Unidata 2018: Transforming Geoscience through Innovative Data Services

Data are the lifeblood of the geosciences. Rapid advances in computing, communications, and observational technologies — along with concomitant advances in high-resolution modeling, ensemble and coupled-systems predictions of the Earth system — are revolutionizing nearly every aspect of our field. The result is a dramatic proliferation of data from diverse sources; data that are consumed by an evolving and ever-broadening community of users and that are becoming the principal engine for driving scientific advances. Data-enabled research has emerged as a Fourth Paradigm of science, alongside experiments, theoretical studies, and computer simulations (Hey *et al.*, 2010).

For more than a quarter century, Unidata has worked in concert with the atmospheric science education and research community to develop and provide innovative data systems, tools, techniques, and resources to support data-enabled science to understand the Earth system. In doing so, Unidata has maintained a close, synergistic relationship with the universities, engaging them in collaborative efforts to exploit data and technologies, and removing roadblocks to data discovery, access, analysis, and effective use.

Unidata's long experience helping geoscientists incorporate data-centric techniques into their scientific workflows positions us perfectly to help our community take advantage of a rapidly changing scientific landscape. Science is shifting in emphasis from pure disciplinary research to a more balanced mix that advances disciplinary knowledge while looking to apply research results to interdisciplinary "grand challenge" problems touching both science and society. Unidata is in an ideal position to adapt to and capitalize on the rapid advances in information technology that will enable the community to incorporate the dramatic explosion of data volumes into their research and education programs. This will ensure that the next generation of students and young researchers in the Unidata community become leaders using state-of-the-art tools, technologies, and techniques.

Unidata's community expects it to both *lead* through innovation and *follow* the community's direction by being responsive to its changing needs. We welcome these challenges, viewing them as opportunities to enhance our community's ability to work together to create knowledge, while recognizing that our own ability to adapt and change as new technologies emerge is crucial to our continued success.

To that end, this proposal charts a course for the next five years, laying out our plans for sustaining and enhancing the Unidata program. While the goals and the action plan articulated here are consistent with and in furtherance of the National Science Foundation's vision and goals to transform the conduct of science, it provides a specific vision in which community members are empowered to tackle emerging data-related scientific and educational problems in more innovative, efficient, and productive ways. Achieving the goals in this proposal will bring our community closer to the vision of *geoscience at the speed of thought* articulated in our strategic plan, dramatically reversing today's situation in which young researchers spend 80 percent of their time performing data-related tasks and only 20 percent "doing meaningful science." Through the proposed activities and concrete steps described here, Unidata commits to work toward enabling that transformation.

Who We Are and What We Do

Unidata (http://www.unidata.ucar.edu) is a community data facility for the atmospheric and related sciences, established in 1984 by U.S. universities with sponsorship from the National Science Foundation (NSF). The Unidata Program Center (UPC), the program office for Unidata and the nexus of activities related to Unidata's mission, is managed by the University Corporation for Atmospheric Research (UCAR), a consortium of over 104 member universities and academic affiliates providing science in service to society.

Unidata exists to engage and serve researchers and educators dedicated to advancing the frontiers of Earth System science. The program's aim is to help transform the conduct of research and education in atmospheric and related sciences by providing well-integrated, end-to-end data services and tools that address many aspects of the scientific data lifecycle, from locating and retrieving useful data, through the process of analyzing and visualizing data either locally or remotely, to curating and sharing the results.

Specifically, the UPC:

- Acquires, distributes, and provides remote access to real-time meteorological data.
- Develops software for accessing, managing, analyzing, visualizing, and effectively using geoscience data.
- Provides comprehensive training and support to users of its products and services.
- In partnership with others, facilitates the advancement of tools, standards and conventions.
- Provides leadership in cyberinfrastructure and fosters adoption of new tools and techniques.
- Assesses and responds to community needs, fostering community interaction and engagement to promote sharing of data, tools, and ideas.
- Advocates on behalf of the community on data matters, negotiating data and software agreements.
- Grants equipment awards to universities to enable and enhance participation in Unidata.

Unidata is governed by its community. Representatives from universities populate standing and *ad hoc* committees that set policies for the program, provide first-hand feedback from users of program software and services, and offer guidance on individual projects. Non-voting representatives from government agencies also provide valuable information and advice. Unidata's governance structure ensures that the program stays in tune with the community it serves, and allows it to quickly adjust program priorities as the technological landscape and community needs change.

While Unidata's primary mission of serving universities engaged in atmospheric science education and research has remained unchanged through the years, the evolution and broad usefulness of its products and services have greatly enlarged its initial user base. Today, the Unidata community includes users from all sectors in over 200 countries, including nearly 2500 academic institutions and more than 80 research labs. Simultaneously, Unidata's activities and responsibilities have also grown as community needs have evolved. Despite the growth in users and enhanced scope of its activities, according to a 2010 survey conducted by the Unidata Users Committee, 97% of the respondents indicated that they were either satisfied or highly satisfied with Unidata's overall service to the community.

Last year, with community input and engagement of its governing committees, Unidata developed a strategic plan (http://bit.ly/121cdE1), creating a roadmap and a vision for the future. The overarching goal embodied in that plan is the creation of a scientific ecosystem in which "data friction" (Edwards, 2010) is reduced and data transparency and ease-of-use are increased. In such an environment, scientists will expend less effort locating, acquiring, and processing data and more time interpreting their data and sharing knowledge.

To accomplish the goals set forth in our strategic plan, Unidata will continue to build and provide infrastructure that makes it easy to discover, access, integrate and use data from disparate geoscience disciplines, allowing investigators to perceive connections that today are obscured by incompatible data formats or the mistaken impression that the data they need for their investigations do not exist. This proposal serves as an implementation plan for accomplishing those goals, moving our community toward the realization of our overarching vision.

Unidata Snapshot

The following tables provide a snapshot of the Unidata program in May 2013, with comparisons to metrics from the previous NSF proposal (where available).

Table 1: Data Services	2008	2013
Institutions Participating in the IDD network[1]	170	263
Host machines on the IDD network[1]	460 (250 unique sites)	520 (232 unique sites)
Data streams in the IDD	22	34
Approximate volume of data ingested into the IDD	100 GB/day	315 GB/day
Volume of data pushed to IDD sites	2.7 TB/day	13 TB/day
Volume of data pulled via remote access protocols	44.3 GB/day	659 GB/Day
Uptime of UPC data and support infrastructure	99.96%	

Table 2: Software Package Downloads	2008-2012
GEMPAK	11,801
IDV	43,944
Local Data Manager	11,455
McIDAS	857
netCDF-C Libraries (includes FORTRAN, C++ support) [2]	408,250
netCDF-Java Libraries (Common Data Model)	89,416
THREDDS Data Server	7,176
UDUNITS	26,265

Table 3: Workshops	2003-2007	2008-2012
Training Workshop Participants	335	475
Training Workshop Courses	38	33
Users Workshop Participants	164	183
Regional Workshop Participants	44	162

Table 4: Miscellaneous	2013
Number of community members registered with Unidata	44,231
Number of countries where Unidata software and services are used	214
Number of academic institutions participating (U.S.)	743
Number of academic institutions participating (worldwide)	2,495
Number of organizations participating worldwide	3,032
Number of community electronic mail lists	58
User support e-mail transactions, 2008-2012	21,388
Number of Community Equipment Awards, 2008-2012	30
Average staff FTEs at the UPC, 2008-2012	25

[1] These metrics are limited to sites of which Unidata is aware. Sites can use the LDM and participate in the IDD without reporting statistics to the UPC; we *suspect* the number of unreported sites is large.
[2] UPC source code downloads only. This number does not include downloads from repositories at the University of Kyoto and on the Github site, or binary distributions available via package managers on UNIX-like systems.

Unidata Focus Activities

Unidata Program Center staff are involved in a wide variety of projects in support of the Unidata community, including data distribution, software development, user support and training, and community service and leadership activities. The activities undertaken by the UPC span years or decades and affect thousands of users around the world. In a different context – without the unifying locus of the Unidata program – one could imagine many of Unidata's projects as candidates for NSF funding in their own right. As it is, our projects are able to build on and leverage each other's capabilities, advancing our community's goals more effectively as a result. The Plan of Action that follows describes how activities in these areas will address Unidata's strategic goals for the empowerment of its broad community.

Data Distribution

The UPC coordinates the Internet Data Distribution system (IDD), in which hundreds of universities cooperate to disseminate near real-time earth observations via the Internet. While the "push" data services provided by the IDD system are the backbone of Unidata's data distribution services, the UPC also provides on-demand "pull" data services via THREDDS, ADDE, and RAMADDA data servers.
The UPC's data servers are not classified as "operational" resources, but they nonetheless have a 99.96% uptime record and are used heavily by educational sites that lack the resources to store IDD-provided data locally, or to operate their own data servers (see Table 1). UPC's servers are housed in a UCAR co-location computer facility for reliability, and share UCAR's Internet2/National Lambda Rail connectivity, which provides access to ample bandwidth for Unidata's needs.

