unidata

Strategic Vision 20/20

What is unidata?

Unidata is a diverse community of education and research institutions with the common goal of sharing geoscience data and the tools to access and visualize that data. For more than 30 years, Unidata has been providing data, software tools, and support to enhance Earth System education and research. Funded primarily by the National Science Foundation (NSF), Unidata is one of the University Corporation for Atmospheric Research (UCAR)'s Community Programs (UCP).

The Unidata program was first conceived in 1983, when the NSF called together representatives from some 80 U.S. university meteorology and atmospheric science departments to conceptualize ways of making "real-time" weather observations and other data available to the university community. As a result of that meeting, the Unidata program was established to facilitate the transmission of weather data, imagery, and forecasts to the university community; to support universities' local interactive computing capabilities; and to improve data access capabilities for researchers and educators.

Since its founding, Unidata has not only succeeded in its original mission to ensure access to geoscience data, but has built a vibrant community of researchers and educators who have communicated their evolving needs so that the program can be of even greater service. As a result of Unidata's robust community governance structure, our long-term efforts to develop and support scientific software for data access, manipulation, analysis, and visualization have enjoyed both support and guidance from those we serve.

Unidata's active support of new technologies helps our community navigate the quickly-changing technological landscape to find tools that increase productivity and reduce the time needed to achieve results. For example,

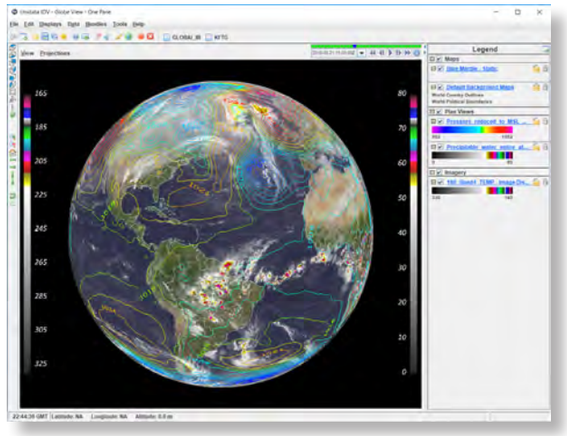
- The Local Data Manager and the Internet Data Distribution system fundamentally transformed the way our university partners gained access to real-time geoscience data. Now, Unidata cloud computing and remote data access technologies are enabling access to new observational and model datasets too voluminous to be "pushed" over the network.
- 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 across geoscience domains.
- An array of robust software tools like the Integrated Data Viewer (IDV), the National Weather Service's Advanced Weather Interactive Processing System (AWIPS) package, and the MetPy library allow users to choose data visualization and analysis tools that fit their scientific workflows.

What are our community's Needs?

Mid-way through its fourth decade of providing geoscience data and tools to the university community, Unidata continues to evaluate and respond to changes in the technological environment in which our community members operate. But in addition to staying abreast of technical advances, we realize that changes in the scientific and educational landscape also have important implications for the researchers and educators who rely on Unidata's software and services. This is why Unidata has focused its resources on, for example, both cloud-computing technologies and helping members navigate the changing data management requirements of funding agencies. Similarly, we understand that the challenges presented by a world of "big data" go beyond the technical issues of finding and using ever-expanding volumes of observational data and model output to encompass problems relating to how the datasets are described, stored, accessed, analyzed, integrated, cited, and shared with other scientists and the public.

Recognizing widespread public interest in the openness and accessibility of the scientific process, the National Science Foundation's Strategic Plan for 2014 – 2018, *Investing in Science, Engineering, and Education for the Nation's Future*, makes a commitment to "the principles underlying open government, including transparency, public participation, and collaboration with other government agencies and private institutions" (NSF 2014a). Unidata hopes to further these NSF goals by focusing on facilitating scientific workflows that combine software tools and data management practices that promote ease of use, reduce "time to science," and enhance both the transparency of the scientific process and the reproducibility of scientific results.

