Unidata: A Program of the Community, by the Community, and for the Community

National Weather Center Seminar
University of Oklahoma
Norman
11 September 2007

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.
Unidata: A Benevolent 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...
Unidata in a Nutshell

We

Facilitate data access

Provide tools for data access, management, analysis and visualization

Provide comprehensive support

Engage in community building and advocacy

Although the principal focus of our activities is on real-time weather data provision, 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.
An Independent Assessment

- Earlier this year, we completed an independent Metrics and Assessment study on Unidata’s impact on education and research in the community;

- The study included many components and looked at many aspects of Unidata
  - Unidata has been truly transformational and is irreplaceable
  - Unidata is a model facility for other communities like hydrology to develop
Over 250 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.
We provide over 30 data streams (surface & upper-air obs., radar, satellite imagery, model output, lightning data, ACARS, NWS bulletins, etc.)
Working definition: an integrated system of interconnected computation/communication/information elements that supports a range of applications

Cyberinfrastructure is the means; “e-Science” is the result
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;
Ray Pierrehumbert, noted climate scientist from the University of Chicago: “I think one mustn't discount a breakthrough of a technological sort in AR4 though: The number of model runs exploring more of scenario and parameter space is vastly increased, and more importantly, it is available in a coherent archive to the full research community for the first time. The amount of good science that will be done with this archive in the next several years is likely to have a significant impact on our understanding of climate. --raypierre
Principal Drivers Shaping Unidata’s Work

- Science
- Education
- Technology
- Social & Organizational Evolution

What does the community need from Unidata?
What is our niche and what are our 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
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.
These days, few traditional meteorology programs exist. There are notable exceptions, including here at OU, FSU, and Penn State.
Education Drivers

- A “holistic” Earth-system science approach to education

- Active, student-centered learning. i.e., learning science by doing science
  - Observations
  - Tools (models, visualization)
  - Discovery
Technology Trends

- Internet & the World Wide Web
- Multi-core processors
- Object-oriented programming
- Open standards
- Web services and Web 2.0
- Global, high-bandwidth and wireless networks
- Digital libraries
- Virtual organizations
- Grid computing/e-Science
- Social networks
- Blogs
- Podcasts
- GIS
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 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
Need for End to End Information Services

GIS Integration

Coastal Environments

Ensemble Predictions

Emergency Response

Need integrated services
We have just developed a new Strategic Plan for Unidata, to be released to the community next week.
**Mission:** To provide the data services, tools, and cyberinfrastructure leadership that advance Earth system science, enhance educational opportunities, and broaden participation.

**Vision:** Unidata is the premier provider of seamless and comprehensive data services that help solve complex environmental problems facing science and society.

- Implicit in this vision is recognition that Unidata will provide a broad array of end-to-end and well-integrated data services and contribute cyberinfrastructure that benefits a broader Earth system science community.

- We align our mission and goals with those of the NSF, our primary sponsor.

- Our goals are congruent with national and community priorities, and will be anchored by the evolving community’s needs and the program’s core competencies.
Strategic Focus Areas

- Community
- Data Services
- Tools
- Communication and Support
- Cyberinfrastructure
- Leadership
- Diversity
Data Services Will Need to Provide Multiple Access Methods

- Given the very high data rates from each GOES-R satellite, 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.
Thematic Real-time Environmental Distributed Data Servers (THREDDS)

User applications: e.g., McIDAS, IDV, LAS, IDL, MatLab...

DLESE
Digital Library for Earth-System Education

Hydrology Data, e.g.

Geophysical Data, e.g.

Satellite Imagery...