Software Development

A variety of software packages are developed, maintained, and supported by the UPC:

NetCDF

Unidata's netCDF (network Common Data Form) is a freely distributed collection of data access libraries that provide a machine-independent data format that is self-describing, portable, scalable, appendable, sharable, and archivable – all important qualities for those who wish to create, access, and share array-oriented scientific data. NetCDF permits easy access to array-based, multi-dimensional datasets, a task that can be difficult when using other common storage schemes. NetCDF has been adopted widely by the atmospheric sciences community, and is especially popular among climate and ocean modelers. For example, model output datasets for the Fifth Assessment Report of the Intergovernmental Panel on Climate Change must be submitted in netCDF format, using the associated Climate and Forecast (CF) metadata conventions. The resulting large base of netCDF users and data has led to support for the format in more than 80 open source packages and many commercial applications including MATLAB and IDL.

Common Data Model & THREDDS Data Server

Unidata's Common Data Model (CDM) provides an interface for reading and writing files in netCDF and a variety of other scientific data formats. The CDM uses metadata to provide a high-level interface to geoscience-specific features of datasets, including geolocation and data subsetting in coordinate space. Unidata's THREDDS Data Server (TDS) builds on the CDM to allow for browsing and accessing collections of scientific data via electronic networks. Data published on a TDS are accessible through a variety of remote data access protocols including OPeNDAP, OGC Web Map Service (WMS) and Web Coverage Service (WCS), NetCDF Subset Service (NCSS), and HTTP.

The CDM and TDS are widely used in the United States (by NOAA, USGS, NASA, and the Earth System Grid, for example) and internationally, and are part of the deep infrastructure on which next generation capabilities are being built by other organizations. Additionally, many other tools build on the

CDM (NOAA's ERDDAP, NASA's Panoply, and CMAS' VERDI, are examples) and on the TDS (NOAA PMEL's LAS and Ferret-TDS, for example).

Integrated Data Viewer

Unidata's Integrated Data Viewer (IDV) is a 3D geoscience visualization and analysis tool that gives users the ability to view and analyze a rich set of geoscience data in an integrated fashion. The IDV brings together the ability to display and analyze satellite imagery, gridded data (such as numerical weather prediction model output), surface observations (METARs), upper air soundings, NWS NEXRAD Level II and Level III RADAR data, NOAA National Profiler Network data, and GIS data, all within a unified interface. The IDV integrates tightly with common scientific data servers (including Unidata's TDS) to provide easy access to many real-time and archive datasets. It also provides collaborative features that enable users to easily share their own data holdings and analysis products with others.

AWIPS II & GEMPAK

AWIPS II is a weather forecasting, display, and analysis package currently being developed by the NWS and NCEP. Because many university meteorology programs are eager to use the same tools used by NWS forecasters, Unidata community interest in AWIPS II is high. UPC staff have worked closely with NCEP staff during AWIPS II development in order to devise a way to make it available to the university community.

NCEP has stated that GEMPAK applications will be migrated from GEMPAK/NAWIPS into AWIPS II for the National Centers. The UPC will likewise facilitate a migration from GEMPAK/NAWIPS to AWIPS II for the university community.

Rosetta

The Rosetta project at the UPC is an effort to improve the quality and accessibility of observational data sets collected via datalogging equipment. The initial goal of Rosetta is to transform unstructured ASCII data files of the type commonly generated by datalogging equipment into the netCDF format, while minimizing disruption to existing scientific workflows.

Local Data Manager

The Unidata Local Data Manager (LDM) system includes network client and server programs designed for event-driven data distribution. It is the fundamental component of the IDD system. The LDM is used by hundreds of sites worldwide, and is integrated into the National Weather Service's AWIPS II package.

McIDAS

The Man-computer Interactive Data Access System (McIDAS) is a large, research-quality suite of applications used for decoding, analyzing, and displaying meteorological data. The older McIDAS-X system, developed by the University of Wisconsin's Space Science Engineering Center and supported by Unidata, is gradually being replaced by the IDV and by McIDAS-V (which is based on the IDV).

UDUNITS

Unidata's UDUNITS supports conversion of unit specifications between formatted and binary forms, arithmetic manipulation of units, and conversion of values between compatible scales of measurement.

RAMADDA

The Repository for Archiving, Managing and Accessing Diverse Data (RAMADDA) is a vibrant and growing technology initially developed by Unidata and now managed and developed as an open source project. Unidata integrates RAMADDA functionality into the IDV, provides training and support, and contributes code to the project. In addition, Unidata makes extensive use of RAMADDA to support community and collaborative projects, and actively facilitates its deployment in the university community.

User Support and Training

Users of Unidata software and data rely on the UPC for comprehensive support services. UPC software developers handle user support directly, together responding to an average of more than 250 technical support questions each month. Developers also create product documentation, training materials, and provide hands-on training workshops each year. The workshops, which have been attended by more than 450 participants from the university, government, and commercial spheres in the past five years, provide UPC developers with an excellent opportunity to interact with software users face-to-face.

Cyberinfrastructure Leadership

Unidata community members look to the UPC not only for technological solutions, but for guidance on emerging trends in cyberinfrastructure and to represent their interests in collaborations with standards bodies and organizations that work across scientific disciplines. As standards-based solutions have become increasingly important to the conduct of international science, Unidata has assumed a central role in identifying and articulating standards, conventions, and data formats. Unidata's standards efforts have enabled ongoing collaboration with dozens of international organizations – especially those represented in the OGC MetOceans, Earth System Science, and Hydrology Domain Working Groups.

Community Service Activities

The UPC undertakes a wide variety of activities aimed at fostering a shared vision for and community ownership of the program and encouraging community input into its operation and direction. Bringing the community and stakeholders together to share knowledge and address problems that are important to them through meetings, workshops, conferences, and other venues is a key aspect of Unidata's community service mission, as are efforts to disseminate information of interest to community members.

Results of Prior Support Under NSF 0833450

A great majority of activities at the UPC are sponsored under the five-year core-funding award "Unidata 2013: A Transformative Community Facility for the Atmospheric and Related Sciences" (NSF 0833450). The period of performance for that award is 1 December 2008 to 30 November 2013.

The core-funding proposal identified six thematic areas as the focus of Unidata's efforts:

1. Broadening participation
2. Advancing data services
3. Developing and deploying useful tools
4. Enhancing user support services
5. Providing leadership in cyberinfrastructure
6. Promoting diversity

During the period of performance for the current award, Unidata met most of the goals described in that proposal, accomplishing many of the stated objectives.

Unidata significantly increased the reliability, volume, and variety of data provided to the community, extended the functionalities of its software, developed and deployed new tools and services, and actively engaged its growing user community. The program's contributions and accomplishments have had a demonstrable and sustained impact on the geosciences community, empowering faculty, students, and researchers to be more productive and enhancing their ability to advance science and learning.

Key Achievements 2008-2013

The following list highlights, in no particular order, Unidata's most significant accomplishments during the past five years.

- **Data Delivery.** The volume of data input to Unidata's IDD cluster has more than tripled since 2008, now averaging over 13 GB/hr; data output averages more than 540 GB/hr. Numerous new datasets were added to the IDD system, including higher resolution WSR-88D products (February 2010), Global Lightning Network (GLN) and North American Precision Lightning Network (NAPLN) (April 2011), Fire Weather products (September 2011), and Rapid Refresh (RAP) grids (replacing RUC grids April 2012).
- **OGC Standards.** In 2011, the netCDF Classic data model was accepted as an OGC core standard. The binary encoding for the classic data model was established as the first extension to the netCDF core standard, followed by the enhanced data model and the CF (Climate and Forecast) conventions extension in 2012 and 2013. These advancements encourage broader international use and greater interoperability among clients and servers interchanging data in netCDF format and will make the large collections of environmental netCDF data more readily accessible and usable.
- **NetCDF.** Unidata released netCDF-4 in late 2008, incorporating an expanded data model, compression, parallel I/O, multiple unlimited dimensions, user-defined data types, and other features supported by the HDF5 format. The NOAA GOES-R project will make all L2 and L2+ products available in netCDF-4 format. NetCDF is also used in other disciplines, including chromatography, neuro-imaging, molecular dynamics, and fusion research.
- **Integrated Data Viewer.** Among a wide variety of improvements to the IDV were support for analysis and visualization of ensemble model output, improved handling of sub-minute data and long range climate data, and addition of trajectory functionality. The IDV is now in use at more than 180 US-based, degree granting universities, colleges and technical schools, and at over 200 academic institutions abroad.
- **AWIPS II.** The UPC has been actively collaborating with NCEP and the NWS Office of Science and Technology to ensure that the AWIPS II software will be available to the academic community soon after it is operational. To that end, the UPC has been working to implement AWIPS II software on the types of computer systems that would typically be available in universities, and has partnered with universities in a beta testing program. The UPC is preparing to release the software more broadly to the university community by the end of 2013.
- **EarthCube Activities.** UPC staff have been active in NSF's EarthCube initiative, contributing several whitepapers, participating in the EarthCube charrettes and meetings, and being involved in four EAGER-award projects that generated roadmaps for EarthCube's further development. In December 2012, the UPC organized and hosted an NSF-sponsored workshop titled "Shaping the Development of EarthCube to Enable Advances in Data Assimilation and Ensemble Prediction."
- **Equipment awards.** During 2009-2013, the UPC provided 30 equipment grants to 27 universities, encouraging 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. The facilities developed with funding from these awards are transforming faculty instruction and student learning at those universities in a demonstrable manner.
- **Community Workshops.** In addition to its annual training workshops, the UPC organized and hosted triennial Users Workshops in 2009 and 2012, each of which brought nearly 100 community members together to share tools, techniques, and educational strategies. UPC staff also helped community universities organize and present regional workshops at California University of Pennsylvania and San Jose State University.

