Unidata 2020: Geoscience at the Speed of Thought

December 2011

Unidata Program Center
UCAR Office of Programs
P.O. Box 3000
Boulder, CO 80307-3000
www.unidata.ucar.edu

Mohan Ramamurthy, Director

Introduction

Data are the lifeblood of the geosciences. Rapid advances in computing, communications, and observational technologies — along with ensemble and coupled-systems approaches to numerical modeling — are revolutionizing nearly every aspect of the geosciences. The result is a dramatic proliferation of data from diverse sources; data that are consumed by an ever broadening community of users and which are becoming the principal engine for driving scientific advances.

Modern data volumes are staggering. The Phase 5 Coupled Model Intercomparison Project (CMIP5) *alone* will generate more than ten petabytes of climate projection data for use in assessments of climate change, including the Intergovernmental Panel on Climate Change Fifth Assessment Report, scheduled for publication in 2014. For researchers and educators, this deluge of data brings challenges along with the opportunities for discovery and scientific breakthroughs. Retrieving relevant data in a usable format from such an archive should not be more time consuming and arduous than the scientific analysis and investigation the data make possible.

Unidata strives to ensure that our community views the ever-expanding volume of data as an asset rather than an impediment. For more than a quarter century, we have worked in concert with the atmospheric science education and research community to develop and share techniques and resources that remove roadblocks to data discovery, access, and analysis. Unidata's active support of new technologies and the community that uses them has served as a focal point for community-driven change. Some examples:

- Innovations like the Local Data Manager and the Internet Data Distribution system have fundamentally transformed the way our university partners gain access to real-time atmospheric science data.

- Data formats like netCDF, together with community-based data standards like the Climate and Forecast metadata convention and the Common Data Model are enhancing the widespread usability and interoperability of scientific datasets.

- Portable and easy-to-use data analysis tools like the Integrated Data Viewer are bringing robust visualization abilities to researchers, educators, and students in freely-available packages.

Unidata has been remarkably successful in aiding our core atmospheric science community's transition to data-centric scientific workflows. But this success — along with the rapid expansion of data collection and the rise of ubiquitous high-speed networks — brings new challenges and increased expectations from our community. This plan attempts to chart a course for the next five to seven years, outlining concrete goals to further enhance our community's ability to create knowledge from geoscience data while recognizing that the scientific data landscape will continue to change around us.

Our Vision for the Future

Unidata and its community operate in an environment of rapid technological change and evolving scientific priorities. Sources of geoscience data are multiplying and data volumes are expanding, straining traditional data management techniques. Where in Unidata's early days our community was primarily concerned with university access to specific, well-defined weather-related data sets, today's Earth System scientists want the ability to expediently access and integrate any multidisciplinary data set — anywhere, at any time — that advances their scientific or educational goals.

Researchers and educators are looking to Unidata for help with all aspects of the scientific data lifecycle, from locating useful data to retrieving full or partial datasets, through the process of visualizing data locally or remotely, to managing and sharing their results. Working together with our community, we hope to create an environment in which scientists expend less effort locating, acquiring, and processing data and more time interpreting their data. Unidata envisions sustaining and enhancing a community that capitalizes on new technology and concepts to advance the understanding of the complexities of the Earth System and provide community leadership in advancing research and education in the geosciences.

To that end, we will continue to build infrastructure that makes it easy to integrate and use data from disparate geoscience disciplines in concert, allowing investigators to perceive connections that today are obscured by incompatible formats or simple lack of knowledge that the data they need for their investigations indeed exist. Our overarching goal is to work toward a scientific ecosystem in which "data friction"[1] is reduced, and data transparency and ease-of-use are significantly increased. In essence, Unidata's vision calls for creating a scientific cyberinfrastructure environment that allows researchers to conduct

Geoscience at the speed of thought through accelerated data discovery, access, analysis, and visualization

While the term "science at the speed of thought" was introduced by Devaney *et al.* (2007)[2] in the context of immersive software environment for data mining and visualization, we construe the idea more broadly, as a general process to remove barriers to scientific discovery.

