A Transformative Community Facility for the Atmospheric and Related Sciences

Five-year plan

2009-2013

Table of Contents

Project Summary

Intellectual Merit

Unidata's mission is to provide the data services, tools, and cyberinfrastructure leadership that advance Earth system science, enhance educational opportunities, and broaden participation. Over 170 institutions worldwide participate in the Unidata data sharing network and many more institutions use Unidata tools and technologies in education, research, and operations. While its primary mission of serving the academic community remains unchanged, the user base has broadened considerably over the years, and its activities and responsibilities have grown as community needs have evolved. During the next five years, Unidata proposes to extend and enhance its well-established role as a transformational community facility. This proposal presents a bold but realistic plan that identifies areas where Unidata can continue to enable and enhance education and research that lead to advances at the frontiers of atmospheric and related sciences.

The proposed plan and endeavors are guided by the results of a comprehensive strategic planning effort over the past year that included strong input from and leadership by Unidata's governing committees and feedback from the broader community. The resulting strategy builds on the successes of the present program, its capabilities and core competencies, and its unique niche in providing robust, reliable, and comprehensive data services and tools to geoscience users.

This proposal presents a plan that is centered around the following six thematic focus areas:

- Broadening participation and expanding community services
- Advancing data services
- Developing and deploying useful tools
- Enhancing user support services
- Providing leadership in cyberinfrastructure
- Promoting diversity by expanding opportunities

In addition to Unidata's traditional role in enabling the meteorology community, the proposal identifies broad priority areas such as facilitating climate and field-project science as well as enhancing diversity in the atmospheric and related sciences. Unidata's guiding vision is well-integrated, end-to-send solutions for its users. By the end of this proposal period, a typical participating institution will be able to use all the technologies proposed in this plan to effectively integrate data in education, research, and operations. To achieve the stated goals and realize the vision, Unidata has put in place a highly proactive community engagement process and strong decision-making, governance, and feedback mechanisms, along with adaptable strategies, and a nimble organizational structure.

It should be emphasized that even as Unidata embarks on this plan to enhance and adapt its tools and services to meet the needs of an evolving community, the program remains deeply committed to meeting its responsibilities to the core atmospheric science community. The quality of services that the core community has come to expect will be maintained and enhanced as a result of the new partnerships and synergies.

Broader Impacts

That Unidata is a service organization is indisputable. Virtually no activity it undertakes is without a broader impact on its community, which encompasses users in many disciplines. The proposed plan

extends and expands Unidata's service mission so that it can continue to have a deep as well as broad impact in advancing NSF goals across all three sectors of the geoscience enterprise: academia, government, and the private sector. In addition to universities, many Unidata technologies are being used by national and international organizations, community projects, and the private sector. For example, data services in NOAA, NWS, NCAR, and other organizations and projects are now built upon the formats and tools that have been, and are being, developed at Unidata. Numerous other metrics are provided in the Project Description, along with examples that are highlighted, as Unidata considers its broader impact.

Project Description

I. Introduction

An integrated approach that transcends discipline and geographic boundaries is needed to understand and address societally important problems such as weather prediction, climate change, and the water cycle. Similarly, an Earth system science approach that employs inquiry-based learning is recommended for teaching geoscience. The success of the geosciences enterprise depends heavily on the availability of a state-of-the-art, robust, and flexible cyberinfrastructure, transparent access to high-quality data from diverse sources, and requisite tools and services to analyze, synthesize, visualize, interpret, and use the data effectively. The university community conceived and established Unidata nearly a quarter of a century ago to meet those needs, specifically to acquire and distribute real-time weather data to U.S. universities.

Unidata: a community-driven facility

Unidata's mission is to provide the data services, tools, and cyberinfrastructure leadership that advance Earth system science, enhance educational opportunities, and broaden participation. The Unidata Program Center (UPC) (http://www.unidata.ucar.edu/) is managed by the University Corporation for Atmospheric Research and sponsored primarily by the National Science Foundation through a proposal process and a longstanding cooperative agreement. Over 170 institutions worldwide participate in the Unidata data sharing network and many more institutions use Unidata tools and technologies in education, research, and operations. While its primary mission of serving the academic community remains unchanged through the years, the user base has broadened and its activities and responsibilities grown as community needs have evolved. The metrics in Table 1 provide a snapshot of Unidata's current influence.

As the enabler of a broad community, the UPC

- Acquires and distributes data to facilitate Earth system education and research
- Develops software for accessing, managing, analyzing, visualizing, and effectively using those data
- Provides comprehensive support to users
- Conducts annual training workshops on Unidata software packages
- Facilitates advancement of standards, conventions, and interoperability
- Provides leadership in geosciences cyberinfrastructure and fosters technological change
- Assesses and responds to community needs
- Advocates on behalf of the university community on data issues and negotiates data agreements
- Fosters community interaction and engagement to promote sharing of data, tools, and ideas
- Grants equipment awards to universities to enable and enhance participation in Unidata

Unidata's hallmark has been democratizing access to data and tools by serving both large and small institutions in higher education. The program benefits from the diversity of its user community, spanning the technological, educational, and scientific spectra. Unidata-provided cyberinfrastructure has enriched university courses by facilitating educators' efforts to incorporate applications of real-time data and state-of-the-art tools into student-centered learning experiences, enhanced productivity of students and researchers, and transformed the culture in atmospheric science departments. Unidata has experienced a gradual but natural evolution from a program focused primarily on synoptic scale

meteorology to one that serves a broader geosciences community. Unidata has attracted a broader community because it has been successful in providing tools and services that are interoperable, extensible, platform independent, and free. The robustness and quality of Unidata tools and services have resulted in their use beyond a community of several hundred universities, by organizations such as the NWS and other weather agencies around the world, NOAA laboratories, NASA, NRL, and ECMWF, as well as many companies in the private sector. In the process, Unidata has matured into a cornerstone facility upon which the university geosciences community and other stakeholders have come to rely.

As a facility for the university community, Unidata focuses on being responsive to current and anticipated user needs and aims to be efficient and nimble in doing so. To achieve that, Unidata has emphasized a community-driven, consensus-building process that provides a shared vision and the ability to make mid-course corrections. Critical decisions affecting Unidata are made only with formal guidance from its two governing committees: the Policy and Users Committees. Committee members represent the broad range of institutions served by Unidata and diverse geoscience disciplines. Unidata maintains a disciplined approach to the prioritization of its objectives as well as adoption of methodologies. An informal credo the UPC follows is to undertake no function that can be performed more effectively by the universities themselves. For these reasons, Unidata has come to be regarded as a forward-looking resource, and viewed as a trusted and knowledgeable partner by its stakeholders.

Research enabled by Unidata

Data and software provided by Unidata have contributed to numerous scientific studies and discoveries, as evidenced by the citation of Unidata and its offerings in 139 peer-reviewed and 277 conference papers in the last 5 years. The following list represents a cross-section of research enabled by Unidata:

- Climatology of mid-latitude cyclones and anticyclones over the United States
- Studies of tropical cyclone and hurricane life cycle and intensity estimation
- Variability of North American monsoon rainfall over complex terrain
- Studies of mesoscale convective vortices, tornadoes, and squall-lines
- Studies of fronts, cold-season precipitation events, and cold-air damming
- Validation of outgoing longwave radiation estimations with the GOES sounder
- Assessment of the ECMWF model cloudiness and surface radiation fields

Based on a recent survey of a dozen doctoral programs in atmospheric sciences, it is estimated, *conservatively*, that more than half of the graduate students used Unidata systems in their research. Unidata's impact on research extends also beyond the atmospheric sciences. For instance, Unidata systems and technologies are integral parts of community projects such as SuomiNet in the GPS applications arena, GEON and EarthScope in solid earth studies, and the multidisciplinary International Polar Year activities.

By providing data, advancing its software suite, organizing opportunities to facilitate exchange of ideas, and making cyberinfrastructure contributions, Unidata has become an essential part of the atmospheric sciences community. An independent study concluded that Unidata's sustained activities have had broad and transformational impact on education and research in the atmospheric sciences. The Committee of Visitors that reviewed the NSF/ATM's UCAR and Lower Atmosphere Facilities Oversight Section (Unidata's sponsor) reported in its 2003 findings that Unidata is a "vital national resource" and added that "if no Unidata system existed, it would have to be invented."

Unidata's value as a foundational facility for the geosciences has been amply demonstrated by a sustained record of innovation, a rich portfolio of tools and services, an actively-engaged community,

strong governance, mutually-beneficial collaborations, and the ubiquitous use of its services. With its solid foundation and core competencies, Unidata is poised to make even greater contributions to the geosciences by 2013. This proposal offers a bold vision and a realistic plan to extend the successes of a national facility that will empower the community and enable, to an even greater extent, transformative advances in the geosciences. To that end, a 5-year plan for achieving the vision is provided through six overarching themes that are shaped by pertinent scientific and education drivers, consistent with community trends and needs, and aligned with the NSF priorities. Implicit in this plan is recognition that Unidata will continue to provide and extend its data services and related tools to enable the community to address important scientific and societal challenges and advance discovery and learning at the frontiers of atmospheric and related sciences.