- **Move to Open Source Development Methods.** The UPC software team has adopted Open Source development methods including community access to source code, issue tracking, and release planning information to encourage community participation and contribution.
- **RAMADDA.** RAMADDA was originally developed by Unidata and released during the award period. RAMADDA is an information management and data repository framework for the geosciences that provides a turnkey system enabling data providers and users to upload, manage, and share data holdings.
- **GOES Data.** The UPC continues to ingest GOES-East, West, and South America imager data for injection into the IDD. GOES imagery remains one of the most-used data services in the Unidata suite, serving over 3 TB of data per month, and providing support for major field campaigns.
- **THREDDS Data Server and Common Data Model.** Unidata's TDS and its underlying CDM library have matured significantly and seen major enhancements in the last five years. As a result, the TDS is now in use in a large number of universities, organizations, and data centers; many third-party tools build on the CDM library.

Additional Highlights 2008-2013

In addition to the above-listed achievements, the following sections provide a snapshot of proposed objectives from NSF 0833450 and a brief summary of achievements in each of the six thematic areas.

Broadening Participation and Expanding Community Services

Objective	Accomplishments
Foster community ownership of the program	Moved several Unidata software projects to the Open Source environment Github. We are already seeing the fruits of this in the form of code contributions from community members.
Entrain community colleges into the Unidata community	Unidata now has 44 community college members. Provided equipment awards to Lyndon State College and Madison Area Technical College.
Bring community together through workshops, meetings, conferences, etc.	Hosted an NSF-sponsored EarthCube workshop focused on the meteorological domain in December 2012. Hosted and participated in several AccessData science education outreach workshops 2008-2010. Hosted OGC Technical Committee Meeting September 2011.
Develop global partnerships with geoscience data providers	Established working relationships with the British Atmospheric Data Center, ESSI Labs (Italy), ECMWF, EUMETSAT, and CPTEC (Brazil).

Advancing Data Services

Objective	Accomplishments
Develop and provide high-level interfaces to geoscience data	The Common Data Model (CDM) allows data servers to provide seamless access to data in a wide variety of geoscience formats (netCDF, GRIB, GEMPAK, *etc*.)
Adopt, develop, and promote open standards, conventions, and protocols	Worked closely with the OGC to establish netCDF as an international data standard. Actively participated in efforts to standardize and implement the Climate and Forecast (CF) conventions for metadata, which are used by the IPCC and many others. Implemented OPeNDAP client access.
Enable users with data holdings to contribute and share their data easily	The TDS now makes it possible to publish a wide range of scientific data for access by any client that implements the OPeNDAP protocol. RAMADDA makes it easy to publish small to medium-sized datasets. Both servers integrate tightly with the IDV.

Provide frameworks for creating dynamic case study datasets	RAMADDA facilitates the creation of dynamic case study materials by providing the ability to "federate" servers so that data from different locations appears as part of a single dataset.

Developing and Deploying Useful Tools

Objective	Accomplishments
Enhance the usefulness of the Integrated Data Viewer (IDV)	Major releases spanning versions 2.6 to 4.0. The IDV can now access and visualize every data stream provided by Unidata, including ensemble model output, sub-minute data such as lightning flash data, and long-range climate data. Collaborations with developers at other facilities have led to improved graphics capabilities.
Continue to package and support GEMPAK for the university community	Major releases spanning versions 5.11 to 6.7. Began distributing GEMPAK releases as source-code rather than binary form in 2009. GEMPAK usage has remained strong, with downloads averaging more than 200 per month in 2012.
Continue to support McIDAS-X for the university community	Unidata continues to support and release versions of McIDAS-X, which is developed at the University of Wisconsin-Madison's Space Science Engineering Center.
Build expertise with the AWIPS II system	AWIPS II has not yet been released to the Unidata community by NCEP. UPC developers are actively collaborating with NCEP to refine the AWIPS II package and prepare a version for release to universities. Five universities are participating in a beta test program.
Deploy RAMADDA	The RAMADDA scientific content management system was released to the Unidata community. Ongoing development has been transitioned to the open source community, and the UPC continues to provide support and training to the university community.
Develop new tools to make it easier to access geoscience data	Initial work on the Rosetta project began in 2011. Rosetta is a web-based service that will provide the ability to read from and write to a wide variety of scientific data formats.

Enhancing Support Services

Objective	Accomplishments
Foster effective communication among community members	In addition to traditional electronic mailing lists, the UPC has begun using blogs and social media channels communicate with community members. Ninety technical articles are now available on the Unidata developers' blog.
Harness advances in online collaboration technologies	The UPC has adopted open source community tools like Github to foster collaboration with software developers individually or in other organizations around the world.
Simplify download, installation, and maintenance of supported packages	Made major improvements to the process of building and using netCDF in the Microsoft Windows environment. Streamlining the process of installing and configuring web-based services like the TDS, RAMADDA, and Rosetta.
Document available datasets, datastreams, and services	Created a series of data flowcharts to clarify the relationship between Unidata tools and different types of data provided via the IDD.

Providing Leadership in Cyberinfrastructure

Objective	Accomplishments
Provide stewardship for standards, data formats, conventions, and protocols	UPC staff have worked closely with the OGC to recognize netCDF as an international standard and provided the C reference client for the OPeNDAP protocol standard. Staff continue to work to extend the netCDF standard.

Take a leadership role in setting future directions for geoscience cyberinfrastructure	UPC staff have been active in the NSF's EarthCube initiative, hosting a workshop in 2012. Staff have also worked with the OGC to define data encoding standards, advocated for solutions to problems with the WMO's GRIB format, and worked to advance adoption of the Climate and Forecast (CF) metadata conventions.
Help guide the evolution of international data system standards	Outreach collaborations by Unidata staff include: Marine Metadata Interoperability (MMI) Project Steering Team, IOOS DMAC Steering Team, CUAHSI Standing Committee, UCAR-wide representative to OGC Technical Committee, AGU ESSI Focus Group Board, ESIN Journal Editorial Board, Liaison to OOI Cyberinfrastructure Project, several collaborations with EarthCube teams, member of Steering Committee for international Ocean Data Interoperability Platform (ODIP), Chair of European Space Agency (ESA) Earth Observation Product Trees Project.
Serve as a testbed for deployment of emerging technologies	The UPC has worked closely with NCEP and Raytheon to test the AWIPS II system in preparation for non-operational university deployment.

Promoting Diversity by Expanding Opportunities

Objective	Accomplishments
Engage a highly diverse population of educators and researchers	Unidata's governing committees include educators and researchers from large and small institutions across the U.S. UPC staff engage with a variety of scientists worldwide through conferences, electronic communications, and organizations such as the AMS, AGU, and OGC. In order to reach out to students, the UPC has begun attending student conferences and career fairs, and continues to have a student representative on its Users committee.
Develop/promote specific opportunities to broaden participation from underrepresented communities	In partnership with UCAR's Community Development Program, Unidata staff made presentations to and worked with collaborators from Tribal College and Alaska Native communities for establishing closer ties between UCAR and those communities. Made equipment awards to HBCU and HSI institutions.
Continue participation in the UCAR-led SOARS program	UPC staff have participated actively as mentors for three SOARS protégés over four of the past five years, and will continue in 2013.