OpenDAP, ADDE, & FTP protocols

IDD

DL interchange protocol

• Combines IDD “push” with several forms of “pull” and DL discovery

• About 25 data providers are partners in THREDDS

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

Connecting People with Documents and Data
THREDDS Interoperability

GIS Client
Applications

OGC or
proprietary GIS
protocols

GIS Servers
Demographic, infrastructure, societal impacts, … datasets

Metadata
crosswalk

THREDDS Client
Applications

OGC or OPeNDAP
ADDE, FTP…
protocols

THREDDS Servers
Satellite, radar, forecast model output, … datasets

Metadata
crosswalk

Open Archives Initiative (OAI) Metadata Harvesting

Digital Library Discovery Systems

OpenGIS Protocols:
WMS, WFS, WCS

GIS Server

THREDDS Server
Common Data Model – TDS

TDS

OGC WCS (Web Coverage Server)

(THREDDS Data Server Interface)

Primary Interfaces

OPeNDAP

THREDDS catalog

NetCDF interface

Underlying Interfaces

OpenDAP

ADDE

netCDF via HTTP

Local/Remote Services

NetCDF

Jgoffs

AREA

DMSP

GRID

Station

NIDS

File Formats

IOsp

GRIB

GINI
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.
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.)
LEAD is connecting the traditional Unidata community (small computers) to the High-performance computing world (TeraGrid)
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.
Let us look at how faculty at a typical university share information with their students and colleagues.

Now let’s ask ourselves how we can integrate THREDDS & IDV to extend this concept.

Voila! Next Generation Case Studies!!!!
Vision: To leverage and bring together the strengths of Unidata and COMET programs to develop a new generation of case studies that are dynamic, integrated, and interactive.

- These case studies will include not only datasets of weather events, but they will, where appropriate, integrate relevant educational modules.
- These new case studies will be “living” or dynamic, allowing for the community to augment and add value to existing case studies by contributing related observations, analyses, educational, curricular and multimedia materials, and other views on the case.

An important element of that vision is to build an easy to use framework in which community members can develop and build on future case studies.
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
Importance of International Partnerships

Hurricane Catarina, 2004

SALLJEX, 2003
- 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;
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 36 awards totaling $560,000, ranging from $5K to $25K.

The review process actively engages the Users Committee and community members.
Training Workshops

- Each year, we hold training workshops on our software.
- This year, 75 people attended our workshops.
- Each year, we get several international participants.

<table>
<thead>
<tr>
<th>Event Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>netCDF for developers</td>
</tr>
<tr>
<td>GEMPAK</td>
</tr>
<tr>
<td>LDM</td>
</tr>
<tr>
<td>THREDDS Data Server Administration</td>
</tr>
<tr>
<td>Introductory IDV</td>
</tr>
<tr>
<td>Advanced IDV</td>
</tr>
<tr>
<td>netCDF JAVA</td>
</tr>
<tr>
<td>McIDAS</td>
</tr>
</tbody>
</table>
Sean Arms, Graduate Student
- School of Meteorology
- University of Oklahoma

Sean's two-year appointment as the student representative to the Users Committee is a historic moment in Unidata's nearly 25-year history.

Sean's selection came after a competitive process that saw the nomination of six highly-qualified students.

Sean uses several Unidata software packages in his daily life as a grad student that include GEMPAK, IDV, and netCDF.
Goal: Seamless Access to and Sharing of Data
- From a device anywhere on the Internet to local and institutional libraries anywhere on the Internet
Tremendous strides have been made in the commercial arena by companies like Amazon, EBay, Yahoo & Google.
Google Earth and Google Maps have unleashed a revolution in light-weight GIS integration. They have immense potential for enhancing geoscience education and spatial awareness, but they lack capabilities for conducting detailed analysis.
New Data Distribution Software

- Project still in the design stage
- Features will include:
  - Support for multiple operating-systems via Java
  - Support for adaptive and redundant connection topology
  - Download initiation via a web browser
  - Support for both static and dynamic datasets (e.g., case-studies, streaming data)
  - Infinitely-extensible product-namespace
  - Built-in security and cryptographic validation of data requests
  - Highly Scalable
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.
- Global scientific partnerships and data sharing activities are crucial for this transformation.
Thank You!

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