Table 1
Metrics 2003-2008

Data	Numbers
Institutions Participating in the IDD	~170
Host Machines on the IDD	460 (250 unique domains)
Data Streams in the IDD	22
Approximate volume of data ingested in the IDD	100 GB/day
Volume of data pushed to the community	2.7 TB/day
Volume of data available from UPC for remote access UPC LEAD Test Bed motherlode.ucar.edu	26.5 TB 24 TB 2.5 TB
Average volume of data pulled via remote access protocols per day	44.3 GB
Up Time of UPC Data Infrastructure	99.96%
Number of Unidata/COMET case study datasets	44
Software	**Downloads (2005 – 2007)**
Analysis & Visualization	
GEMPAK	5,029
IDV	10,933
McIDAS	874
Data Distribution	
LDM	5,760
LDM-McIDAS	948
Data Access	
THREDDS Data Server	2,214
Data Management, Infrastructure, and Conversion Software/Middleware	
netCDF	111,550
netCDF-JAVA	20,752
netCDF-Perl	2,873
UDUNITS	3,280
Workshops	**Number of Participants**
Training Workshops (5)	335
Users Workshops (2)	164
Regional Workshops (2)	44
Total Workshop Attendees	543
Total Number of Training Courses	38
Other Metrics	**Numbers**
Number of Users Registered with Unidata	19,492
Equipment Awards Made in the past 5 years	36 ($600K)
Number of community Email Lists managed	34
User Support Email Transactions (Jan 2006 to Jan 2008)	10,400
Staff at the UPC	25

II. Results of prior support for NSF 0317610 (1 October 2003 – 30 September 2008)

The UPC's activities are funded primarily under the proposal entitled "Unidata 2008: Shaping the Future of Data Use in the Geosciences." During the period of performance for the current award, the UPC met most of the goals stated in that proposal and accomplished many of the objectives. The myriad activities of the program were organized under six broadly-themed endeavors:

- Responding to a broader and more diverse community
- Comprehensive support services
- Real-time, self-managing data flows
- Software to analyze and visualize geoscience data
- Distributed, organized collections of digital material
- Improved data access infrastructure

Unidata significantly increased the reliability and volume of data flowing to the community, enhanced the capabilities of its software, developed and deployed new tools and services, actively engaged the community, and served an increasingly diverse group of users in the geosciences. Collectively, the accomplishments and activities have enhanced the transformational impact of Unidata on the community. The following list highlights Unidata's most significant accomplishments during the past five years, with Table 2 providing a snapshot and summary of each endeavor.

Unidata's key accomplishments and successes during 2003-2008

1. Significantly increased both data volumes and the number of participants in the data distribution system. The volume of data flowing has increased nearly five fold, while the number of participants in the IDD has more than doubled. As a result of improvements to the LDM, the reliability and timeliness of data delivery has seen a marked improvement.
2. Made available new types of data, previously unavailable to a broad community, in real-time, including Level II radar data, high-resolution NCEP model output, and COSMIC global soundings.
3. Achieved advances in the LDM that have resulted in its adoption as an important operational infrastructure within the NWS, NCDC, and a dozen other organizations worldwide. The improvements have resulted in its use by THORPEX to support TIGGE, which requires aggregate data transfer rates in excess of 250 GB per day between 10 operational weather prediction centers across the world and the three archive centers at NCAR, ECMWF, and CMA.
4. Developed and released the IDV, a platform-independent analysis and visualization application and framework that supports access to and integration of diverse remote and local data sets, collaboration among users, and new views of data. Built entirely in Java, the IDV opened up personal computers (Windows and Macintosh OS X) as platforms for Unidata analysis and visualization applications. The IDV framework is now in use beyond the Unidata community by projects such as GEON, McIDAS-V, and organizations such as EUMETSAT. IDV availability has resulted in the use of Unidata tools in non-traditional communities, including community and four-year colleges and universities that may not have UNIX expertise.
5. Made advances in both push (LDM) and pull (OPeNDAP, ADDE and THREDDS) data services.
6. Developed Java middleware such as THREDDS and netCDF that are now being used by a growing number of data providers in academia and government, including NOAA, NASA, and DoE. NetCDF advances resulted in its adoption as a standard for data access in the WCRP CMIP3 multi-model data archive and facilitated its broad use in the IPCC Fourth Assessment.

7. Facilitated data exchange with GIS client applications and communities through enhancements to netCDF, TDS, and the IDV. One result was ESRI's adoption of netCDF as a data format in ArcGIS, enabling GIS data integration and adding a million potential new users of netCDF data.
8. Provided funds for equipment purchases to three dozen universities to encourage new members from diverse disciplinary backgrounds in the geosciences to join the Unidata community and to allow existing members to continue and enhance their active participation.
9. In partnership with other stakeholders, proposed and helped to organize Earth and Space Science Informatics sections at the AGU and EGU to facilitate communications and coordinate pertinent activities. Convened and organized several stakeholders and users meetings on Unidata-related projects such as CRAFT, CONDUIT, THREDDS, OPeNDAP and GALEON.
10. Conducted Triennial Users Workshops in 2003 and 2006, as well as annual Training Workshops on Unidata applications. The theme of the 2003 Users workshop was "Expanding Horizons: Using Environmental Data for Education, Research, and Decision Making," while the theme of the 2006 Workshop was "Expanding the Use of Models in the Atmospheric and Related Sciences." Nearly 500 participants attended these Unidata-organized meetings. Helped community members organize Unidata Regional workshops at Millersville University in 2004 (20 participants), and University of Oklahoma in 2007 (24 participants).
11. Provided comprehensive support to users via diverse avenues. On average, Unidata now provides responses for over 5000 support inquiries each year.
12. Developed capabilities to support the use of Unidata tools in field projects (RICO and T-REX).
13. Provided leadership on data standards, conventions, and related cyberinfrastructure topics through advocacy, collaboration, and ongoing efforts within the program

Table 2:
Highlights of Accomplishments under the Previous Five-Year Award

Proposed Objectives	**Work Accomplished**
Endeavor 1. Responding to a broader and more diverse community	
Community governance	Facilitated and coordinated over two dozen meetings of the Policy and Users Committees and Community Equipment Award panels
Representation on governing committees	Diversified committee representation by adding hydrology, oceanography, air quality, GIS, GPS, and field project scientists to Policy Committee membership and hydrology and climate scientists as well as a student member to the Users Committee.
Increased participation in the IDD	The number of participants in the IDD continues to grow steadily. At this writing, 460 host machines at approximately 250 unique network domains are receiving real-time data via the IDD.
Community Equipment Awards	Made 36 equipment awards to universities for a total of approximately

	$600K.
Broadening via provision of platform-independent software	Developed and deployed the IDV, a platform-independent tool for data access, analysis and visualization, permitting the use of Unidata tools in institutions that do not have Unix expertise.
Bringing community together	In partnership with other stakeholders, advocated for instituting Earth and Space Science Informatics sections at the AGU and EGU to bring the community together to address important cyberinfrastructure issues.
Community engagement	Convened stakeholders and users meetings and workshops on the following projects: CONDUIT, CRAFT, MeteoForum, GALEON, OPeNDAP, and THREDDS.
Interdisciplinary users	Continued to engage users in other geosciences including the hydrology, oceanography, air quality, and GIS communities. Established collaborations with CUAHSI, IOOS, OGC, UNAVCO, GEON, NSDL, DLESE, and the ESIP Federation.
Nonacademic users and collaborators	Collaborated with several government organizations: NOAA Labs, NESDIS, NCDC, NWS, NODC, NASA, CMA, OGC, ESRI, and ITT Vis.
International users and collaborators	Extended use of Unidata applications and data by establishing strategic international partnerships. Engaged international community through collaborative projects. Universities in 10 countries are now participating in the IDD. NetCDF is used in over 70 countries.
Enhancing diversity	Fostered diversity by providing science, community, and writing mentors for SOARS students.
Endeavor 2. Comprehensive support services	
Support emails	Answered roughly 140 support inquiries per week on about 50 topics ranging from analysis and display applications to community outreach
Simplifying software installation	Simplified Unidata software installation by using Webstart installers for IDV. Eased binary installations for other Unidata applications.
Web-based training modules	Created training modules for the IDV using screencast technology.
Data-flow description	Created an interactive, multi-layered web document to assist users in locating, understanding, and using all data available through Unidata.
Training workshops	Conducted training workshops each year on Unidata data software. Total attendance over 5 years: 335

