

## Unidata 5-year Proposal

**Some ideas that have emerged from internal staff discussions toward the development of the next 5-year proposal. These are provided mainly to stimulate discussion in the break out groups. They are not meant to be comprehensive, nor do they at this stage map directly to the strategic plan goals.**

Identify areas where Unidata can be an agent of transformation; For example,

- Advancing meteorology education and research,
- Facilitating climate science and services
- Supporting real-time weather prediction
- Enabling research at the interfaces of geosciences (hydrology, oceanography, air quality)
- Providing data and tools to support field projects
- Providing technologies for tool developers and data providers
- Playing a stewardship role in data format standards and conventions
- Broadening participation in cyberinfrastructure
- Democratizing access to data
- Promoting partnerships across the three sectors: government, academia and industry (e.g., NEXRAD Level II data distribution), use of Unidata-developed technologies by federal agencies and other institutions

### 1. Data Discovery

There are many aspects of data that need to be considered. Some areas for consideration are:

What is the appropriate role for Unidata in data discovery? How do users search for data? Topic is too researchy - field in infancy, but Unidata can still play a facilitating role. Unidata should facilitate development of metadata. Several ideas were discussed.

- DataPedia has been recommended in the past, but there has not been progress. The infrastructure should be in place with some prototype examples so it will catch on. Perhaps identifying specific people within the community could be tapped to provide input.
  - Descriptions of data streams need to be provided and should include metadata content, for search/discovery
  - Unidata could build the infrastructure to facilitate metadata submission
  - Link the collected metadata to datasets
- A thematic case study search mechanism should be considered
- Ontologies have not yet fulfilled their promise. Still a research topic and the concept may not be ready for operational use.
- Questions re. granularity, what level to access data, etc
- Data server aggregation-no one else is doing this-Unidata can be proud of this

- Need to document and describe data sets with key words
- Data should be available via APIs (application programming interfaces), not just by filling out a form on a web page
- Need to build up sets of data and train others to do the same thing – so far, in THREDDS, 25 NCEP model data sets under control—what’s the difference between the 25 models—do we have metadata in there?
- Who’s looking at the data, what are they using it for, and why? Use cases are needed.
- Need to begin with a strawman, ask community for additional ideas and information, create the framework and request various individuals for assistance with case studies, open to total community.
- Need to remember that our community might be using the data, but may not be using Unidata’s visualization/analysis tools.

## 2. Unidata’s role in Web 2.0-world of mash-ups

What are mash-ups? A **mashup** is a website or application that combines content from more than one source into an integrated experience.

[http://en.wikipedia.org/wiki/Mashup\\_%28web\\_application\\_hybrid%29](http://en.wikipedia.org/wiki/Mashup_%28web_application_hybrid%29)

- See Unidata e-Letter article “Web 2.0, Mashups, and the Unidata Community” by Rudolf B. Husar and Erin M. Robinson, Washington University in St. Louis
- University of Utah’s Mesowest data displays are another example: <http://www.met.utah.edu/cgi-bin/droman/mesomap.cgi>
- must provide an API to use data
- publish API and make available to others
- adding things such as Mesowest to THREDDS capabilities is a Unidata role, use standard interfaces, compatibility desired to interface with Google Maps/Earth
- Benefit: Democratize access to data and tools
- Unidata needs to listen to both sides of the community—users as well as developers.

## 3. Turning Data into Knowledge

Unidata facilitates the use of data and provides examples of how the tool works with the data, but the community needs to provide creativity for use in the classroom and research.

Data needs to be self-describing and geo-referenced so users can place where the data belongs. Automatic data enhancement tools-metadata-can feed into discovery center, e.g. GCMD – value added data and stored data – this is viewed as metadata. Need to focus on usefulness to researcher or educator.

Question of handling 30 users in a classroom-should Unidata be in the middle of such an activity with a prototype? Can Unidata contribute in the area of data mining and knowledge discovery?

Automatic data enhancement capabilities need to be considered.

We should provide a framework for standards and provide data services based on data that application can accept-provide data, metadata in several forms-Dublin Core, FGDC—we should not develop – can be done through partners.

Unidata technologies should be used in the archive of model outputs for the next IPCC, which are collected and served from the PCMDI at Lawrence Livermore National Labs for the [IPCC](#). (This topic also comes up later when discussing connections to the climate community)

#### **4. Role of Databases**

Reference: Unidata 2008 Panel Review Finding and Recommendation:

*Finding: It might be useful for Unidata to explore some cross- fertilization with the database research community. Databases have long standing as supporting application independence, persistent data storage, fast access, standard query methodology and data integrity constraints. NetCDF as the standard format is trying to serve this role but may in the long term be a limiting solution. Databases have yet to provide comprehensive support for georeferenced image data types but it is improving.*

*Recommendation: The panel recommends that Unidata seek interaction with the database research community to explore possible database solutions that could help address Unidata's data delivery objectives as well as promoting database solutions for the broader scientific community,*

Unidata has made some progress in this area.