Broader Impacts

Unidata's predominant impetus is service to its community. As shown in Table 4 (page 3), Unidata's products and services are in use by over 3000 organizations including nearly 2500 academic institutions in over 200 countries (see a map online at http://www.unidata.ucar.edu/about/usage/index.html#map). Many dozens of research labs, national and international agencies, weather services around the world, and projects large and small benefit from Unidata technologies. Unidata's scientific software libraries are routinely used by scientists and service providers in most geoscience disciplines; they are embedded in more than 20 commercial and 100 open source software packages.

Although our core activities focus on serving scientists and educators in the atmospheric and related sciences, virtually every project Unidata undertakes has a broader impact on the geosciences community and society at large. This section touches on some of the ways Unidata enables research in the geosciences, enhances education in universities and colleges, provides support to our collaborators and users of our software, and strives to build a larger and more active community.

Impact on Research

The impact of Unidata's data systems, software, and services on scientific research extends far beyond our core atmospheric science community. The number and diversity of publications in peer-reviewed journals that cite Unidata or its software is one compelling measure of the growing impact of Unidata's services on research.

For example, a survey of papers published between 2008 and 2012 in journals of the American Meteorological Society shows 110 citations of Unidata software and data services. In the same period, an additional 52 papers published in journals of the American Geophysical Union cited Unidata software and data services. In both cases, papers published in the past five years account for nearly half the total number of citations recorded over more than a quarter century of Unidata history – an indicator of the growing reach of Unidata's software and services.

A review of citations reported by the Google Scholar search engine revealed similar results. Between 2008 and 2012, Unidata software and data services were cited 1150 times in the full range of scholarly literature encompassed by the search engine. Interestingly, Google Scholar returned an additional 3690 articles mentioning netCDF, but which did not include mention of its connection to Unidata. This correlates with anecdotal evidence of widespread netCDF use beyond the communities traditionally served by Unidata. Again, comparing the number of citations recorded in the past five years with a complete history indicates that roughly half of the citations have come in since 2008.

In 2012, the UPC surveyed longstanding academic community members to find what types of research Unidata software and data services are enabling today. Among the comments were:

- *We use Unidata software/facilities quite extensively from getting global model analysis/forecast fields in real time for high-resolution coupled model development and forecasting experiments, getting satellite data for analysis, and to display our model results in IDV.*
- *Several Professors are involved with field programs that have intensive observing periods that include aircraft, etc. Unidata-provided weather and model information is integral (anywhere from minor to absolutely essential) to the field program and operational decisions. In addition, software and real-time data provided by Unidata is part of the post-field program analysis.*

Looking beyond the self-reported activities of Unidata community members, a survey of some recent academic journal citations highlights the use of Unidata technologies in a wide range of contexts including:

- The Gulf of Mexico Coastal Ocean Observing System (Simoniello *et al*., 2011)
- Algorithms for Detecting and Tracking Tropical Cloud Clusters (Hennon *et al*., 2011)
- Bridging the gap between Hydrologic and Atmospheric communities (Boldrini *et al*. 2012)

Impact on Education

As mentioned earlier, Unidata software and data services are in use at nearly 750 U.S. colleges and universities and more than twice that many in other countries. Unidata reaches across geoscience disciplines to provide data and tools to researchers, educators, and students in the atmospheric sciences, hydrology, and oceanography fields, among others. From research universities to community colleges, Unidata provides timely support and service to the academic organizations that are training the next generation of geoscientists. In the process, it is estimated that more than 100,000 users are exposed to products generated using Unidata software and systems, and more than three quarters of all graduate students in the atmospheric and related sciences now use software provided by Unidata.

For example, Pennsylvania State University considers Unidata's IDD feed, along with analysis/display software packages including GEMPAK/NAWIPS and IDV, to be vital tools for research, instruction and outreach. Unidata software allows students to explore current and past weather scenarios as part of upper-level undergraduate meteorology courses, and is also used for the generation of graphics for the popular and publicly-available Penn State electronic map wall.

Unidata strives to promote diversity in the geosciences by supporting use of its technologies in a wide array of educational institutions. Unidata software and data services are used at universities in all 31 EPSCoR jurisdictions, including many institutions that have a large number of students from underrepresented communities.

Community Equipment Awards

Each year, the UPC sets aside $100,000 to fund the Unidata Community Equipment Awards program. The program provides funds to encourage new geoscience departments to join the Unidata community and to allow existing members to continue and enhance their participation. During the past five years, 30 awards were made to a diverse group of institutions, from major U.S. research universities to community colleges to educational institutions in Europe and Central America.

Projects funded in the past five years include:

- **Madison Area Technical College** used award funding to introduce modern weather software into its meteorology curriculum, ingesting data from Unidata's IDD for display using McIDAS and IDV.
- The **Coastal Ocean Observation Lab at Rutgers University** (RU-COOL) used award funding to install Unidata and other open source technologies to collect, process, and make available a wide range of ocean data for use by students and researchers.
- **Iowa State University** used award funding to establish THREDDS and RAMADDA servers that provide access to the university's weather data archive to students and the Unidata community.

A complete list of projects funded under the Community Equipment Awards program and the many creative applications of Unidata software and systems by the recipient universities to advance education and research is available online at http://www.unidata.ucar.edu/community/equipaward/.

Impact on Other Organizations and Projects

Unidata-developed cyberinfrastructure, in addition to being used widely in universities to advance education and research, also provides a substrate for other stakeholders in federal agencies, the private sector, and many non-governmental and international organizations. For instance, many data services in NOAA, NWS, NASA, USGS, DOE, DOD, NCAR, ECMWF, EUMETSAT, CMA, and CPTEC are built on the formats, software, and data systems that Unidata has developed. Unidata systems and technologies are integral parts of numerous large-scale projects, including SuomiNet, THORPEX, GEON, EarthScope, IPY, and others.

The letters of commitment to collaborate from some of Unidata's partners, provided in Appendix A, provide a view of how Unidata's work is having a positive impact on those organizations.

Some additional examples not previously mentioned:

- **ESRI** has made support for netCDF and the Climate and Forecast metadata conventions an important component of its ArcGIS software package.
- Numerous **NASA EOSDIS Data Centers** (PO.DAAC, ORNL DAAC, GES DISC *etc*.) are using the THREDDS Data Server to make geoscience datasets available to researchers and the public.
- **Weather Decision Technologies** of Norman, OK uses Unidata's LDM to retrieve weather data that is fed to its family of iMap mobile and online weather applications.
- Over the course of **NASA's Space Shuttle program**, the LDM was used to transport observational and experimental data from a variety of sources. The UPC received a commendation from the NWS's Spaceflight Meteorology Group in 2011 for providing LDM software and technical support for both the LDM and McIDAS packages. (Read more online at http://bit.ly/nl4e1o)

- Unidata, in collaboration with sister UCAR entities and external institutions, participates in the **Google Africa Initiative** Meningitis Project. Unidata technologies including LDM, IDV, RAMADDA, THREDDS, and McIDAS ADDE are used in the collection and dissemination of data relevant to the project. (Read more online at http://bit.ly/103dVls)
- The US **Integrated Ocean Observing System** (IOOS) has deployed a total of 17 THREDDS Data Servers at both academic and government institutions. At each location, NcML is used to deliver model output via OPeNDAP as CF-compliant datasets.
- Within days after the Fukushima earthquake and tsunami, the OceanNOMADS group at the **National Coastal Data Development Center** deployed a THREDDS Data Server that made it possible to deliver output from a US Navy model (predicting the transport of contaminated material in the ocean) to ships in the region. The TDS let scientists retrieve just the data in the vicinity of their ship, allowing them to predict where the water they sampled would be the following day.
- The **MyOcean project**, a part of the European Commission's Global Monitoring for Environment and Security (GMES) program, uses THREDDS Data Servers to provide WMS, OPeNDAP, and other data access and viewing capabilities for users of data collected by MyOcean provider sites. (Read more online at http://bit.ly/Yerfqz)

Building Community

Unidata undertakes a variety of activities with the goal of building a vibrant community in the geosciences and beyond. The following are a sampling of activities not previously mentioned:

- **Scientific Society Meetings**
  Unidata staff are active in convening sessions and making presentations at AGU, AMS, and EGU meetings as well as at other national and international conferences and workshops. UPC staff members helped create AGU's Earth and Space Science Informatics session in 2004; participation in the AGU ESSI program has grown to over 400 abstracts for the 2012 Fall Meeting. The EGU ESSI Division was formally launched in 2008 with the active involvement of UPC staff; it has grown significantly, receiving over 300 papers at the 2013 EGU Geosciences Meeting.
- **GALEON**
  Unidata has been a core participant in the Open Geospatial Consortium GALEON (Geo-interface for Air, Land, Earth, Oceans NetCDF) Interoperability Experiment. GALEON aims to specify and use standard interfaces to foster interoperability between data systems used by the traditional GIS community and those in the atmospheric and oceanographic science or Fluid Earth Systems (FES) communities. (Read more online at http://bit.ly/12WI75U)
- **Helping Community Members Reach the Public**
  Unidata technologies help community members reach out to their own communities, facilitating the provision of meteorological data and displays through dozens of popular web sites. For example, the College of DuPage, Iowa State University, University of Wyoming, University of Oklahoma, and University of Utah's Mesowest all make extensive use of Unidata services in their outreach. In addition, several museums (the Boston Museum of Science and San Francisco's Exploratorium among them) make use of either data or software provided by Unidata.