Triennial Users Workshops and Regional Workshops	Conducted Triennial Users workshops in 2003 and 2006, which were attended by a total of 164 users through Users Committee collaboration.
Web Site as Information Commons	Redesigned the Unidata web site to provide a consistent look-and-feel for all packages and activities, and easier software downloads. Expanded capabilities and on-line forums for community interaction.
A New tracking system	In 2005 a new, commercial inquiry tracking system, SupportSuite. Since its implementation over 10,400 user-support interactions have taken place via the system.
Electronic newsletter	Launched monthly online newsletter, CommunitE-letters, in April 2004 whose primary goal is informing the community, while providing a forum for showcasing member accomplishments and activities.
Seminars	Instituted periodic seminars on topics of interest to a broad community. Provided real-time and archived access to seminars via a web browser.
Collaboration tools	Researched and deployed the Access Grid technology at the UPC to facilitate remote collaborations.
Advocacy	Advocated on behalf of community on data-related matters.
Endeavor 3. Real-time self-managing data flows	
LDM advances	Made significant advances in the LDM that resulted in improved performance, throughput, and scalability. In addition to its deployment in over 250 academic and research institutions, the LDM has now become part of the data infrastructure in numerous national and international agencies and large projects.
Data stream advances	Provided over two dozen data streams to the community via the IDD and other point-to-point topologies. In addition to significant increases in data volumes in existing streams, new data streams have been added: NEXRAD Level II radar data, HYDRO-NEXRAD feed that includes value-added multi-parameter metadata of each radar scan, and GPS-Met data from COSMIC.
LDM/IDD data volumes and throughput	Increased to nearly 100 Gigabytes amount of data ingested into the IDD each day, a five-fold increase over the last 5 years. The LDM systems at the UPC now transmit approximately 2.7 Terabytes each day to community members. As a result, the LDM moves more data via the Internet 2 backbone than any other advanced application, surpassing data volume via FTP since 2004.
Decoders	Offered and maintained a suite of decoders that interface to and accept data

	from LDM feeds and make them usable.
NNTP-based data distribution	Developed and demonstrated an end-to-end NNTP-based approach to data distribution, including relay, reception, decoding, and visualization. A platform-independent receive-only client was developed for Windows. This work has served to inform the development of the next generation LDM.
Endeavor 4. Software to analyze and visualize geoscience data	
GEMPAK	Grew GEMPAK use in over 300 academic institutions worldwide. Added support for WRF model and ensemble output, WSR-88D Level II data, QuickSCAT, and COSMIC.
McIDAS	Provided a new release each year. Used by 45 institutions. Significant interest and use continues in the international community. Improved functionality and ease of use that included updates to accommodate changes in NOAAport datastreams, ports to new Linux distributions and new development environments, and simplified the GUI. The server-side capability, ADDE, has enabled a network of Cooperating Community Servers that provides substantial volumes of data to McIDAS and IDV users worldwide.
IDV	Released initial version of IDV in 2003. Innovations in the IDV framework and reference application include platform independence, 3D visualization, remote and local data access via multiple protocols, plug-in framework for customization and use in different communities, support for GIS integration, and collaboration tools. Now used by 200 institutions and projects well beyond the academic community with its adoption by the solid earth community (GEON-IDV), SSEC for next generation McIDAS (McIDAS-V), EUMETSAT, IRIS, EOL to support field projects, the LEAD project, and CMA's Shanghai Typhoon Institute.
Endeavor 5. Distributed, organized collections of digital material	
THREDDS	Originally funded by NSF/EHR (NSF Grant #DUE-0121623, NSF #0333600) as an NSDL project, THREDDS has become an integral part of Unidata's core activities. Using THREDDS, datasets are conveniently accessible from a collection of THREDDS-enabled analysis and display tools. The breadth of data served by and integrated into THREDDS technologies continues to grow, actively engaging new disciplines and incorporating new tools that make data more useful at all educational levels. THREDDS technologies are deployed not only by the UPC but also by an increasing number of data providers like NOAA, NCAR, and NASA.

Dataset inventory catalogs	The THREDDS catalogs allow data providers to list available online datasets. The catalogs can also be searched via discovery mechanisms.
THREDDS Data Server (TDS)	The TDS provides remote access to many types of real-time and archived data, and it has been deployed, among other locations, at the UPC (motherlode.ucar.edu), NCDC, and NCAR.
GIS interoperability	THREDDS has also been focused on facilitating data exchange with organizations in the GIS communities including hydrology. In effect, the TDS makes Unidata collections available to an entirely new set of client application programs that observe common protocols such as OPeNDAP and WCS. Among these are ESRI's ArcGIS, IDL, MATLAB, and many others.
Next Generation Case Studies	This project has developed a prototype and framework for delivering a new generation of case studies that are dynamic and incorporate learning modules from COMET and other sources, as well as a full-suite of atmospheric data delivered via the TDS and ADDE and visualized by the IDV. The case studies can evolve and be added to through community contributions.
Endeavor 6. Improved data access infrastructure	
NetCDF advances	Major advances were made to the netCDF Java and C-based interfaces, including the development of a new netCDF-4 interface. While providing backward compatibility, netCDF-4 allows users to read and write HDF5 files with the netCDF API that includes support for groups, user-defined types, per-variable compression, multiple unlimited dimensions, and parallel I/O. The merger of netCDF and HDF5 was originally funded by NASA, but is now advanced through core funding. The enhanced netCDF-Java interface allows users to access remote data via OPeNDAP and HTTP and provides support for subsetting and dataset aggregation. Created and released a test version of a new software package, libcf, for supporting interoperability and compliance with the widely-used CF conventions.
NetCDF use	Downloaded by users in over 70 countries making it Unidata's most widely used software. Incorporated into more than 50 open source and over a dozen commercial software packages. As a result, netCDF has become a key infrastructure element for data providers and users of oceanographic and atmospheric science data. Several research projects have adopted netCDF as a standard for data access and archives, including the WCRP CMIP3 multi-model dataset used for the IPCC Fourth Assessment.
Common Data Model	Developed and implemented a Common Data Model in the netCDF Java library that generalizes the data models of OPeNDAP, netCDF, and HDF5. The CDM provides users transparent access to data stored in multiple file formats through a single interface. The CDM establishes a set of abstract data types

	which have been mapped to the ISO abstract model for coverages.
NetCDF Markup Language	NcML is a natural augmentation of netCDF with extensions encapsulating descriptions of the structure and content of netCDF objects in an XML form. NcML allows for describing "virtual netCDF" files that may be aggregations of data from several existing netCDF files, or it can represent a target dataset to be created by transforming existing netCDF files into a new form described in the NcML language. The NcML-G extension provides a means for fusing the data models of the traditional netCDF atmospheric science community with those of the GIS community.

III. Key drivers of the Unidata Program

The key drivers of the Unidata program are presented here in light of the evolving landscape in which Unidata operates. Highlighted are three high-level drivers that inform this proposal – science, education, and information technology.

A. Science

Science is a fundamental driver in defining Unidata's mission, and enabling science through the provision of data and tools is its primary responsibility. As described in the introduction, Unidata has enabled research in many atmospheric science areas, but most notably in weather prediction and mesoscale meteorology. Unidata's contributions to advancing science in those areas will undoubtedly continue and be enhanced, but here we highlight a few additional areas of importance that will shape the program's future and focus its work.

For more than a decade, there has been a growing trend in Earth and environmental sciences toward understanding the Earth as a synergistic system involving complex, multi-scale, interrelated phenomena and processes in the atmosphere, lithosphere, cryosphere, hydrosphere, and biosphere. Studying Earth system phenomena such as El Nino, climate change, and the water cycle requires approaches that transcend disciplinary as well as geographic boundaries.

Strong consensus now exists that human activities are altering the Earth's climate system at an accelerated pace (IPCC 2007, AGU 2007, and AMS 2008). Recognition of climate change has brought to the fore a need for increasing our understanding of both natural climate variability and anthropogenic changes, its societal and ecological impacts, as well as linking those findings to policy decisions. The scientific community is recognizing that coupled climate system models, which have long been the primary tools for diagnosing and projecting future climate changes, lack important features that are crucial for understanding how climate change will affect the world's ecosystems and biogeochemical cycles, and their feedback on the climate system. A new generation of Earth System models, enabled by corresponding software frameworks, is being developed, and the models are expected to make significant improvements in climate simulations and predictions.

Meanwhile, the weather prediction community has seen a major paradigm shift over the past fifteen years (Tilmann Gneiting and Adrian E. Raftery, 2005). Until recently, meteorologists viewed numerical weather prediction (NWP) as an intrinsically deterministic endeavor: that is, for a given initial condition,

a single "best" forecast is generated. However, as high-performance computing capabilities advanced, the community began using ensemble forecasting techniques to deal with the myriad uncertainties of NWP systems. Ensemble forecasting is based on statistical representation of collections of dynamical forecasts, and it has become the primary method of forecasting at operational centers. An ensemble forecast comprises multiple (typically between 5 and 100) runs of NWP models, which differ in the initial conditions and/or the numerical representation of the atmosphere, thereby addressing the two major sources of forecast uncertainty. As a result of the success of ensemble approaches, they have now become an integral part of prediction on many time scales, including short and medium-range weather prediction (Kalnay, 2002) for example.

Advances in remote and in-situ sensor technologies and modeling systems are not only driving a revolution in science, they are also generating vast quantities of data. To make progress in those areas will require collecting, assimilating, and synthesizing observations across disciplinary boundaries. Comprehensive cyberinfrastructure, including information architectures for the discovery, retrieval, integration, analysis, visualization, and knowledge extraction, is needed. As data volumes grow, it is imperative that scalable, flexible, and efficient solutions are developed. The global and interdisciplinary nature of geoscientific problems also require that science be conducted collaboratively by distributed teams of investigators, often involving the use of community models, common frameworks, and shared data sets, as exemplified by IPCC assessment activities.

B. Education

The Earth system approach views our planet as a single integrated system. Such a holistic view provides a powerful perspective for understanding the interconnected processes that drive the Earth and the many relationships between it and its inhabitants.

Earth science education is also uniquely suited to drawing connections between the dynamic Earth system and societal issues. Catastrophic events like the 2004 Indian Ocean tsunami and Hurricane Katrina provide ample evidence of this relevance. Other types of severe weather events (e.g., tornadoes, floods, and blizzards) and their impacts also demonstrate the connection between science and society. These events underscore the importance of timely and interdisciplinary integration and synthesis of data.