“[One size does not fit all](#)” paper by MIT professor Michael Stonebraker about data bases.

- Databases make more sense for point data; not suitable for gridded data
- Interoperability needs to be emphasized with others
- Unidata needs to have some database for metadata, need indexing system-LDM and scouring side.
- THREDDS group is exploring the use of object-oriented relational databases for point data

#### **5. Data and Web Services:**

- Data services-Web services-coordination, authorization of access, how do we do that?

- Modular web services are important
- Need to consider data assimilation as a Web service issue
- Data services-low hanging fruit-select data in netCDF, etc.
- Decoders as web services
- Technological side and organizational side-data issues, strategic partnerships need to be considered
- Partners are important aspect of providing Web Services and WRF
- Unidata needs a good approach to working with super computing center groups
- Security issues can haunt web services-Unidata has no direct role but will leverage available technologies
- Every data stream provided by Unidata (excluding proprietary data sets) should be accessible via TDS and viewable using IDV
- WRF should be connected to THREDDS at both ends; input and output
- THREDDS is a key element - Create a prototype in Unidata and provide it to SCD and other super computing centers
- Migrating to clustered environment-serving data with THREDDS and moving toward clustered development to serve data using more than one machine
- IDV runs and TDS talks to it and provides a JPEG in return
- Use case-person is not always sure, but if TDS exists they can preview the data set in advance and then they can get the data.
- Continue to extend the Common Data Model – e.g., to support HDF4 and HDF-EOS (Unidata has received many requests for this)

## 6. Climate

- This is an important area for Unidata in the future
- A recommendation in the Metrics and Assessment report calls for increased emphasis in Climate - to study Climate via a weather prism
- Also mentioned by Cliff Jacobs in one of his comments on the Strategic Plan: “In recognition that the more members of the community are becoming interested in climate, as well, I think plan might look for some middle ground in terms of its scientific focus. One approach is to recognize the merging of weather and climate as a science focus and Unidata position itself to serve both communities --- serving as a special role as a facilitate to facilitate the merger.”
- Climate – need to begin researching to serve community –ask climate people what is missing in Unidata tools.
- Is 3D visualization useful for climate
- IPCC runs for processing
- Need a calendar library
- Server end for climate data - our tools are being adopted.
- Check with Community Data Portal, CCSM people, and NOAA/ESRL - now is the time to work with next set of IPCC runs to integrate Unidata technologies
  - Netcdf, netcdf operators on server or web services-Charlie Zender.
- Need to be sure IPCC chooses standards and technologies for storing and accessing the output from all the climate models that will run for the 5th

assessment report (due in 2012 or 2013) they choose netCDF, OPeNDAP, and our SOA rather than some other set of technologies

- Climate event notifications
- Need additional functionalities on both the server and client end

## **7. Next Generation Case Studies**

Vision: 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. In addition, they 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.

Pilot effort currently funded by UOP funds. Unidata is working closely with COMET and University of North Carolina, Charlotte to develop a prototype.

- integrating data with multi-media materials from COMET - should continue to develop framework for creation of the next generation of case studies-discovery setting-enabling creating of documents that people would be building

- Publishing capabilities could be developed.

WikiPedia-extended for data and case studies-case studies take you into a set of tools to be used for needs-connecting literature to data provided

**Goal** (Russ asked to add after Retreat): Update of the current proposal, deleting tasks already completed and updating proposed tasks with what we know now that we didn't know in 2003. This might give us an idea of what developments we need to continue working on and whether we already have some bold ideas that are extensions of our current work. It would also give us an idea if the 6 endeavors are still a good way to organize the proposed work.

## **8. Next Generation LDM, a key goal in the next proposal**

A replacement system for the LDM needs to be developed. It will combine the best features of the LDM, Inter Network News (i.e., netnews), and more modern peer-to-peer systems such as BitTorrent. The work is currently in the design stage within the UPC. Design features include

Support for multiple operating-systems via Java

Reduced bandwidth usage by allowing every site to receive only desired data

Support for dynamic and redundant connection topology

Download initiation via web browsers

Support for both static and dynamic datasets (e.g., case-studies, streaming data)

Infinitely-extensible product-namespace

Built-in (but optional) security via secure-socket-layer (SSL) connections and cryptographic validation of data requests

Graphical user interface

Scalable to the world

The proposed system will need to be integrated with notification services as well as TDS

Notification services should be incorporated into our data services

- THREDDS data server to notify people when things are deleted or changed
- OGC push notification
- Next version of LDM in Java - requires more staff input in design

LDM-notification- e.g., THREDDS needs to know when a new GRIB file or a METAR observation arrived

Need both Push-pull approaches

- can the next generation of LDM integrate push-pull for THREDDS?
- document how LDM fits with THREDDS for users in one stop shopping

## **9. THREDDS Data Repository:**

Archive to store and move data, notes, IDV bundles, etc. that are organized into case studies and other collections, and made accessible to the outside world.