Synergistic Activities

For the benefit of the geoscience community, the UPC participates in projects and undertakes activities that are funded through non-core awards. These projects are always consistent with Unidata's mission, and synergistic with ongoing core-funded activities. We believe that these collaborations are essential to maintaining a healthy program, and that many benefits accrue to our community. In addition to the intrinsic merit of the projects, non-core funding brings modest additional resources to the UPC (contributing about 10% of overall funding), relieving some of the pressure on Unidata's budget. Both the Unidata Policy Committee and NSF have encouraged Unidata to participate in synergistic opportunities as appropriate. The UPC undertakes non-core projects only after careful analysis of their merit and benefits to our core community, and endorsement by the Policy Committee.

Historically, synergistic projects have created new capabilities, provided new datasets for the community, and leveraged ongoing activities. Here are some examples of previous projects:

- The NWS provided funding for creation of COMET-Unidata case study data sets.
- Initial THREDDS development was funded by NSF/EHR under the NSDL initiative.
- The LEAD project motivated the development of uploading capabilities in RAMADDA.
- The integration of HDF5 into netCDF occurred as a result of a NASA-funded project.

Currently, Unidata is participating in the following two non-core projects:

OPULS: The OPeNDAP-Unidata Linked Servers project, funded by NOAA, reflects the intentions of the two organizations to better align, link, and eventually integrate software that they independently offer as open source. The union of the capabilities provided by the TDS and Hyrax software offerings has become increasingly important as basic infrastructure for scientific data exchange.

ACADIS: The Advanced Cooperative Arctic Data and Information Service is a collaborative effort between NSIDC, NCAR, and Unidata. It provides data management support and archival services for the Arctic Observing, Arctic System Sciences, Arctic Natural Sciences, and the Arctic Social Sciences Program. ACADIS is providing data ingest and access services to scientists, decision-makers, and other Arctic stakeholders, as well as archival services to ensure data accessibility through the coming years and decades. Unidata's Rosetta project grew out of work for the ACADIS project.

Cyberinfrastructure in Context

Facilitating and advancing research and education in Earth System sciences in general – and atmospheric sciences in particular – has been and will remain Unidata's raison d'être. While our overarching mission remains unchanged, Unidata must be cognizant of and responsive to current and future global science priorities that shape the nature and scope of research; the changing technological landscape and resource-constrained environment in which research and education are conducted; the backgrounds, skill sets, and expectations of students entering the profession; and the educational goals of all stakeholders. In proposing to both sustain and extend the Unidata Program as a cornerstone data facility for the atmospheric and related sciences, we present here a brief scan of the environment that influences not only our current activities, but also helps to shape the future direction and goals of the Program.

Science

Society is grappling with abrupt climate change and its effects, extreme weather and impacts, and water cycle changes. In response to these "grand challenge" problems, the geoscience community is shifting its emphasis from pure disciplinary research to a more balanced mix that advances disciplinary knowledge while looking to apply research results to interdisciplinary questions touching both science and society.

In 2009, UCAR surveyed more than 15,000 members of the atmospheric and related sciences community. Of the over 2000 responses, nearly 70% suggested increasing the emphasis on interdisciplinary research. The top request for additional UCAR/NCAR services was for additional datastreams and datasets, ahead of community models, educational materials, community workshops, and observing facilities.

Given the changing scientific landscape and emphasis, it is imperative that Unidata develop and support cyberinfrastructure that not only enables researchers to advance the frontiers of science, but transcends traditional disciplinary boundaries. The main challenge we see is providing easy access to the right data, in the right format, to the right software applications.

Education

For decades, the research community has harnessed the power of data and computers to better understand Earth System problems through complex models, visualizations, and analysis techniques. Educators are increasingly integrating data-driven exploration into the learning environment, and easy access to data for student exploration is crucial to making this transformation. Professors realize that authentic learning and effective pedagogy focus on solving real-world problems that engage student interest and increase understanding of phenomena and processes. Investigation, research, analysis, and discovery of natural phenomena make the geosciences an ideal platform for integrating authentic learning activities into the curriculum, and many atmospheric science programs have successfully done so.

Data-enabled learning encourages student projects to enhance the collection and analysis of data through smart tools and sensors that automate the capture, recording, processing, and sharing of results. As an example, the Android operating system now provides support for atmospheric sensors that can monitor a variety of environmental properties. With these sensors and the geolocation facilities of their cell phones, students can engage in micro field studies and data-collection activities. Intercomparison of the crowd-sourced data gathered with such sensors is an ingenious way to advance data literacy and active learning.

It is critical to the health of our profession that future science and engineering leaders understand modern cyberinfrastructure and be trained using state-of-the-art tools, technologies, and techniques. Fluency is required not only in the geosciences and mathematical/statistical disciplines, but also in computational and information technology areas. Twenty-first century scientists must be data literate through many aspects of the data life cycle, including collection, management, analysis, and sharing of scientific data.

Data

Data-intensive science has emerged as the Fourth Paradigm of scientific discovery after empirical, theoretical, and computational methods. This is particularly true in the geosciences, where data have become increasingly important in scientific research. Modern data volumes from high-resolution ensemble prediction/projection/simulation systems and next-generation remote-sensing systems like hyper-spectral satellite sensors and phased-array radars are staggering. For example, CMIP5 alone will generate more than ten petabytes of climate projection data for use in assessments of climate change. NOAA’s National Climatic Data Center projects that it will archive over 350 petabytes by 2030.

For researchers and educators, this deluge and increasing complexity of data brings challenges along with the opportunities for discovery and scientific breakthroughs. Retrieving relevant data in a usable format from an archive should not be more time consuming and arduous than the scientific analysis and investigation the data make possible. At the other end of the spectrum, the majority of agency-funded research is conducted by scientists in relatively small projects with one lead researcher, typically a faculty member with part-time commitment to the project, and one or two graduate students or part-time post-docs. While great care is frequently devoted to the collection, preservation, and reuse of data on large, multi-investigator projects, relatively little attention is paid to curating and sharing data that is being generated by these smaller projects (Heidorn, 2008), resulting in large amounts of “long tail” or “dark” data. As a result, it is difficult to discover unpublished “dark data,” which might remain unshared, underutilized, or in some cases even be lost. By some estimates, only 5% of the data generated by individual PIs in the geosciences is shared with the broader community (Killeen, 2011).

Publication of research data presents unique challenges, sociological and technical, for the science community. Society has begun to demand increased scientific transparency, but researchers face a lack of incentives, inadequate resources, intellectual property issues, a culture of protectiveness, and the absence of supporting cyberinfrastructure, tools, and data repositories for sharing data. In 2011, NSF issued a new guideline requiring that every proposal submitted include a Data Management Plan describing how the project will conform to NSF policy on the dissemination and sharing of research results. Recently, the Office of Science and Technology Policy (OSTP) directed federal agencies to develop a plan to support increased public access to the results of research funded by that agency (read online at http://1.usa.gov/VBkFJv). Among other elements, such a plan must provide a strategy for improving the public's ability to locate and access data resulting from federally funded research.

Campus Information Technology Infrastructure

Educational enterprises are in a state of rapid transformation. Since the introduction of personal computers in the 1980s and high-speed networking on college campuses in the 1990s, there has been a sea change in the IT environment in which students and faculty operate. The World Wide Web and the availability of ubiquitous wireless access, smartphones, tablets, and cloud-based services have accelerated the shift. The expectations, modalities, and skills of students have dramatically changed. As Conford (2008, http://bit.ly/18PahAH) argues, "in the past, universities and colleges were often the institutions that provided students with their first experiences of networked information technology services such as email and easy access to the web. Today, however, students arrive at universities and colleges with years of experience of these technologies. As a result, the ways in which individuals use technologies, and their expectations about how they are going to use those technologies, are already well established. Institutions no longer introduce users to information technology; instead, information technology is often the main context in which users are introduced to the institution."

Increasingly, students are not only allowed but encouraged to bring their own devices and connect them to campus networks. Students expect content and services to be delivered through their devices; they won't accept a step back in technology when they step onto campus. They expect to use interactive, intuitive, and collaborative tools to learn and communicate just as they do in their off-campus lives.