Earth science education is best advanced by incorporating new teaching techniques, active learning strategies, and real-world data into curricula. Students need opportunities for authentic, hands-on inquiry so that they experience the scientific process and excitement of discovery. A critical component of scientific inquiry includes learning to collect, analyze, synthesize, and interpret data to understand the underlying relationships, complexities, and uncertainties of the processes that control and shape the planet. The university community has responded in many ways to challenges surrounding global warming and is taking a leadership role in addressing climate change issues. As pointed out by Rappaport and Creighton (2007), “for faculty in a very wide range of disciplines, climate change presents rich opportunities for teaching, research, and community action.”

The richness of students’ exploration and experience depends on the quality of the data available and the tools they use. Cyberinfrastructure provided by Unidata allows students to access the same databases and tools used by scientists, providing an important pathway to integrate research and education.

Unidata recognizes that sustained efforts in enabling technologies are critical to realizing the vision of a well-developed workforce. This proposal promotes such perspectives by democratizing access to data

and facilitating their integration via platform-independent tools and services, and common data formats and conventions.

C. Information Technology

Throughout its history, information technology (IT) advances have shaped Unidata's activities and the ways in which the program has served its community. The following list identifies some of those IT developments that have informed Unidata's work:

- The Internet and the World Wide Web (WWW)
- Personal computers and commodity microprocessors
- Client-server architectures and remote access protocols
- Object-oriented programming
- Open standards and open source approaches
- Web services, Web 2.0, and Sensor Web
- Geographic Information Systems
- Digital libraries
- Virtual organizations
- Social networks
- Grid/Cloud Computing

The Internet has radically transformed the conduct of science and education. The WWW has enriched and enhanced the provision of services by shifting away from centralized approaches toward a more scalable, user-centric model. It has democratized Internet use while providing interactive and customized services and changing the paradigm for searching, sharing, and delivering data, products, and educational materials. Platform-independent Web services, usually implemented in service-oriented architectures, have emerged as tools for providing a new generation of data services. Such services can be standalone, performing simple tasks, or they can be linked to create more complex services.

The advent of web services, digital libraries, and open standards and protocols has been important in shaping a new generation of data services. The second generation web services referred to as Web 2.0 harnesses the network using it as a platform for deploying new kinds of applications. Through creative integration of tools and services, Web 2.0 has made it possible to provide richer, customized services, such as interactive mapping and tagging, and overlaying of scientific and societal data.

The UPC has tracked, evaluated, implemented, and incorporated many of these technologies for community benefit. This plan is also informed by and consistent with the NSF's cyberinfrastructure vision (NSF, 2007). Collectively, the above drivers present both exciting opportunities as well as interesting challenges. It is envisioned that a future shaped by these drivers will lead to transformational changes in the operation of and services provided by the UPC and will make important contributions to advance the geosciences.

IV. The plan for the next five years

As stated in the introduction, Unidata was founded by the meteorological community so that universities could acquire and use real-time weather data in education and research. Since then, both the Unidata community and the scope of its activities have grown, and participation in the program has broadened both geographically and to other geoscience disciplines. This organic growth has occurred with modest incremental resources due largely to the leveraging of capabilities in the program's traditional areas.

During the next five years, Unidata proposes to extend and enhance its role as a community facility and to leverage the following core competencies:

- Low-latency distribution of real-time data
- Integrated geoscience data analysis and visualization
- Data access infrastructure, including services for managing and serving archived data
- Community engagement, bringing stakeholders together to address common needs
- Comprehensive software support
- Advocacy on behalf of the community on data and related matters
- Flexible and platform-independent software solutions

The plan identifies areas where Unidata can facilitate educational and research activities that lead to advances at the frontiers of atmospheric and related sciences.

Even though Unidata has a tradition of *proactively* enabling research and education related to weather, the climate science community has only *indirectly* benefited from many of the capabilities developed by the UPC. As universities respond to challenges presented by climate change, Unidata must accommodate needs in that area. As documented in the Results of Prior Support section, the climate community has been using Unidata software (e.g., netCDF and TDS), although climate science has not thus far been a primary focus for Unidata. In this proposal, facilitating research and education in climate is singled out as a new priority area for Unidata.

As noted earlier, a second notable scientific trend in the atmospheric science community over the past decade has been the steady move toward probabilistic forecasting, using ensemble weather prediction techniques. This proposal recognizes that change and presents an approach for addressing the attendant challenges and community needs.

The proposed plan and endeavors therein are guided by the results of a comprehensive strategic planning effort over the past year that included strong input from and leadership by Unidata's governing committees and feedback from the broader community. That effort helped to clarify Unidata's goals and connect them to Unidata's overarching mission to enable research and education in the geosciences. The resulting strategy builds on the successes of the present program, its capabilities and core competencies, and its unique niche in providing robust, reliable, and comprehensive data services and tools to geoscience users. It should be emphasized that even as Unidata embarks on this plan to enhance and adapt its tools and services to meet the needs of an evolving community, the program remains deeply committed to meeting its responsibilities to the core atmospheric science community. The quality of services that the core community has come to expect will be maintained and enhanced as a result of the new partnerships and synergies.

In this section, we present a plan that is centered around the following six thematic focus areas:

- Broadening participation and expanding community services
- Advancing data services
- Developing and deploying useful tools
- Enhancing user support services
- Providing leadership in cyberinfrastructure
- Promoting diversity by expanding opportunities

The proposed endeavors and advances are so important that the greatest risk would be not to pursue them, leaving Unidata universities with static tools and services in a dynamic environment of rapid technological and scientific advances.

A. Broadening participation and expanding community services

Since Unidata's inception, there has been a gradual evolution in the Unidata academic community as many traditional meteorology departments in universities broadened their mission, scope, and role in the scientific and educational enterprise. This gradual metamorphosis has been shaped by the changing landscape as environmental, scientific, educational, and societal priorities have evolved due to a growing shift toward integrative and collaborative science. Even as atmospheric science programs in universities evolved, Unidata has attracted users in other geoscience and engineering disciplines to its tools and services. Today, the Unidata community includes researchers, educators, and students in many geoscience and engineering departments, unconstrained by international boundaries. This organic broadening of the community is expected to continue, and the proposed plan positions Unidata to respond and adapt to the evolving needs of a growing community.

International activities

Increasingly, the conduct of science requires strong international scientific partnerships and the sharing of information, knowledge, and other assets. This is particularly true in the geosciences where the highly coupled nature of the Earth system and the need to understand global environmental processes and their regional linkages have heightened the importance of international efforts. The climate system, for example, is far too complex a puzzle to be unraveled by individual nations. As science becomes increasingly global in nature, it is critical that focus is placed on full, open, and timely access to and sharing of ESS data and related analysis tools.

The Unidata Program recognizes the benefits of a global cyberinfrastructure and the power of networked communities, as institutions and people exchange knowledge and resources. Unidata's international activities began modestly as the MeteoForum project in 2001, funded with internal UCAR funds. Since then, Unidata as a community has developed a growing portfolio of international outreach activities, conducted in close collaboration with academic, research, and operational institutions on several continents. The portfolio includes providing data, tools, support, and training as well as activities that bring various stakeholders together to address important issues, all toward advancing the goals of building a globally-engaged community of educators and researchers. Real-time atmospheric science data delivered to Latin America has helped initiate teaching innovations in universities in Argentina, Brazil, Chile, and Costa Rica. As a result of these efforts, data are also flowing from other continents back to the U.S. for educational, research and operational use.

Priorities for the next five years:

- Continue to foster a shared vision for and community ownership of the program, creating a framework for capturing community input
- Provide a transformative learning environment for the community by involving users as both contributors to and stakeholders in Unidata
- Broaden participation in Unidata from closely-related geoscience disciplines such as hydrology and oceanography
- Adapt services to the needs of the climate community
- Leverage partnerships with universities to entrain nearby community colleges into the Unidata community
- Continue to broaden participation in Unidata governing committees

- Bring the community and stakeholders together to share knowledge and address problems that are important to them through meetings, workshops, conferences, and other venues
- Develop global partnerships with geoscience data holders for free and open sharing of data
- Advocate community interests in areas pertaining to data needs, policies regarding data access and use, and related technological matters.
- Bridge communities across a range of computing resources and capabilities; e.g., connect Unidata systems and services with NSF HPC resources like the TeraGrid
- Manage and expand the community equipment awards program to serve the evolving needs of the community

The program will monitor community growth through its management and governance structure and avoid overextending itself as it responds to broader needs that may include different data types, larger data volumes, and new tools and services. To serve the broader community without reducing the level of service to the current users, the UPC will continue its approach of leveraging partnerships and collaborations with other organizations and working with contacts in the new communities to set up support systems for their respective users.

B. Advancing data services

Earth system research and education are inherently multidisciplinary in nature. They require comprehensive data services that provide integrated access to diverse sources and formats of data for advancing our scientific understanding and creating new knowledge. Providing easy access to data via reliable, flexible, and efficient data services is a cornerstone of Unidata's mission and remains an overarching goal of this proposal.