- Programmatic interface
  - invoked by portal and web services
- Can have very large data volumes, inputs and outputs to ensemble forecasts (even greater than Unidata workshop level of stress)
- Interactive web form interface
- Archive designer may interactively browse, create, edit, and delete metadata and catalogs to create desired logical structure
- Safe and secure world access

- Move and copy operations to allow users to more easily structure and maintain the repository.
- Enhanced performance, including support for multiple simultaneous clients and very large data volumes
- TDR will
  1. Find and manage storage space
    1. Implement space management policies
    2. May involve various storage devices
    3. Could include replication
  2. Move the data into the repository, return a unique ID
    1. http, gridftp, scp
  3. Incorporate existing metadata
  4. Generate additional metadata where possible
  5. Integrate with a THREDDS Data Server

Long-term TDR goals:

- Secure long term data storage
  - Data movement
  - Supporting a variety of storage media
    - Mass storage devices
- Metadata and cataloguing support
  - Using THREDDS metadata and cataloguing system
- Both interactive and programmatic interfaces
- Integration with the Unidata THREDDS Data Server
  - Provides all applicable TDS access capabilities
- Invoke as a Web service

## 10. Visualization/analysis tools and architecture

- AWIPS Service oriented architecture—GEMPAK, AWIPS and merge into SOA by 2012
- Continue tracking AWIPS and NOAA visualization and analysis development
- IDV and visualization services-server that accepts bundles
- WMS-takes the data and returns a JPEG
- An important policy issue pertains to the continued development and support of all three (GEMPAK, McIDAS, and IDV) analysis and display tools.
  1. How long should UPC provide support for all three packages? A white paper developed by UPC and presented to Polcomm in April 2006 recommended that we continue support for all three packages as long as there are users. Does it still hold?
  2. NWS/NCEP moving toward AWIPS II; Long term development and support of NAWIPS (a.k.a. GEMPAK) with NWS is uncertain; AWIPS II, which will use open source software, Java-based, and fit into a service-

oriented architecture, is supposed to provide functionality currently available in NAWIPS/GEMPAK

- Should Unidata provide support for AWIPS II if and when it is available to the community. [AWIPS was proprietary, so Unidata decided to not bring it to the academic community, but there was community interest in that system]
- GIS integration will continue to be important
- Provide exchange mechanisms with commonly used tools (e.g., Google Earth, GIS, Excel, IDL, MatLab, etc.)
- Tools will need to facilitate Web 2.0 and mash-ups – provide frameworks and components to build useful tools
  
- a document that describes needs – use case needed –GIS-Google Earth and Google Maps and GIS integration generally needs to be considered
  
- Export and import GIS into tools – hydrology and oceanography it is useful for those communities

### **Community Equipment Awards**

- Continue to solicit proposals and make awards to new institutions from other GEO programs
- May 2007 Users Committee Action Item: Unidata should explore seeking additional equipment awards funding from other NSF divisions, due to the broadening efforts and collaborations

### **11. Miscellaneous**

- COSMIC data and Climate data need to be researched and made available to the community.
- Users Committee Action Item: Mohan Ramamurthy and Linda Miller will develop a short survey to explore the Unidata community's interest in receiving COSMIC GPS/Met data in real-time using the LDM.
- 

### **Community Broadening**

- Continue to broaden the Unidata community organically.
- Engage community and leverage partnerships to accomplish community broadening
- Continue efforts to enhance international participation

**Question:** Should Unidata make a concerted effort bring Unidata systems to community colleges? This is both a technology as well as a resource issue. (Unix expertise not readily available in community colleges, so they are unable to install the current LDM. Need a platform independent LDM that can run on Windows, Macs and Unix.

## **Sensor Web and Dynamic Adaptivity of Observations**

Should Unidata play a role in the sensor web area?

Environmental monitoring community is taking the lead in sensor web research and development. NEON, OOS and CUAHSI are key communities.

Dynamic adaptation to weather and dynamically data driven applications are becoming increasingly important. They are event driven.

## **Promoting Diversity**

Democratize access to data and tools that have been available to large institutions, and we make data and tools available to all. Encourage participation from underrepresented communities.

- Broad impact is successful, e.g. assessment.
- Democratizing-moving away from Unix-only systems.
  - PC version
- For example, Howard University is participating in LEAD.
- How can Unidata be more proactive in Historically Black Colleges and Universities and Minority Institutions, UCAR/SOARS is working with native American tribes-they would like real-time data
  - UCAR is working on this and we participate in SOARS
  - Gender diversity on staff; lacking diversity in under represented groups
  - We have been working to make tools/data available to minority schools, but we have not proactively tried to make it available. We need an advocate at the school to facilitate
    - Could consider a Visitor appointment, targeted workshops, targeted equipment grants, tap into a different pot of money at NSF?