The computing environment in departments and faculty research labs is also changing. Anecdotal evidence suggests that reduced budgets mean diminished system administration and maintenance support for many departments' scientific computing infrastructure. University administrators are realizing that the traditional model, in which discrete, dedicated computer systems perform specialized tasks, cannot be sustained and that new ways of delivering information technology and data services are needed. Choices driven by budget pressures include consolidation or centralization of IT systems, increased virtualization, and the adoption of cloud computing technologies to deliver services.

EarthCube

EarthCube, a multi-year NSF CIF21 initiative that began in 2011, is described as "a collaboration between the NSF and Earth, atmosphere, ocean, computer, information, and social scientists, educators, data managers, and more." The EarthCube vision is to transform the conduct of research through the development of community-guided cyberinfrastructure to integrate information and data across the geosciences. EarthCube is aiming to build a unified, adaptive, and scalable cyberinfrastructure framework that integrates all Earth system and human dimensions data in an open, transparent, and inclusive manner.

UPC staff have been actively involved in EarthCube since its unveiling; contributing four white papers (Domenico 2011, Miller *et al.* 2011, Ramamurthy 2011, Rew 2011), attending both EarthCube charrettes, organizing a community workshop "*Shaping the Development of EarthCube to Enable Advances in Data Assimilation and Ensemble Prediction,*" participating in four EarthCube EAGER projects that developed corresponding roadmaps for Governance, Brokering, Service-based Integration, and Cross-domain Interoperability. With a quarter-century's experience providing data services to the atmospheric science community, Unidata's involvement in the EarthCube effort is both necessary and mutually beneficial.

Vision for the Future: Moving Unidata's Services and Software to "the Cloud"

We have identified some of the challenges universities are facing: shrinking budgets, rapidly evolving information technologies, growing data volumes, multidisciplinary science requirements, and high student expectations. Most faculty and researchers would prefer to focus on teaching and doing science rather than setting up computer systems. These changes are upending traditional approaches to accessing and using data and software; Unidata's products and services must also evolve to support modern approaches to research and education. In this section, we present a vision for Unidata's future that will provide a transformative community platform for collaborative development and an array of innovative data services to our users.

After years of hype and ambiguity, cloud computing technologies have matured. Their promise is now being realized in many areas of commerce, science, and education, bringing the benefits of virtualized and elastic remote services to infrastructure, software, computation, and data. Cloud environments can reduce the amount of time and money spent to procure, install, and maintain new hardware and software, reduce costs through resource pooling and shared infrastructure, and provide greater security. Cloud services aimed at providing *any resource, at any time, from any place, using any device* are increasingly being embraced by all types of organizations. NOAA, NASA, and other federal science agencies are establishing cloud computing services. Universities are no exception; the University of Washington, University of Illinois, Cornell University, and George Washington University are some of the universities that have already set up cloud services for scientific and academic computing.

Given this trend and the enormous potential of cloud-based services, we propose to gradually augment Unidata's products and services to align with the cloud-computing paradigm. Specifically, we will work to establish a community-based development environment that supports the creation and use of software services to build end-to-end data workflows. The design will encourage the creation of services that can be broken into small, independent chunks that provide simple capabilities. Chunks could be used individually to perform a task, or chained into simple or elaborate workflows. The services we envision will be loosely coupled to meet user needs rather than tightly coupled into a monolithic system. The services will be portable, allowing their use in researchers' own cloud-based computing environments.

Unidata recognizes that its community is not monolithic. Our users have diverse needs and access to a range of cloud-computing resources. Users will be able to implement these services in conjunction with their own workflows in ways they want; for example, by leveraging the capabilities of Python, R, or other workflow systems they use. This approach permits greater flexibility and interoperability. We envision users being able to invoke and use the services from an array of computing devices, including laptops, high-powered workstations, tablets, and even smartphones. The proposed vision is not about building a system, but an environment with a collection of capabilities, using a standardized approach.

Categories of Services

Here we present a list of candidate cloud-based data services that can be enabled as extensions of Unidata's current activities. In no particular order:

- Remote access to real-time data streamed via the IDD, as a cloud service (*e.g.*, Unidata's "thredds" server in a cloud)
- Data discovery, access, and extraction services (*e.g.*, subsetting services, catalog search services)
- Catalog and metadata generation services
- Data manipulation and transformation services (*e.g.*, Rosetta, decoders, Unit conversion)
- Brokering services for data and metadata (*e.g.*, ESSI labs)

- Server-side data analysis and operations (*e.g.*, netCDF+ operators, time-series generation, other mathematical calculations)
- Data-proximate display and visualization services that provide products to thin clients
- Data publishing services allowing users to publish results, analyses, and visualizations
- Subset subscription services, providing delivery of specified custom products

Community Collaboration

Unidata will *not* build all of the needed "chunks," but provide a platform for creating a range of cloud-enabled geoscience data services. Our objective is to harness the community's vast scientific and technical expertise in making this transformation. In partnership with the community and collaborators, the UPC will define the broad directions and design the general architecture/framework for composable services, provide a platform for developing services, establish a governance process for the envisioned environment, and develop and add component services to the suite incrementally and systematically.

As always, Unidata will actively engage its community, providing encouragement to community members to contribute applications and services to the collection. The UPC will provide help in "wrapping" scientists' applications into web-enabled, pluggable services. Unidata will also develop and demonstrate end-to-end prototype data services and make them available to the community.

We recognize that the governance of this process is extremely important and will pose some challenges. We have already put in place key elements to enable this transformation: by developing several web services, moving to Open Source environments like Github and Redmine for collaborative development of software, and providing remote data access for more than 15 years.

Managing Change for Our Community

Unidata and its community are at a crucial juncture in history. The IT landscape is changing dramatically even as users face rapidly shrinking budgets; our supporting community is expecting Unidata to lead by innovating and enabling new capabilities that let them do more with the resources they have. It is imperative that Unidata present a bold vision in response to these needs. In doing so, Unidata assumes the responsibility to help community members make the transition as their workflows and *modus operandi* change. To maintain the trust of our community, we firmly believe we must follow the spirit of the medical profession's Hippocratic Oath: *First, do no harm*.

We propose that the UPC align its future activities around this vision, but take a measured and disciplined approach that eases the community's evolutionary transition to newer, more powerful ways of working.

Our long-term goal is to create rich, self-sustaining scientific research ecosystems in which data, services, and tools are nurtured by an engaged geosciences community. The full realization of this vision will likely take a decade or longer, but we expect to make demonstrable progress by the end of the proposal period. We believe that the cyberinfrastructure environment that will arise from the steps we are proposing has the potential to profoundly transform the conduct of research and education in the geosciences and beyond.

Plan of Action

As described in the Results of Prior Support section, Unidata has made many continuing contributions and initiated valuable new endeavors in the past five years. We propose to maintain those critical programs while embarking on initiatives in response to the community needs described above.

A facility like Unidata, which has been providing reliable services for a quarter century, faces a healthy tension between *sustaining* support for its successful legacy products, extending and *enhancing* products that hold promise for the future, and *innovating* via new, forward-looking endeavors that capitalize on advances in technology and respond to the evolving scientific landscape. Under tight budgets, it may be tempting to sustain only ongoing services and delay undertaking new development or initiating new endeavors. That approach risks products and services becoming obsolete or irrelevant. Allocating resources across these three areas requires not only careful consideration of the community's current needs, but an ability to anticipate emergent needs, and an appreciation of the delicate balance of priorities as the program fulfills its obligation to both lead and follow.

Unidata's strategic plan defines the program's mission, vision, and goals, and informs our decisions about how to allocate our limited resources. This section describes Unidata's strategic goals along with concrete actions planned for the near term and activities envisioned to take place over a longer time span. (You can read the Unidata strategic plan online at: http://bit.ly/121cdE1.) The descriptions below indicate whether the activity supports an ongoing service (*sustaining*), adds new features to an existing package or service (*enhancing*), or adds a new package or service to complement Unidata's existing portfolio (*innovating*).

Strategic Goal: Enable widespread, efficient access to geoscience data

Seamless access to data is essential for advancing education and research. To ensure that the geoscience research and education community gains access to the data it needs, Unidata will:

Distribute atmospheric and other geoscience data in real time

Distributing real-time meteorological data to the community is a foundational Unidata activity and remains one of its highest priorities (*sustaining*, *enhancing*, *innovating*). Planned activities include:

- Continue to operate robust IDD system top-level nodes, inserting data from NOAAPORT, CONDUIT, and GOES satellites (among others).
- Enhance the quality, quantity, and diversity of data available from UPC's remote access servers.