Rapid advances in computing, communication, sensing, and information technologies have revolutionized the provision and use of data, tools, and services and continue to shape the development of data services. At the same time, the unrelenting increase in geoscience data volumes and their increasing diversity as well as complexity pose important challenges for data services. This proposal presents a plan to facilitate the end-to-end data needs of the community—from data processing through, decoding, management, distribution, discovery, access, integration, analysis, visualization, and interpretation, to publication and sharing. Combined with the shift toward ensemble models, the impending data deluge from new kinds of ground-based and satellite instruments (e.g., Phased Array Radars, and NPOESS and GOES-R sounders) will result in staggering increases in data volumes, requiring new strategies for providing scalable data services and efficient access.

In response to those challenges and to meet evolving community needs, Unidata will continue to apply sound software engineering with innovative technologies and software approaches—web services enabled via service oriented architectures and scientific workflows, for example—to foster the creation of useful products and services. To support current users and attract new users in other communities, Unidata must continue to adapt its services to evolving computing environments and user requirements.

To facilitate discovery, enhance the value, and maximize the effective use of geoscience data in research and education, this plan also recognizes that those data must be enriched with appropriate metadata and linked with information services related to those data.

Modern applications like the IDV need well-defined and well-documented APIs to access data and metadata. Interoperable use of such APIs requires agreement on data conventions. Unidata is well positioned to make valuable contributions to agreement on suitable data conventions and to facilitate the governance needed to sustain them. The netCDF-CF conventions have proved to be a successful

approach to specifying metadata for gridded model outputs, but more work is needed for conventions for observational data.

Priorities for the next five years

- Develop and provide high-level interfaces to geoscience data as well as simple mechanisms for locating, accessing, and distributing real-time and thematic data, creating and publishing metadata, and providing data availability notifications
- Adopt, develop, and promote open standards, conventions and protocols for data formats, access, and metadata to enhance interoperability of data services
- Facilitate seamless integration and analysis of data from diverse sources, including GIS information
- Enable users and partner institutions with data holdings to contribute and share their data easily
- Empower community members to create and deploy innovative data services using web services and mash-up technologies
- Provide extensible frameworks for creating next generation, dynamic case study datasets

The overarching data services emphasis is that Unidata will provide a well-integrated, end-to-end solution to its users. By the end of this proposal period, a typical participating institution will be able to use all the technologies proposed here to effectively integrate data in education, research and operations. Concrete objectives for advancement in specific areas are outlined in the following subsections.

NetCDF

- Remote access: improve compatibility between netCDF APIs and OPeNDAP protocols in ways that will enhance the ability to represent the intent of data providers and for the ability to efficiently access subsets of remote data.
- Performance: Enhance netCDF with support for stream access, modifications to make libraries thread-safe, robust benchmarks to preserve and improve performance, and guidance about use of compression, chunking, and data structures for efficient access.
- Support for CF conventions: Develop C-based LibCF as a reference implementation supporting the continued evolution of CF conventions, including support for structured grids, new observational data conventions, standard name attributes for quantities, climate modeling calendars, and interoperability with OGC and ISO data models.
- Support for standard names: Provide interfaces for CF standard that offer query functionality beyond what is available from OGC or ISO interfaces.
- Improved language interfaces: Upgrade Fortran and C++ interfaces to support netCDF-4 and to exploit language advances such as portable inter-language calls, namespaces, exceptions, templates, and iterators.
- NetCDF Utilities: Enhance utility programs with human-readable time coordinates, performance-related attribute representations, and NcML compliance.
- Documentation and usability: Provide generic examples (such as copying) for each language interface and documentation for Common Data Model interfaces.

THREDDS and Common Data Model

- Provide full data access to the HDF family of file formats (HDF4/HDF-EOS, HDF5/HDF5-EOS/NPOESS) through netCDF-Java/CDM software .
- Provide efficient access to output from ensemble forecast models and output on unstructured grids.
- Complete the implementation of "scientific data types," including Point/Station/Trajectory, Radial, Swath, and Unstructured Grids in the CDM. Standardize conventions for representing these types.
- Make Unidata/LDM data feeds, as well as case study data, available through the THREDDS Data Server (TDS) and accessible by the IDV.
- Make the TDS efficiently scale to large collections of datasets and files.
- Create web services to subset large collections of data, returning data in convenient formats.
- Support OGC and ISO standards (WCS, WFS and WMS) to allow access to meteorological datasets from users in other disciplines. In particular, target development that benefits users in the GIS, data assimilation, and numerical modeling communities.
- Map the resulting CF conventions to international standard specifications such as the ISO "coverage."
- Make progress, through appropriate collaborations, towards "Scientific Databases" to allow complex search capabilities over very large collections of scientific datasets.
- Enhance the CDM and TDS to provide support for aggregating, accessing as a collection, fusing, and analyzing data from ensemble model output.
- Facilitate the installation of the TDS in universities so that faculty and students can share their local data holdings
- Enable users in other communities to access weather and climate data using their own client applications instrumented with standards-based interface protocols.

THREDDS Data Repository

The TDR is a tool for collaborative development of repository content, complementing the TDS for serving collections and RAMADDA (see below) for metadata management. The TDR is envisioned to be used by the community to archive, manage, and move data, annotations, IDV bundles, case studies, and other collections. Along with an interactive user interface, the TDR provides a programmatic interface (API) so that it can also be invoked by portals and web services.

Using the TDR, clients can populate a repository by uploading data to a server and creating catalogs in an automated fashion, a functionality that addresses a gap in Unidata's current software suite. The TDR also provides ways to edit repository content, including restructuring the repository and editing metadata. Furthermore, the TDR facilitates collaboration by supporting content owners role-based authorization to control write and edit access to their content. It supports integration of distributed resources via the notion of a remote resource, which resides elsewhere but is cataloged and linked to a repository.

Databases and RAMADDA

The area of scientific databases remains a nascent one with different view points on their suitability for storing and accessing large volumes of scientific data, especially gridded datasets. Column-oriented databases are an alternative that provide access two orders of magnitude faster than traditional relational row-oriented implementations used in enterprise contexts. NetCDF-4 and HDF5 data models

are capable of representing relational data as well as column-oriented and gridded data, and query languages for accessing such data are emerging. Unidata proposes continued research and development work in this important area involving technologies more suitable for scientific data access than either low-level file systems or high-level enterprise database systems.

In that regard, a new UPC effort called RAMADDA, Repository for Archiving, Managing and Accessing Diverse Data, an open data management framework that enables a data provider or a community of users to upload, manage, and share large data holdings, has been designed to run in multiple contexts, ranging from local desktop use to real time and case-study data archives. Its HTML interface provides support for configuration, browsing, search, creation, update and management of the data repository. A Java library and web-service API support integration with external data management systems, data provision services (e.g., TDS), and end-user clients (e.g., IDV).

RAMADDA incorporates a number of rich information management facilities, including an extensible metadata framework, grouping, tagging, annotations, and associations. Based on an extensible Model-View-Control (MVC) framework, RAMADDA offers the ability to plug-in a variety of ways to access and deliver information—e.g., THREDDS catalogs, RSS feeds, and real-time event notification.

While the TDR was developed to address functionality missing from the TDS, RAMADDA was developed to explore a database approach for managing and querying data collections. The UPC currently is exploring how best to move TDR, TDS, and RAMADDA technologies forward and to realize the benefits of their complementary functionality. Efforts include putting the TDR service interface in front of RAMADDA, and also creating an interface and accompanying functionality that would allow a TDS to retrieve needed server side information from RAMADDA that is currently stored in THREDDS catalogs.

Next Generation LDM

The LDM has grown to be a robust, reliable and portable base on which to build data distribution networks. Despite the widespread use and success of the current LDM, its limitations (small number of feed types, static topology, high-bandwidth use by top-level relay sites, and lack of Windows support) inhibit use and adoption by other communities.

The need for a more scalable technology is apparent from several trends:

- The volume of useful data is growing faster than increases in network bandwidth
- Approaches to setting priorities for bandwidth by "shaping" its use, charging departments for bandwidth, or other institutional data policies may disrupt the model of essentially free bandwidth currently in use for the IDD
- Competition for network bandwidth from the explosive growth in video distribution may demand more dynamic routing than is now possible with current LDM technologies

To avoid interruption in data delivery, future needs must be anticipated and reliable software must be developed before the need is apparent. The current LDM has advanced about as far as it can given the constraints of its architecture and the underlying technologies it uses.

Minimizing disruption—both to individual sites and to the flow of data -- is a major concern of any new implementation and deployment. The LDM software is open source, so no one will be forced to upgrade to different and incompatible software. In fact, there will be two competing distribution systems running in parallel for some time, during which the new system will have to prove itself capable of

satisfying users of the current LDM as well as providing solutions to problems the current LDM cannot handle.

Next generation case study data services

The UPC is currently developing a framework for a new generation of case study datasets that are dynamic, integrated, and interactive. These case studies will include datasets of important meteorological events, but will also permit integration of relevant educational modules. The vision is "living" or dynamic case studies, allowing the community to augment and enhance them by contributing related observations, analyses, and curricular and multimedia materials. An important element of that vision is providing a framework in which community members can build on existing case studies, contribute new case studies to collections, and connect conference papers, journal articles to their underlying data, and other literature.

Unidata, in collaboration with COMET and Prof. Brian Etherton at the University of North Carolina-Charlotte, is developing a prototype new generation case study that incorporates learning modules from COMET and other external sources, and atmospheric data delivered via THREDDS catalogs and visualized in the Unidata IDV. Three interactive, educational modules related to the 2005 Atlantic tropical cyclone season are being built using these tools (Holmberg et al., 2007). These case studies will eventually be open for contributions via the TDR/RAMADDA (under development).

