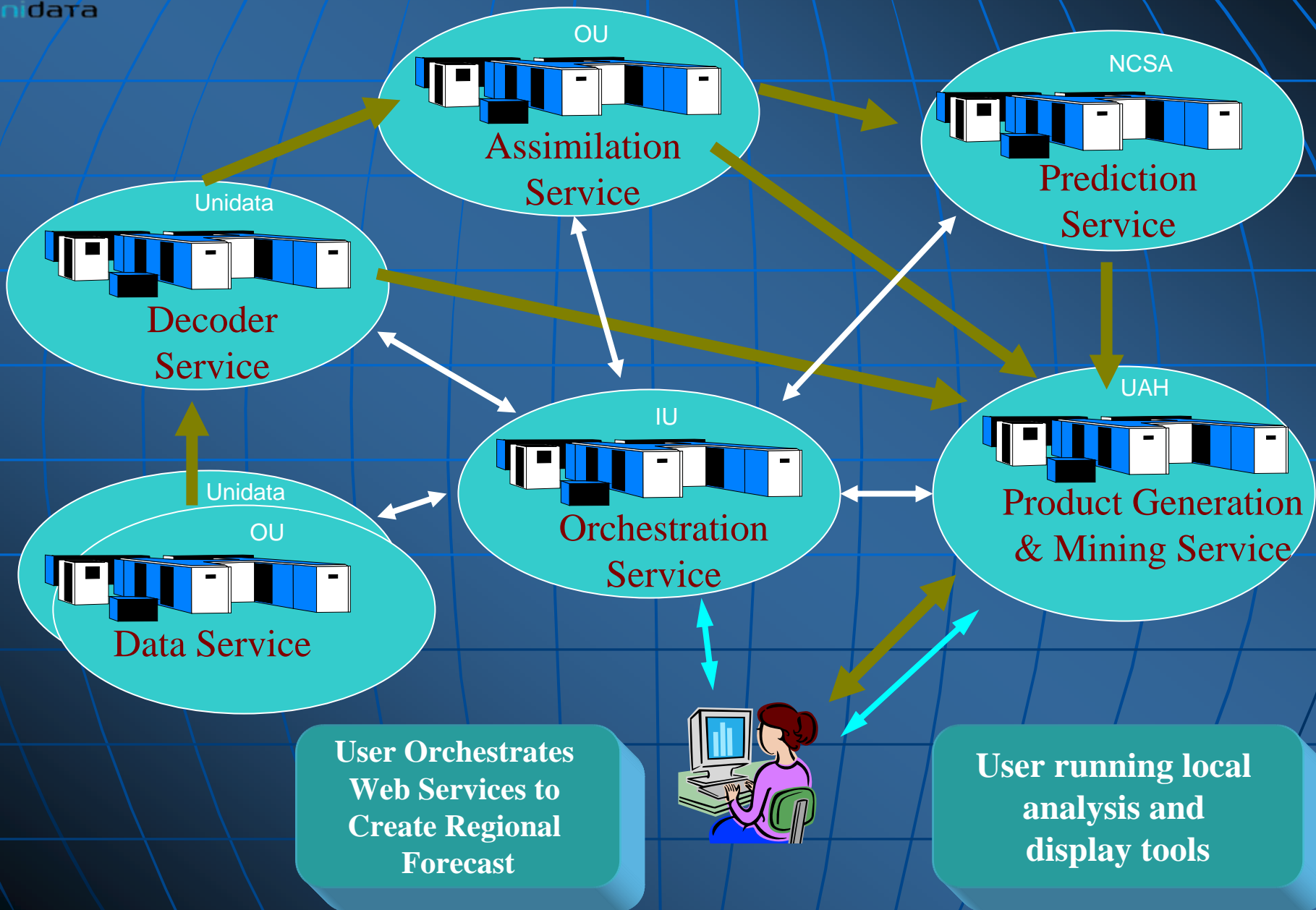


**LEAD Mantra:**  
Everything is a service

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



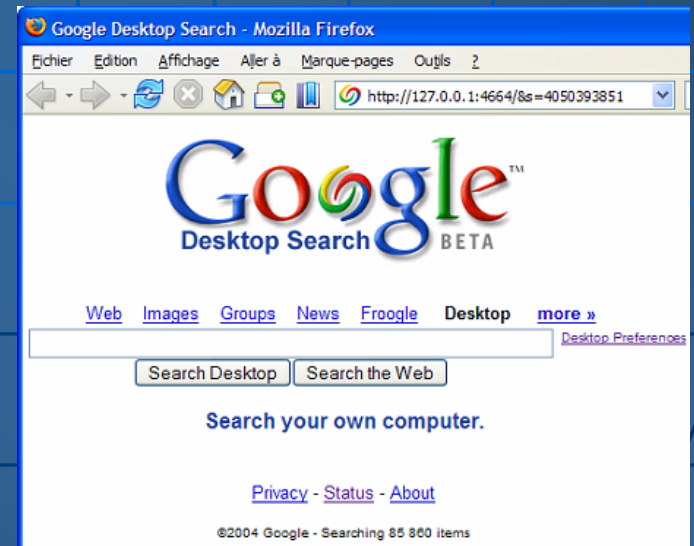
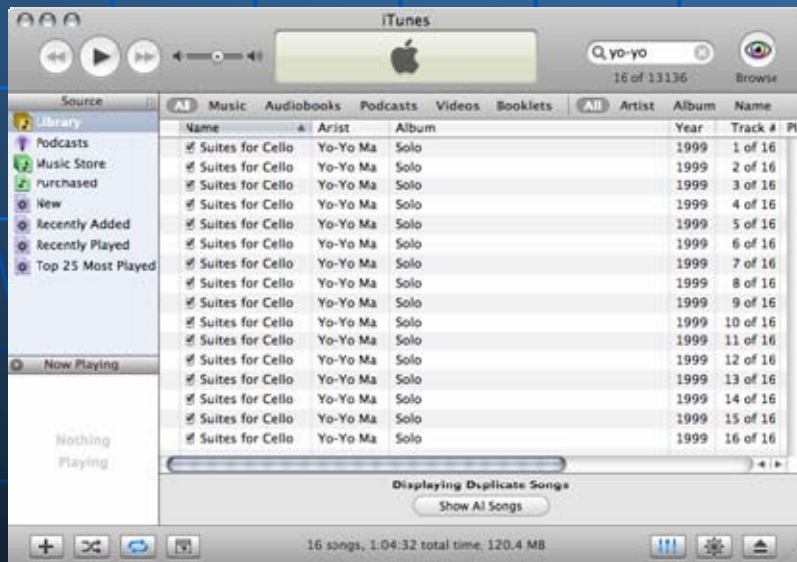
# LEAD: Data Services for NWP