Develop innovative cyberinfrastructure solutions to facilitate dissemination of scientific data

Unidata's vision of future data-centric scientific workflows requires us to move beyond the current reliance on push technologies to deliver data (*enhancing*, *innovating*). Planned activities include:

- Work to make Unidata technologies function more effectively as server-side processing engines to facilitate data-proximate computations and analyses.
- Find ways to reduce the amount of data that must be transmitted over electronic networks before researchers can "do science," for example through subsetting and progressive disclosure.
- Integrate changes made by NOAA's AWIPS-II project into the Unidata LDM so that a single data-distribution system can be used in both contexts.

Work with data providers to make geoscience data freely available

Advocating for full and open access to scientific data for the university community has always been a core activity for Unidata (*enhancing*). Planned activities include:

- Work with governing committees and data providers to negotiate for access to data and model output deemed important for teaching and research.
- Provide easy mechanisms (e.g., RAMADDA and TDS) for investigators to share or publish their data for use by colleagues.

Develop and maintain the computing and networking infrastructure necessary to keep the growing volume of data flowing reliably and in a timely manner

While the UPC is not classified as an "operational" data center, we strive to provide reliable data service to the Unidata community (*sustaining*). Planned activities include:

- Continue operating UPC clusters and remote-access data servers, which have a 99.96% uptime record.
- Work with UCAR/NCAR information technology groups to ensure continued access to UCAR hardware and networking resources.

Strategic Goal: Develop and provide open-source tools for effective use of geoscience data

Faced with an abundance of scientific data, researchers and educators need well-integrated, state-of-the-art tools to access, analyze, manage, and visualize the data. Because our experience shows us that robust solutions arise from collaborative efforts, we will develop and support open-source software solutions to:

Analyze, integrate, and visualize geoscience data in two, three, and four dimensions

Unidata's scientific visualization tools are designed to make it easier for researchers, educators, and students to understand what heterogeneous geoscience data are telling them about the world (*enhancing*, *innovating*). Planned activities include:

- Integrate existing McIDAS-V functionality into the IDV to better handle data from new observation platforms such as the GOES-R satellite network. McIDAS-V excels at accessing and displaying satellite imagery; this development will bring similar capabilities to IDV users.
- Create a library of user functions for use in the statistical analysis of ensemble output.
- Collaborate with the SSEC VisAD team to update the library's 3D rendering API to incorporate a more modern, flexible, and maintainable rendering engine for the benefit of VisAD-based software including the IDV and McIDAS-V.
- Prepare for release the first publicly available version of AWIPS II, which will be bundled with the AWIPS II Development Environment for community developers who wish to create or edit plugins.
- Assist NCEP in testing GEMPAK applications within AWIPS II.

Enable visualization and effective use of very large data sets

As data volumes grow, Unidata visualization tools must take advantage of vast data holdings and new technologies that make those holdings accessible (*enhancing*, *innovating*). Planned activities include:

- Collaborate with the EarthCube community and others to enable end users to create, analyze, and visualize customized ensemble datasets from a variety of model outputs.
- Enhance the IDV's server-side functionality to make IDV capabilities available in environments that demand a "thinner" visualization client (tablets and other mobile devices, for example), expanding access to IDV display and analysis facilities to students and others who lack access to a traditional desktop scientific computing environment.
- Evaluate modifying or adding to existing AWIPS II software components to facilitate transfer of low-bandwidth graphics (images, contours, overlays, etc.) from a remote data server.

Access, manage, and share collections of data from diverse sources

Unidata's data access technologies are integrated into its visualization software products (*enhancing*, *innovating*). Planned activities include:

- Enhance the IDV's ability to interact with local and remote datasets in ways that facilitate access to and use of very large data collections. This will entail development of improved client and server infrastructure and enhanced data access algorithms.
- Develop functionality to allow regridding of model output onto a single common grid designated by the user, and to enable the aggregation of individual datasets into ensemble collections.
- Evaluate the addition of OPeNDAP to the AWIPS II system and the possibility of using the OPeNDAP protocol to access THREDDS Data Servers, similar to remote access through the IDV.

- Foster an early-adopter group of AWIPS II data server administrators to evaluate feasibility of serving data to dozens or hundreds of remote clients at once.
- Add a WRF initialization (and boundary condition specification) service to the TDS that subsets model data, automatically selecting the parameters needed for WRF initialization and prediction.

Strategic Goal: Provide CI leadership in data discovery, access, and use

The tools and techniques of distributed scientific computing are continually evolving. In order for the Unidata community to benefit from changes in the scientific cyberinfrastructure landscape, we will:

Develop useful data models, frameworks, and protocols for geoscience data

Unidata strives to advance data science for the good of the geoscience community (*enhancing*, *innovating*). Planned activities include:

- Create a streaming protocol to allow use of netCDF in web services, reducing the need to move large datasets to local machines.
- Reduce "Big Data" performance bottlenecks by improving chunking strategies, facilitating use of parallel I/O, using new metadata access strategies, and leveraging HDF5 storage layer advances.
- Add data access abstractions like swaths and unstructured grids, based on geospatial coordinates.
- Improve Python support across the Unidata software suite.
- Speed up the THREDDS catalog parse/build and improving the associated developer API.
- Provide a TDS configuration web interface and other tools to increase ease of use, improve scalability, and support dynamic reinitialization of servers.

Advance geoscience data and metadata standards and conventions

Standards and conventions allow investigators to fit the best technology for a given task into their own scientific workflow. Unidata strives to "play nice" with existing and emerging technologies, while advancing the state of the art (*enhancing*, *innovating*). Planned activities include:

- Create netCDF APIs and web services that allow access to data by geospatial coordinates.
- Provide full client (CDM), server (TDS), and library (netCDF) support for the new DAP4 data access protocol (once it is finalized).
- Work toward furtherance of netCDF technologies as international standards.
- Advocate for improvements or alternatives to the GRIB format for weather data and model output archives.

Evaluate emerging cyberinfrastructure trends and technologies, providing information and guidance to community members

Unidata must remain fully apprised of trends in technology that will affect our community in the future (*innovating*). Planned activities include:

- Explore and implement cloud computing solutions by gradually migrating Unidata applications to become services that are available from the cloud.
- Investigate faster compression techniques that achieve better compression than zlib for netCDF.

Facilitate discovery mechanisms for quickly finding and accessing geoscience data

As data volumes increase, it is ever more important that researchers be able to find and use the data that is available (*enhancing*, *innovating*). Planned activities include:

- Support access services from emerging computing platforms, including cloud, mobile, parallel, and data-intensive architectures.
- Add support for scientific workflows involving multiple datasets and virtual machine platforms.

- Improve handling of large collections of files (feature collections) and gridded datasets.
- Promote the TDS netCDF Subset Service (NCSS) and other existing data subsetting services from semi-experimental quality to production-quality services, improving their reliability, performance, and range of response types.
- Improve automated extraction and harvesting of metadata from datasets.

Strategic Goal: Build, support, and advocate for the geoscience community

Unidata works to help community members learn from each other by providing opportunities for collaboration, discussion, and knowledge sharing. To monitor the pulse of the community, track user needs, and build community relationships, Unidata will:

Provide expertise and resources to researchers in designing and implementing effective data management plans

Unidata plans a variety of new activities aimed at supporting members in creating data management plans for proposals (*innovating*). Planned activities include:

- Provide information on data management best practices and information on NSF requirements.
- Create template data management plans for different types of projects, and provide training on their use.
- Promote the use of Unidata technologies such as TDS and RAMADDA at the PI's institution for dissemination of scientific research data.
- Serve as an information clearinghouse to help PIs publicize availability of their data.
- Investigate partnerships with potential providers of archival data storage for community members.
- Work with UCAR's Data Citation effort to help PIs issue DOIs for their data collections.

Represent the academic community in partnerships with agencies and stakeholders

Unidata works closely with agencies such as NOAA and NASA, as well as with commercial entities, to secure access to data and software of interest to the community (*sustaining*). Planned activities include:

- Collaborate with NCEP and Raytheon to make the AWIPS II package available.
- Collaborate with NOAA and NASA to make GOES-R satellite data available.
- Partner with NCEP to provide access to model data via CONDUIT.
- Collaborate with NOAA to make HRRR and FIM data more broadly available.

Conduct workshops related to current community interests and needs

Unidata assists its community to sponsor and organize workshops that bring community members and UPC staff together to learn from each other and share experiences (*sustaining*). Planned activities include:

- Work with the Unidata Users committee to organize Users workshops in 2015 and 2018.
- Increase the number of regional software training workshops arranged by the UPC, hosted by community members, and taught by UPC developers.