Unidata proposes to continue and expand this prototype effort. To realize the vision for the Next Generation Case Study project, the UPC will create a mechanism for the community to share and collaborate in creating and enhancing case studies.

C. Developing and deploying useful tools

Unidata's reputation is most strongly linked to software tools that are used to access, analyze, visualize, integrate, interpret, and explore geoscience data. Unidata currently develops and maintains three tools for analysis and visualization: GEMPAK, McIDAS, and IDV. These tools are being employed widely in education and research as well as operational meteorology. The IDV is also being used in hydrology, oceanography, and other geosciences.

As part of its software development philosophy, Unidata emphasizes object-oriented, platform-independent, and open-source approaches to tool development, and use of open standards. Some tools (e.g., IDV) are developed at the Unidata Program Center, while others (McIDAS and GEMPAK) originate elsewhere, but are modified, maintained, and supported by the UPC. Each application has its strengths and limitations and a community of users, but all three packages have become mainstays in university classrooms and labs. Collectively, the unique capabilities in these tools are helping to enable transformative research in the atmospheric and related sciences.

From the outset, there was a realization that no single application package met all the requirements of the Unidata community, from quick looks at current weather and visualizations in classrooms to sophisticated analyses for advanced research and high-quality plots for publications. Typically, Unidata sites use a combination of GEMPAK, McIDAS, and IDV to meet different needs.

An important consideration for Unidata is the future of GEMPAK and McIDAS in the organizations where they originate—NCEP and the University of Wisconsin/SSEC, respectively. Both organizations are planning to migrate their respective applications. Based on interactions with and feedback from its community members, the UPC has learned that once users become proficient at accomplishing a set of tasks with a specific application, it is difficult for them to transition to different application. Therefore, it

is important to develop a smooth transition path for users, and these plans are described in the following sections.

IDV

The IDV, now used by over 200 institutions, has been very successful in integrating data from various sources, in providing useful visualizations of georeferenced data from other geoscience disciplines, in demonstrating the usefulness of innovative display technologies, and in providing an end-to-end application for testing and improving Unidata's data collections and data access infrastructure software. Because of its many unique capabilities (e.g., 3D views and remote data access via multiple data access protocols), the community of IDV users has grown significantly in the last five years. The IDV also permits seamless integration of data with educational materials, as exemplified by LEAD-to-LEARN https://portal.leadproject.org/gridsphere/gridsphere?cid=modules) and Next Generation Case Study modules (Weber 2008)

The UPC will continue to focus on developing and enhancing the IDV and its software framework to provide multidisciplinary, collaborative analysis and visualization capabilities to facilitate integrative science and interactive, inquiry-based learning. That vision calls for Unidata to advance the IDV to support scalable analysis and visualization of data on a range of spatial and temporal scales – from the mesoscale to global scale and from convective time scale to climate scales.

Specifically, we propose to enhance the IDV in the following ways:

- Ability to access and visualize every datastream provided by Unidata.
- Support for analysis and visualization of ensemble model output (e.g., probabilistic threat and multivariate probabilities, mean-spread analysis, spaghetti diagrams, and aggregated remote access to large ensemble archives).
- Improved handling of sub-minute data (lightning and model) as well as long range climate data.
- Support for field project operations and datasets.
- Adaptation to GIS frameworks and integration of GIS data.
- Enhancements to the IDV framework and plug-in capabilities for a range of customized applications (e.g., lightweight IDV for use in educational settings).

GEMPAK

GEMPAK is an externally developed application that presently leverages the work of several developers at NCEP. Known as N-AWIPS within NCEP, it has the largest user-base (over 300 institutions) of any Unidata analysis and visualization tool, due to its many unique capabilities that include up-to-date decoders, publication-quality output, and comprehensive model output diagnostics. GEMPAK also has limitations: it does not run on Windows, and it cannot directly access data from remote servers. GEMPAK was designed for use in weather analysis and forecasting, so it is difficult to apply to other kinds of geoscience applications.

The NWS is embarking on a project to develop the next generation Advanced Weather Interactive Processing System, AWIPS-II, to be completed in 2011 (Tuell et al., 2008). An important component of AWIPS-II is integration of N-AWIPS functionality to AWIPS-II. Because of the development of the migrated N-AWIPS the current N-AWIPS package will be frozen in summer 2008 (Schotz et al., 2008).

The impending moratorium on further N-AWIPS development is expected to impact GEMPAK users. In response, the UPC is working with N-AWIPS and AWIPS-II developers on strategies for a smooth transition plan for GEMPAK users. In the near term, the UPC will continue support of GEMPAK's use in

universities. In the long term, the UPC will announce a transition away from GEMPAK, while providing support for the final NCEP release as long as feasible. The UPC will stop adding new capabilities, maintain GEMPAK in the face of data stream changes for a period of time, and set up an open source repository to permit continuing community support. GEMPAK users will be encouraged to transition to the IDV by enhanced user training and the addition of GEMPAK-like capabilities in the IDV. Simultaneously, the UPC will work with the NWS and NCEP to investigate the possibility of bringing AWIPS-II to the university community and providing support for it. Given its expected capabilities, the academic community is likely to be interested in using AWIPS II. The UPC will seek guidance from our governing committees and input from our users in making decisions on AWIPS-II support.

McIDAS

Unidata McIDAS is used in about 45 institutions, a decrease from over 100 sites that used it in 2003. The majority of new interest in Unidata McIDAS comes from organizations outside of the U.S., although significant use continues in U.S. universities. Like GEMPAK, McIDAS will also undergo a major transition in the coming years. SSEC is working to transition McIDAS to the IDV/VisAD-based application called McIDAS-V. Unidata's McIDAS plans will remain aligned with SSEC's plans to transition the current McIDAS based software into McIDAS-V. The UPC proposes to continue support for McIDAS (along with McIDAS decoders and ADDE server software) as long as feasible, but gradually encourage current users to migrate to McIDAS-V or other Unidata applications.

Other client applications supported via standard interfaces

A key benefit of broader deployment of the TDS (described in "Advancing Data Services") will be the ability to access weather and climate data via standard protocols. As a result, ArcGIS, IDL, and MatLab users will be able to access data locally via the netCDF interface and from remote servers via the WCS and OPeNDAP interfaces on TDS servers. Continued adoption and adaptation of these protocols for use in the TDS will make Unidata datasets available to a much wider range of applications software including applications supported by other organizations.

D. Enhancing user support services

Timely and effective communication and provision of comprehensive support for users are integral to Unidata's mission. Strategic communication not only informs and educates the community about the program's activities, it also engages community members as active participants.

Unidata's communication methods include traditional print media, electronic newsletters, an interactive website, topical mailing lists for targeted audiences, Really Simple Syndication (RSS) feeds, online forums, and webcasts. Unidata staff routinely present program activities and results and receive feedback at various meetings and community workshops.

Users of data and software also need comprehensive support services, including easy access, enhancements, and bug fixes, consultation, documentation, and training. Unidata is widely known for its exemplary support to its university community and hands-on training to optimize the use of its tools and services. In a 2006 survey conducted by Nelson and Nelson, LLC, 95% of the respondents stated that user inquiries were answered in a timely manner.

As the Unidata community grows, maintaining excellent support is crucial. That goal will be achieved by enhancing Unidata's inquiry-tracking system, creating an information commons that includes a searchable community knowledge base, providing online discussion forums, and web-based training materials and self-education tutorials.

Priorities for the next five years

- Maintain and advance effective communication among community members to augment and advance program goals. Specifically, enhance user experience and web site usability while sustaining a manageable and secure infrastructure.
- Create and maintain an authoring environment that allows easy content contribution for our developers and staff as well as community members.
- Improve support to the community by harnessing advances in online collaboration technologies and use of online forums for support.
- Augment training for Unidata tools and services through innovative methods such as web-based training modules.
- Cultivate a mutually supportive environment to foster expertise and knowledge sharing; specifically, nurture self-organizing support communities in which users help each other.
- Explore and, if appropriate, make use of new modes of community interaction that leverage social networking sites such as MySpace, Facebook, and Second Life.
- Increase use of web-based conferencing systems to support remote collaboration.
- Simplify downloading, installation, and maintenance of all supported packages
- Improve documentation with enhanced descriptions of available datasets, datastreams, and services

E. Providing leadership in cyberinfrastructure

Internet technologies have amply demonstrated the compounding benefits of a global cyberinfrastructure and the power of networked communities as people and institutions share knowledge and resources. As articulated by NSF Director Dr. Arden Bement (2007), cyberinfrastructure is a potent tool for knowledge creation—one that must be shaped with foresight and strategy, while at the same time considering human, organizational, and social factors.

Unidata recognizes that providing robust, comprehensive, and persistent cyberinfrastructure is essential to transforming geosciences education and research, as well as the community's culture. By developing and providing state-of-the-art data services, tools and middleware, Unidata has become a leader in the establishment of geosciences cyberinfrastructure. That leadership role goes beyond the day-to-day operations and management of the program, and includes advocating both within and outside of the Unidata community for adoption and diffusion of new innovations and approaches to data access, sharing, and use.