# 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

amazon.com | Mohan's Store | Electronics | See All 32 Product Categories | Your Account | Cart | Wish List | Help |

Browse Brands & Products | Top Sellers | Camera & Photo | Computers | Software | Audio & Video | Today's Deals | Outlet, Used & Refurbished

Search Electronics

Join Amazon Prime and ship Two-Day for free and Overnight for \$3.99.

**ITEM INFORMATION**  
Explore this item  
[buying info](#)  
[technical data](#)  
[customer reviews](#)  
[product description](#)  
[accessories](#)  
 Help us help others  
[Submit a manual](#)  
 Share your thoughts  
[write a review](#)  
[write a So You'd Like to... guide](#)  
[tell a friend about this item](#)

**RATE THIS ITEM**  
  
 Not interested  
 I own it

**Logitech 961400-0403 Quickcam for Notebooks Deluxe**  
Other products by [Logitech](#)

 **SALE**  
 List Price: ~~\$59.99~~  
 Price: **\$47.49** & this item ships for **FREE** with Super Saver Shipping. [See details](#)  
 You Save: **\$12.50 (21%)**  
**Rebate forms for recent purchases**  
 Availability: Usually ships within 24 hours. Ships from and sold by Amazon.com.

[See larger image and other views](#) **22 used & new** available from **\$32.99**  
[Share your own customer images](#)

Manufacturers, merchants, and enthusiasts: [Submit a product manual](#) for this item.

Get peace of mind and protect your purchase with a service contract today. To order, click checkbox, then click Add to Cart: **Select a Plan**

- **1-Year Replacement Plan for Electronics Products for only \$5.99**

[Technical Data](#) | [Customer Reviews](#) | [Product Description](#) | [Accessories](#)

**Customers who viewed this item also viewed**

- [Logitech Quickcam for Notebooks](#) Other products by [Logitech](#)
- [Creative Labs Webcam Notebook Camera with Clip](#) Other products by [Creative Labs](#)
- [Logitech Quickcam Fusion \(961403-0403\)](#) Other products by [Logitech](#)
- [Logitech QuickCam Communicate STX](#) Other products by [Logitech](#)

**READY TO BUY?**  
**Amazon.com**  
 Price: **\$47.49**  
 Availability: Usually ships in 24 hours  
 **or**  
[Sign in](#) to turn on 1-Click ordering.  
 A9.com users **save 1.57%** on Amazon. [Learn how.](#)

**MORE BUYING CHOICES**

**antonline\_com**  
 Price: **\$48.49**  
 Availability: Usually ships in 1-2 business days

**J&R Music and Computer World**  
 Price: **\$49.99**  
 Availability: Usually ships in 1-2 business days

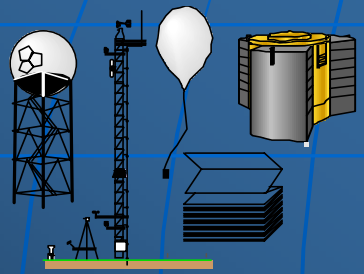
**TigerDirect**  
 Price: **\$49.99**  
 Availability: Usually ships in 1-2 business days

**22 used & new** from **\$32.99**  
 Have one to sell?

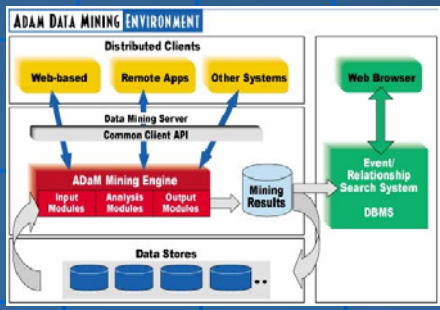
Tremendous strides have been made in the commercial space by companies like Amazon, eBay, Yahoo & Google to integrate diverse content with data



# Dynamically Adaptive Data Services: Next Frontier



Streaming Observations



Data Mining



Forecast Model



On-Demand Grid Computing



Monitoring and notification services play a crucial role