Our Program's Mission

The Unidata program exists to serve a community of researchers and educators dedicated to advancing the frontiers of Earth System science. While we share a set of long-term goals with our community, we are keenly aware that we play a significant but supporting role in the ongoing scientific and educational endeavors. As a practical matter, we look for things we can do *now* to help build the future our community seeks to achieve, realizing that the goals will evolve and our approach must be flexible. We also aim to sustain and enhance a community that capitalizes on new

technology and concepts to advance our understanding of the Earth System, providing community leadership and support. With these things in mind, Unidata's mission is

To transform the geosciences community, research, and education by providing innovative data services and tools

Our Strategic and Performance Goals

Unidata serves diverse geoscience education and research communities. Our goals in support of this mission are to:

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, we will:

- Distribute atmospheric and other geoscience data in real time
- Develop innovative cyberinfrastructure solutions to facilitate dissemination of scientific data
- Work with data providers to make geoscience data freely available for advancing research and education
- Develop and maintain the computing and networking infrastructure necessary to keep the growing volume of data flowing reliably and in a timely manner

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 community and collaborative efforts, we will foster an open-source environment that encourages collaborative software development. In this context, and in cooperation with community members and other partners, we will develop and support open-source development approaches and software solutions to:

- Analyze, integrate, and visualize heterogeneous geoscience data in two, three, and four dimensions
- Enable visualization and effective use of very large data sets
- Access, manage, and share collections of data from diverse sources

Provide cyberinfrastructure leadership in data discovery, access, and use

The tools and techniques of distributed scientific computing are continually evolving. Unidata provides information and leadership that allow community members to better anticipate, react to, or influence new developments. 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
- Advance geoscience data and metadata standards and conventions
- Facilitate data discovery mechanisms for quickly finding and accessing geoscience data
- Evaluate emerging cyberinfrastructure trends and technologies, providing information and guidance to community members

Build, support, and advocate for the diverse 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, we will:

- Provide expertise and resources to researchers in designing and implementing effective data management plans
- Represent the academic community in partnerships with agencies and other stakeholders
- Conduct workshops related to current community interests and needs
- Offer training and support for Unidata products and services
- Provide reference implementations and demonstration systems to allow evaluation of Unidata tools and technologies, and assist with deployment of those tools and technologies in the field
- Foster interactions between community members through meetings and other opportunities for collaboration and communication
- Present Unidata community perspectives and experiences at scientific meetings, conferences, and other venues
- Use our community-based governance mechanisms to ensure that Unidata program efforts continue to align with the needs of community members

The vision and goals outlined in this plan are entirely congruent with the National Science Foundation's strategic plans "Empowering the Nation Through Discovery and

Innovation" and "GEO vision." We believe achieving the goals defined here will help Unidata and its community realize the vision of *geoscience at the speed of thought*. This simple statement asks us to work toward a transformation in the conduct of data-centric 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 his or her time dealing with data discovery, access, and processing, and only 20 percent "doing science" by way of interpretation, synthesis, and knowledge creation.

We acknowledge that this is an ambitious plan with many inter-related goals, but we believe this kind of systemic thinking is required in order to tackle the scientific, educational, and cyberinfrastructure challenges facing the geosciences community. Yet even as we strive to address 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. Achieving the goals we are setting ourselves will require careful planning, priority-setting, and allocation of resources. Sustained and strong engagement by our community, close partnerships and collaboration with geoscience data providers, tool developers, and other stakeholders, and the informed guidance of our governing committees will all be important catalysts for Unidata's success.

As we implement the strategic goals described above, it will be important to formulate a clear action plan with associated performance goals and measures, monitoring them closely to gauge our progress. In addition to the traditional metrics for output and outcomes, it will also require new approaches to evaluation and assessment of the impact of the tools, technologies, and services provided by Unidata on STEM education and scientific innovation.

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 changing technologies and shifting priorities. As a result, the underlying theme of our long-term planning must be 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 concrete scientific problems today while setting the stage for future advancements both evolutionary and revolutionary.

---

[1] Edwards, Paul N. (2010), *A VAST MACHINE: Computer Models, Climate Data, and the Politics of Global Warming*, 518 pp., The MIT Press, Boston.

[2] Devaney, J.E., Satterfield, S.G., Hagedorn, J.G., Kelso, J.T., Peskin, A.P., George, W.L., Griffin, T.J., Hung, H.K., and Kriz, R.D. (2004), Science at the Speed of Thought, *Proceedings of Ambient Intelligence for Scientific Discovery*, 1-24.