The program takes an active role in building a virtual, distributed knowledge community and bringing stakeholders together to address important cyberinfrastructure issues by organizing national and international meetings. The increasing emphasis on standards-based solutions has led to Unidata's assuming a central role in identifying and articulating standards, conventions, and data formats and in facilitating efforts to reach consensus based on a broad perspective. Over the years, Unidata, through its commitment to developing and sharing new knowledge, has served as an intellectual commons, providing stimulation of ideas, and appropriate models for community interaction while developing a strong culture for technological change in academia and government.

Data providers in NOAA, NCAR, and other organizations rely increasingly on technologies and services built by Unidata. The Unidata LDM is the underpinning of a system that distributes terabytes of data in real-time, to universities, government organizations, and the private sector. It has become the foundation for the operational distribution of Level II radar data in the National Weather Service. The CDM, proposed by Unidata, has sharpened the discussion on the representation of different types of

observations so that they can be both stored and transported in a consistent fashion. Likewise, the Unidata-led GALEON Project, for example, is taking the lead in developing standards to facilitate integration of geosciences data sets with GIS databases.

Priorities for the next five years

- Take a leadership role in helping to set future directions for geoscience cyberinfrastructure
- Continue to advance geosciences cyberinfrastructure by developing and providing useful data services and tools
- Help guide the development and evolution of community and formal international data system standards to ensure the resulting specifications support the needs of ESS data systems.
- Provide stewardship for and facilitate development, adoption, and implementation of standards, data formats, conventions, and protocols
- Bring users and stakeholders together to address important cyberinfrastructure issues by organizing workshops, meetings, and other gatherings
- Advocate on behalf of the geosciences community on cyberinfrastructure issues
- Serve as a testbed for the practical deployment of emerging technologies

F. Promoting diversity by expanding opportunities

Attracting students from diverse backgrounds to the field is a fundamental challenge for geosciences. At this time, representation in geosciences education and research by minority populations does not reflect society (Czujko, 2002). This problem is only expected to worsen as demographic projections suggest that minority population is increasing more rapidly than the rest of the population. A large body of research has enumerated the importance of and the immediate need for enhancing diversity in Science, Technology, Engineering and Mathematics (STEM fields), including the geosciences (Snow et al., 2002, AGU, 2002; http://www.agu.org/sci_soc/education/diversity.html).

Cultural, social, and technological factors (Whitney et al., 2005 – EOS, Vol. 86, pg. 277) are among the suggested reasons for the lack of diversity in STEM fields. Additionally, inadequate access to infrastructure and resources has been cited as a formidable obstacle, (AMS 2007, http://www.ametsoc.org/amsedu/online/info/diversity.html).

Promoting diversity is an imperative for Unidata and the UPC is well-positioned to serve as a catalyst for change because of its long-standing role in providing easy access to data, simple-to-use tools, and enhancing communication. Diversity is woven into the fabric of Unidata's activities, and is intended not only to enhance the vitality of the program through new perspectives and applications, but to sustain and contribute to its long term value as well. Any effort undertaken by Unidata will require a careful strategy, close partnerships with universities and other stakeholders, coordinated efforts on multiple fronts, additional resources, and realistic goals. This proposal offers a pathway and a few modest steps to address the diversity challenge.

Providing broad access to technology and services with minimal entry barriers is a key strategy in enhancing diversity. To that end, the UPC proposes to leverage its close relationship with its community to expand opportunities for universities with underrepresented populations (e.g., MSIs and HBCUs) to participate in Unidata activities. Implicit in that strategy is the recognition that Unidata resources can play a key role in making science relevant to local environments and experiences; but for any effort to have long term success and be sustained, it will be important to train a cadre of experts in the new communities and institutions to help build capacity in those communities.

Priorities for the next five years

- Engage, proactively, a highly diverse population of educators and researchers as core constituents targeting efforts such as community-specific regional workshops, and increase diversity on Unidata's governing committees.
- Foster alliances between traditional Unidata universities and community colleges and minority-serving institutions to address under-representation. Leverage synergistic activities within UCAR and the university community and appoint short-term visitors at the UPC with the express goal of building ties and developing concrete activities to entrain users in new communities.
- Lower barriers to using Unidata tools and services by developing platform-independent data systems and analysis and visualization tools that run on Windows, Mac OS, and Unix/Linux. The IDV, GIS capabilities of Unidata data services, and the proposed next generation data transport software are all expected to facilitate the advancement of this goal.
- Develop and promote specific opportunities that broaden participation from underrepresented-communities, targeting minority serving departments in the Equipment Award process, for example.
- Continue to take advantage of initiatives that make communication systems and information technology available in parts of the world where it is now inaccessible.
- Continue active participation in the UCAR-led SOARS program that has been highly successful in recruiting and retaining minority students to the atmospheric sciences.

V. Broader Impacts of the proposed activity

That Unidata is a service organization is indisputable. Virtually no activity it undertakes is without a broader impact on its community, which encompasses users in many disciplines, including atmospheric sciences, oceanography, hydrology, and solid earth sciences. Unidata continues to attract new users because it provides data, tools, and services that are interoperable, extensible, collaborative, and platform independent.

The proposed plan extends and expands Unidata's service mission so that it can continue to have a deep as well as broad impact in advancing NSF goals across all three sectors of the geoscience enterprise: academia, government, and the private sector. Below, a few examples are highlighted as Unidata considers its broader impact.

A. Enhance infrastructure for research and education, advancing discovery and understanding

Unidata-provided cyberinfrastructure is used widely for research and education at over 300 universities world wide. Unidata has fostered diversity by including non-traditional disciplines, expanding its users in community colleges, and democratizing data and tool provision to international institutions. As concluded in an independent Metrics Assessment of the Unidata Program (Nelson and Nelson, 2007), "the Unidata program has established a reliable socio-technical environment that leverages developments in data delivery to maximize creativity and learning in the geosciences."

Over the past five years, Unidata software and data systems have contributed to the publication of over 500 scientific papers on a broad range of topics in the geosciences.

Unidata systems and technologies are integral parts of many projects including but not limited to SuomiNet, THORPEX, GEON, EarthScope, and IPY. That impact is expected to increase as Unidata systems are enhanced and become the fabric of geosciences cyberinfrastructure.

One of the most shining examples of Unidata's broader scientific impact is the use of the netCDF-CF data format in the IPCC Fourth Assessment activities. NetCDF advances resulted in its adoption as a standard for data access in the WCRP CMIP3 multi-model data archive at PCMDI. As pointed out by Meehl et al (2007) its adoption precipitated a new era in climate research. Researchers at all levels and in diverse geographic locations now can use multi-model output in their work, further enabling them to contribute to the global research community.

The adoption of the IDV framework by GEON is another example of a Unidata tool that has had a broad impact on geoscience research. GEON IDV, an extension of the Unidata IDV, is a tool for exploration and display of data in solid earth geophysics. It has demonstrated IDV's value in seismic tomography and mantle geodynamics by providing new insights in the exploration of geophysical data (Meertens et al., 2006).

The democratization of access to data and tools is resulting in attracting and educating tomorrow's scientists. At one institution, the transformation enabled by Unidata led to a tripling of the undergraduate majors in the meteorology program (Nelson and Nelson, 2007).

Unidata systems have facilitated experiential as well as hands-on learning with real-world and real-time data. The software Unidata provides has enabled students to use the same tools of the trade that scientists and operational practioners use. In the process, Unidata facilitates integration of research and education, a long-standing NSF goal.

The triennial workshops Unidata conducts are another broad impacts activity. The last such workshop in 2006 on "Expanding the use of models in the atmospheric and related sciences" focused on disseminating models and pedagogic approaches to their integration in teaching.

Unidata-developed cyberinfrastructure, in addition to being used widely in universities, is broadly adopted and used by other stakeholders in government and the private sector. Many data services in NOAA, NWS, NCAR, and other organizations and projects are now built upon the formats and tools that have are being developed at Unidata.

B. Broaden the participation of underrepresented groups

Unidata infrastructure is being used at universities in 23 out of 27 EPSCoR states, including many institutions that have a large number of students from underrepresented sectors of society. Unidata-provided products and services help contribute to and promote the development of research activities in areas of strategic importance to the NSF mission.

Through universities, Unidata assists in preparing a diverse workforce for the different sectors of the atmospheric sciences enterprise.

Unidata technologies have improved course offerings, providing venues for professional development workshops for high school science teachers, and ultimately to enriching a summer program called Minority Outreach Science Enrichment Program at the University of Missouri-KC.

C. Disseminate information and data to enhance scientific and technological understanding

The UPC uses many avenues for disseminating information to enhance scientific, technological, and educational impact of the program. These include a comprehensive website, e-letters, RSS feeds, online forums, webcasts of presentations, and targeted mailing lists.

In addition to direct dissemination by the UPC, Unidata universities and colleges carry out their own outreach activities facilitated by the Unidata infrastructure. College of DuPage, the Midwest's largest single campus community college, makes considerable use of Unidata's datastreams and tools for disseminating real-time weather products to outside users. An Automated Volcanic Ash Forecast System at the University of Alaska-Fairbanks, uses Unidata systems in its operations. At the University of Hawaii, Unidata systems are pivotal in supplying the Mauna Kea Observatory with custom forecasts to facilitate astronomers in their work.

Unidata infrastructure is being used at NASA Space Centers and in emergency preparedness and management during Katrina, the Challenger disaster, NTSB investigations, and the Columbia Scientific Balloon facility for mission launches.

At the Boston Museum of Science, the WeatherWise exhibit successfully uses data delivered by Unidata and visualizes it using a Unidata-supported tool.