Offer training and support for Unidata products and services

Unidata provides training and support in a variety of modes (*enhancing*). Planned activities include:

- Create a series of short online video demonstrations of Unidata software features.
- Investigate and implement virtual training options to better serve those who cannot travel to workshops at the UPC or other locations.
- Support wider use of AWIPS II/GEMPAK by offering on-line and in-person training.
- Investigate the use of social-networking technologies to leverage knowledge within the Unidata community for solving problems.

- Facilitate and coordinate community use case contributions.
- Respond to questions not handled by our user community and conduct annual training workshops.
- Maintain comprehensive documentation, examples, FAQs, and best practices materials online.

Provide demonstration systems to allow evaluation of Unidata tools and technologies, and assist with deployment of those tools and technologies in the field

As Unidata develops new technologies, we create working installations at the UPC and make them available for the community to test and experiment (*innovating*, *enhancing*). Planned activities include:

- Make it easier to install, configure, update, and administer Unidata server technologies, allowing more member sites to operate data services for the benefit of the larger community.
- Demonstrate an increasing number of Unidata technologies that operate "in the cloud," reducing the need for local processing of large datasets or local administration of software.
- Provide easy access to new server-side technologies, building familiarity with remote processing workflows in the community and enthusiasm among community members for contributing additional server-side tools.

Foster interactions between community members through meetings and other opportunities for collaboration and communication

In addition to organizing conferences and workshops for community members, Unidata works to bring community members into contact at events organized by others (*sustaining*). Planned activities include:

- Coordinate UPC participation in scientific meetings and conferences.
- Inform community members about opportunities to meet with UPC staff and each other.
- Use the Unidata web site, electronic mailing lists, and other electronic channels to disseminate information on data streams, software releases, and program center and other community news.

Present Unidata community perspectives and experiences at scientific meetings, conferences, and other venues

UPC staff play a dual role in the geoscience community, serving both as community representatives and as scientific contributors in their own right (*sustaining*). Planned activities include:

- Participate as community representatives and scientists at annual meetings of the AMS and AGU.
- Engage actively with large-scale, long-term projects like NSF's EarthCube and the OGC's standardization efforts.

Use our community-based governance mechanisms to ensure that Unidata program efforts continue to align with the needs of community members

Unidata relies on advice and governance from representatives of member universities (*sustaining*). Planned activities include:

- Continue regular meetings of the Unidata Policy and Users committees.
- Continue efforts to ensure diverse community representation on committees and working groups.

With close cooperation between the UPC staff and the community, and continuing guidance from our governing committees, we will work toward enabling a gradual transition to the new era in cyber-enabled research and education, in ways that do not disrupt community members' current workflows.

Accomplishing these goals will require careful planning, prioritization, resource allocation, and nimble program management. As it has done throughout its history, the UPC will remain vigilant to the shifting landscape and the evolving community needs, reviewing and refining our plans on a regular basis, and making necessary course corrections along the way.

Setting Priorities

The activities outlined in the Action Plan represent our best judgment, *today*, of the actions that will bring us closer to our stated goals. The rapidly-changing technology environment we work in makes it imperative that we remain nimble and ready to adjust our course as circumstances – and the needs of our community – change. Unidata's governance structure provides a mechanism for setting priorities and brings us ongoing guidance regarding what is most important to our community, and we use this feedback as our touchstone in evaluating and re-evaluating our activities and services over time. Our agile software development methodologies are designed to accommodate, and benefit from, these reprioritizations.

Two core beliefs guide our planning and setting of priorities:

- Our community depends on Unidata to provide reliable and robust data services, tools for analysis and visualization, and user support services. These are our highest priorities.
- Transition to a remote-access "cloud" environment for geoscience data is vital to our community's ability to pursue research and education in the 21st century.

Our open-ended technical plans and interconnected products make traditional project management approaches (Gantt charts, *etc.*) less useful than the agile methods we've adopted. Still, resources are not unlimited. To free up resources for new activities, some legacy software such as the LDM, McIDAS-X, and GEMPAK will not be actively developed, but only maintained and supported, in the future. The following table attempts to delineate activities described in our plan that we deem to be critical to our community, along with those we feel could be delayed if resources are constrained.

Strategic Goal	Activities Critical to the Core Effort	Activities to Delay if Necessary
Enable widespread, efficient access to geoscience data	• Continue management of the IDD, adding new datastreams as needed. • Advance server-side processing in TDS, RAMADDA • Create easier data-sharing mechanisms	• Build new data subsetting and progressive disclosure functionality • Integrate AWIPS II data requirements into Unidata technologies
Develop and provide open-source tools for effective use of geoscience data	• Release AWIPS II to the community • Enhance IDV server-side functionality and large dataset access • Integrate McIDAS-V satellite features into IDV • Build tools for easy data file transformation (Rosetta)	• Add OPeNDAP access, low-bandwidth features to AWIPS II • Add WRF initialization services to TDS • Build tools for regridding/aggregation of model output • Enable customized ensemble analysis and visualization • Modernize IDV's rendering engine
Provide CI leadership in data discovery, access, and use	• Advocate for community standards • Develop netCDF streaming protocol, geospatial coordinate access • Begin implementing Unidata applications as "cloud" services • Improve Python support	• Provide support for DAP4 • Reduce "big data" bottlenecks • Improve TDS speed and configuration process • Build automated metadata harvesting features
Build, support, and advocate for the geoscience community	• Promote diverse participation in governing committees • Provide information on and tools for creating data management plans • Continue collaboration with agency partners to provide software and data	• Form partnerships with data archives • Enhance electronic support offerings • Create instructional materials for online virtual training

Closing Remarks

Data services, software, and committed support are critical components of geosciences cyber-infrastructure that can help scientists address problems of unprecedented complexity, scale, and scope. In this proposal we have presented innovative ideas, new paradigms, and novel techniques to complement and extend Unidata's offerings. Our goal is to empower users so that they can tackle major, heretofore difficult problems.

The goals and the action plan outlined in this proposal are in furtherance of and congruent with the National Science Foundation's strategic plans "Empowering the Nation Through Discovery and Innovation: The National Science Foundation Strategic Plan for Fiscal Years 2011-2016" (2011), "GEO vision" (2009), "Strategic Frameworks for Education & Diversity, Facilities, International Activities, and Data & Informatics in the Geosciences" (2012), and "A Vision and Strategy for Data in Science, Engineering, and Education – Cyberinfrastructure Framework for the 21st Century" (2012). We are building on the foundation Unidata has laid over the years, providing cyberinfrastructure for researchers to address frontier science questions, enabling new discoveries, and building the geoscience community's capability to educate the next generation of scientists. Our commitment goes beyond technology. By creating an environment where community members can work together, we hope not only to lower technological barriers to solving multidisciplinary grand challenge problems, but to develop the profession's human capacity and transform the conduct of science.

We acknowledge that this is an ambitious plan with many inter-related goals, but we believe bold thinking is required to address the emergent scientific, educational, and cyberinfrastructure challenges facing the community. As a cornerstone geoscience CI facility with an established record in creating ground-breaking software and services that are in use far beyond the intended audience in academia, the geoscience community expects Unidata to both *lead* users by providing innovative solutions as well as *follow* the community's direction in setting priorities and being responsive to its current and emerging needs. Yet even as we strive to engage these broad challenges, the Unidata program remains firmly committed to meeting its responsibilities to and addressing the evolving needs of its core atmospheric sciences community. Sustained and strong engagement by our community, close partnerships and collaboration with geoscience data providers, tool developers, and other stakeholders, and informed guidance from our governing committees will all be important catalysts for Unidata's success.

Our community's desire for revolutionary ways of wringing knowledge from an ever-expanding pool of Earth System science data presents Unidata with multiple, quickly moving targets. At the same time, the reality of constrained resources means we must choose the problems we will tackle with care and prudence. To succeed in dramatically improving the way data-centric geoscience is conducted will require an approach that is flexible in the face of evolving technologies and shifting priorities. As a result, the underlying theme of our long-term planning is to remain nimble. We must parlay creative, out-of-the box thinking and ongoing collaboration with the community we serve into pragmatic projects that solve today's scientific problems while setting the stage for future advancements.

Achieving these goals will help our community realize the vision of *geoscience at the speed of thought*. This simple statement asks the UPC, in partnership with the community, to work toward a transformation in the practice of data-intensive research and education in the geosciences, enabling researchers and educators to carry out their work in more innovative, efficient, and productive ways, pushing beyond the boundaries of their current knowledge and approaches. In the process, we envision a future that dramatically reverses today's situation in which a researcher may spend 80 percent of their time dealing with data discovery, access, and processing, and only 20 percent "doing science" by way of interpretation, synthesis, and knowledge creation (Michener, 2012). Unidata is firmly committed to work towards realizing this transformation.