D. Metrics and Assessment Study

In addition to the metrics provided in Table 1, this section describes the results of a formal metrics and assessment study that concluded in 2007. Prompted by a recommendation from the NSF panel that reviewed the last five-year proposal and framed by questions posed by NSF, the UPC initiated an independent evaluation of Unidata in 2005. NSF asked if Unidata has been a successful investment, and if so, why. Additionally, it asked if Unidata has been transformational in the way community members conduct teaching and research in the atmospheric and related sciences.

An independent firm experienced in program evaluations, Nelson Consulting, LLC, conducted the study. The evaluation included a community survey, focus group sessions, individual interviews of users, and a few exemplar case studies of Unidata's impact on the field. Participation in the study was far and wide: 432 participants took part in the survey that included 57 questions. Several dozen community and governing committee members participated in focus group discussions and individual interviews, providing ample opportunity for the consultants to evaluate Unidata's activities and their impact on the community.

The study concluded that the quality of software and support offered by the UPC is exceptional and that Unidata has stayed true to its core values. It also concluded that Unidata was a model program and an irreplaceable national facility, and that its establishment was among UCAR's most important accomplishments. The participants stated that without the data, software. and support provided by Unidata, atmospheric science programs would be severely limited in their ability to carry out their missions. The study was unequivocal in its overall conclusion: Unidata has indeed been a worthwhile investment for NSF, and it has noticeably transformed the way universities conduct education and research in the atmospheric sciences. Another key finding is that Unidata's activities are in alignment with NSF strategic goals. The study found that there are many reasons for Unidata's success, but foremost among them are: a dedicated and well-informed staff, a highly engaged community, and an

energetic governance structure. One community member described Unidata's significance as follows: "The Consortium of Universities for the Advancement of Hydrologic Science (CUAHSI) is an organization representing more than a hundred USA universities. In recent years, with NSF support, CUAHSI has been working on providing hydrologists with better access to hydrology data. Most hydrologists would agree that the ultimate goal is to have an equivalent of Unidata for hydrology."

The full report is available at
http://www.unidata.ucar.edu/2006assessment/Final_Report_021707_Nelson.pdf

VI. Strategies for achieving the goals

The rapidly changing technological and scientific environment in which Unidata and its community operates makes detailed long-term planning impractical. For that reason, Unidata has put in place a highly proactive community engagement process and strong decision-making, governance, and feedback mechanisms, along with adaptable strategies, and a nimble organizational structure. While the proposal describes Unidata's future directions, the program's acknowledged vigilance in monitoring the community's pulse provides the flexibility needed to make course corrections, overcome unforeseen problems, and seize unanticipated opportunities.

The priorities proposed here can only be realized with a clearly defined integrative strategy. The following list is not comprehensive, but it establishes broad strategies for future actions and reflects Unidata's overall philosophy and guiding principles.

- Engage and empower the community so that members are both participants and contributors to the overall effort
- Forge strategic alliances with key stakeholders and leverage partnerships
- Focus on the Unidata niche, sustaining and building on the strengths of the core program
- Provide effective service by constantly searching for ways to improve organizational efficiencies
- Emphasize platform-independent and object oriented approaches to technology development so software can be reused in multiple contexts
- Adopt and adapt existing technologies, developing new ones only as a last resort
- Where appropriate, adopt community and formal standards and influence their evolution

A. Synergistic activities that complement the core effort

To maintain a vibrant program, the UPC, from time to time, participates in certain research and development projects that are tied to its overall mission but funded separately from the core program. Such synergistic activities are both essential and complementary to the core effort, and both the Policy Committee and NSF encourage them. The UPC undertakes non-core projects only after careful analysis of their merit and benefits to the community and upon endorsement by the Policy Committee.

Such projects have played a vital role in advancing the program in new directions, creating new capabilities for the community, enhancing interoperability of Unidata software, providing new datasets to the community, and entraining and diffusing innovative ideas and technologies into the community. Almost always, synergistic projects leverage ongoing activities in the core program, are conducted in collaboration with other partners, and allow the UPC to offer its technologies and expertise for the greater benefit of the community (e.g., the CADIS IPY project). In addition to their intrinsic merit, the projects have also helped to relieve pressure on the ATM-funded portion of the overall Unidata budget. Over the past five years, such projects have been kept to a modest level of about 15% of the overall

effort. Below we highlight a few examples of synergistic activities that have greatly benefited the Unidata community.

The project to develop case study datasets is an early example of a synergistic effort that greatly benefited universities. That effort, originally funded by the NWS, was carried out in collaboration with COMET and JOSS, and the result was the availability of 44 case study datasets that are still in wide use. The THREDDS project began as a separately sponsored non-core effort, with funding from the NSF EHR's NSDL program. The project develops middleware and infrastructure to bridge the gap between data providers and data users. The THREDDS effort is regarded as an effective way to broaden the user community and expand data services. For those reasons, in 2003 the NSF review panel recommended that THREDDS be funded as a core Unidata activity when its funding from the NSDL program ended. Today, THREDDS is an integral part of the UPC portfolio and an important community resource.

Similarly, the effort to merge netCDF and HDF5, a capability long desired in the atmospheric science community, was possible only with funding from NASA. That effort also provided the motivation for developing the CDM, a key underpinning for a new version of the netCDF software and the TDS.

Linked Environments for Atmospheric Discovery (LEAD) is a collaborative, NSF-funded Information Technology Research (ITR) grant. LEAD involves nine institutions that bring different expertise (e.g., meteorology, computer science, grid computing, pedagogy, and data services) to develop a new paradigm for on-demand weather prediction. To that end, LEAD is creating an integrated, scalable framework in which meteorological analysis tools, forecast models, and data repositories can operate as dynamically adaptive, on-demand, grid-enabled systems. Unidata's role in LEAD is to a) extend Unidata technologies to work in Grid and Web services environments; b) help deploy LEAD in the Unidata community; and c) set up and manage one of the four test bed sites. In addition to providing many of the technical underpinnings like LDM, THREDDS, and IDV, Unidata is playing a pivotal role in LEAD's deployment in the meteorological community. The Unidata LEAD testbed has become a valuable community resource, providing a 180-day archive of many IDD-delivered data sets. While Unidata's involvement is crucial to LEAD'S success, it is not without rewards that feed back to Unidata and its community. Unidata has been able to leverage LEAD-sponsored work to develop the TDR technology. Meanwhile, the LEAD testbed continues to provide an excellent crucible for stress testing Unidata's technologies in areas such as the scalability of catalogs. The LEAD use cases have provided opportunities for enhancement of tools like the IDV and TDS. And most importantly, Unidata universities have begun to use the LEAD system for making on-demand weather predictions for a number of different applications, including the WxChallenge forecast contest.

In summary, the UPC recognizes the mutual benefits of synergistic activities that harness the strengths of innovative projects that are funded outside the core program through targeted opportunities. Unidata will continue to pursue such opportunities as they arise during the period of performance of this proposal, but as always, any new activity will be reviewed and carefully balanced against the core effort.

B. Partnerships and collaborations with other organizations and projects

Since its inception, Unidata has forged strategic partnerships within the university community and with other organizations. Mutually beneficial collaborations enable Unidata to fulfill its mission, while the resulting leveraging leads to increased efficiencies and contributes to new capabilities. Such partnerships have resulted in the entrainment of innovative tools and data assets of importance to Earth system science research and education, and advanced standards and interoperability of Unidata systems.

Two noteworthy examples are provided to illustrate the value of such collaborations: Software such as GEMPAK, McIDAS, and VisAD libraries, which underpin the IDV tool, originated externally, but the UPC enhances, maintains, distributes and supports them on behalf of the university community. Similarly, to meet the needs of the university community, the UPC has worked closely with NOAA, NWS, NCEP, and other organizations to gain access to important datasets, such as observations from NEXRAD and National Profiler networks, NLDN, ACARS platforms (meteorological data from commercial aircraft), MADIS data bases (including surface, upper-air and GPS measurements), and output from operational models from NOAA and other agencies (e.g., CONDUIT, NOGAPS, COAMPS, and GEM).

Such collaborations are also leading to broader adoption and use of Unidata technologies. Many data services in NOAA, NWS, NCAR, and other organizations and projects are now built upon the formats and tools that have been, and are being, developed at Unidata. These partnerships have contributed toward advancing the overall value of the projects and the vast communities they serve. In this proposal, we have put forth a plan to continue and enhance these collaborations.

VII. Priorities and schedule

Aside from identifying broad priority areas such as climate science and diversity along with concrete technology development activities (e.g., next-generation LDM and RAMADDA), the decision to provide neither a single ranking of priorities for the proposed activities nor a detailed timetable for completion of many of the tasks described here requires some justification. Since this is a proposal to continue and extend the activities of a long standing facility, traditional approaches to presenting priorities and schedules may not be most appropriate. As in the past, priorities will be set based on input from users, advice from governing committees, discussions with the NSF, and available resources. Experience shows that success in realizing the goals requires carefully balancing competing priorities dynamically to match community needs. Unidata's staff, management, governance structure, and community engagement have been built with the recognition that continuous rebalancing of priorities is the key to remaining successful.

The UPC has embraced agile software development methodologies that value individuals and interactions over processes, as well as responding to change over following a detailed plan. There is good reason to believe these principles will continue to serve the Unidata community well over the next five years.