A recent National Academy of Sciences report, *Envisioning the Data Science Discipline: The Undergraduate Perspective*, calls attention to the importance of building a workforce that understands how to work effectively in the modern data environment. "Scientists, engineers, and executives routinely encounter enormous volumes of data," write the authors, who note that "The nation's ability to make use of these data depends on the availability of an educated workforce with necessary expertise." At Unidata, we couldn't agree more with the panel's assertion that "Using real data will expose students to the messiness they will confront when solving real-world problems. Selecting applications with broad impact will make instruction more compelling, helping to attract and retain students" (NAS 2017). These sentiments validate Unidata's activities aimed at bringing geoscience data and modern tools to the university community.



Similarly, NSF's *Dynamic Earth: GEO Imperatives & Frontiers 2015-2020* prioritizes developing "community-driven cyberinfrastructure to advance data-enabled science and education" (NSF 2014b) Unidata works closely with its community to build cyberinfrastructure that makes it easy to integrate and use heterogeneous data from disparate geoscience disciplines in concert to advance transformative and data-intensive disciplinary and multidisciplinary studies. Going forward, we also hope to continue contributing to NSF's Dynamic Earth vision of a "coherent, distributed framework for the open and easy discovery of, and access to, data; software and services; information; and computational resources." Unidata is working to make data- and user-centered, end-to-end workflows more easily accessible, designing for re-use and modification. By participating in NSF's cyberinfrastructure initiatives like EarthCube and leveraging advances in web-based analysis tools, the abstraction and encapsulation enabled by contemporary virtualization techniques, and emerging ideas like the FAIR Data Principles (Findability, Accessibility, Interoperability, and Reusability) (Wilkinson 2016), we hope to help make the scientific process more open and available to all geoscientists. Our overarching goal is to work toward a scientific ecosystem in which "data friction" (Devaney 2004) is reduced, and at the same time data, software, and workflow transparency and ease-of-sharing are significantly increased.

What is Unidata's Role?

Our Vision

Our overarching goal is to work toward a scientific ecosystem in which "time to science" is shrunk by reducing "data friction," 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

Our 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 their 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 approaches 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 Contributions to the Community

While we see many opportunities to aid the geoscience community by reducing the "time to science," we see our contributions in three broad categories: Data, Tools, and People.

Managing Geoscience Data

Unidata has decades of experience building cyberinfrastructure that makes it easy to integrate and use data from disparate disciplines in concert. Our aim is to help investigators perceive connections that may otherwise be obscured by difficulty in finding, accessing, or working with geoscience data.

Providing Useful Tools

As researchers and educators depend ever more heavily on software tools to pursue their scientific goals, Unidata's aim is to provide cyberinfrastructure that is easy to use, freely available, and build around accepted community data standards.

Supporting People

People — individuals and groups — make science happen. Our focus is on supporting people who pursue scientific inquiry not only with tools and data, but with technical support, training, and community-building activities and opportunities for collaboration and sharing of ideas and expertise that help unlock the potential of new technologies and data sources.

Managing Geoscience Data

Data volumes have expanded tremendously in the past decade. For example, the volume of data emanating from each GOES-R class satellite will be roughly 20 times the volume from the older GOES-IP satellites. Numerical Weather Prediction models have increased their resolution in both spatial and temporal domains, with resulting increases in the volume of their output. Unidata has worked to help its university community members, who have been integral participants in the push-based Internet Data Distribution network, to adapt to rising data volumes by fine-tuning the data sent to them and by switching to the use of pull-based remote access workflows based on the OPeNDAP and other remote data access protocols. Unidata's efforts to enable cloud-based, data-proximate scientific workflows continue this trend, helping community members access the data they need.

Making Geoscience Data Accessible

Unidata will work to make geoscience data more accessible by:

- Promoting remote access methods that make it possible to use the relevant portions of a dataset without the need to store the entire dataset locally
- Improving the reliability/scalability of remote access servers to better handle terabyte- and petabyte-scale datasets, allowing access by a wider range of users

Making Geoscience Data Discoverable

Unidata will work to make existing geoscience datasets easier to find by:

- Fostering community adoption of Data Service Conventions, such as a common OPeNDAP URL format, to make data locations easier to discern and use
- Working towards creation of a Data Services Discovery Protocol that will help standardize the way data users locate and acquire data appropriate for their projects
- Investigating mechanisms that allow the metadata associated with geoscience data to be exposed via common search engines and other widely used tools

Making Geoscience Data Usable

Unidata will work to make geoscience data more usable by:

- Promoting the use of Data Standards such as netCDF, HDF, and the Climate and Forecast metadata conventions (CF) that allow scientists to quickly understand the shape and provenance of datasets
- Improving data-proximate workflows, in which processing and analysis are done close to the data itself, rather than moving data close to the processing tools
- Educating community members in the effective use of these standards and workflows

Enhancing Community Access to Data

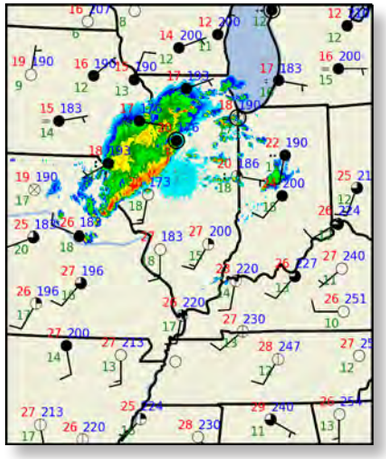
Unidata's ongoing efforts to ensure access to important geoscience datasets will continue, with special emphasis on:

- Collaborations with government science agencies to gain community access to datasets like those generated by the GOES-R series satellites
- Collaborations with the commercial sector to secure access to datasets available through public-private partnerships like that between NOAA and Amazon Web Services
- Advocating with stakeholders such as scientific societies and universities for clear community standards around data sharing and access

Providing Useful Tools

To pursue meaningful scientific inquiry using the ever-expanding volume of available data, researchers and educators need usable and freely available software to access, analyze, manage, and visualize data. Evidence indicates that the most robust solutions come about through "agile" software development processes — that is, incremental and iterative evolution involving feedback from an engaged user community. Unidata has long served as such a center of collaboration, building open source and freely available software tools in a process guided by input from researchers and educators who use Unidata tools daily.

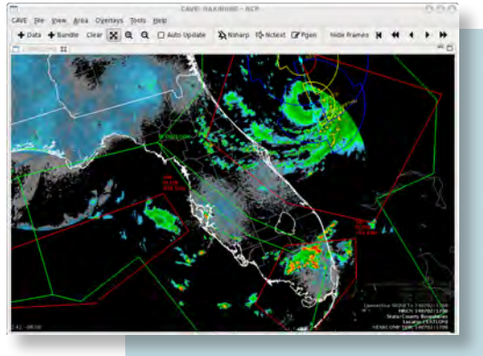
One refrain we hear from our community is *make it simpler*. Unidata has focused on building tools that let practitioners spend less time on tasks like searching for data or managing software and more time studying data and investigating the information it contains. Our efforts to bring robust and standardized tools for working with geoscience data to the Python ecosystem, to reconfigure the National Weather Service's AWIPS package for use in the university environment, and to help researchers design smooth and useful data management workflows that promote collaboration and reproducibility are all examples of Unidata's efforts aimed at reducing *time to science*.



Promoting 21st Century Scientific Workflows

Unidata will work to broaden community adoption of modern analysis and visualization tools by:

- Enhancing offerings of web-based tools that minimize the need for local hardware and software
- Expanding offerings of software that promotes the use of cloud-computing workflows
- Providing professional-grade and cloud-hosted software tools like AWIPS, IDV, and geoscience-focused Python libraries that leverage modern cyberinfrastructure such as Jupyter notebooks



Creating Modern Data Management Workflows

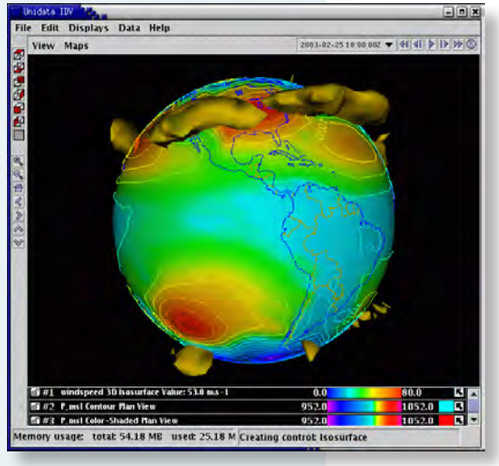
Unidata will work to foster adoption of modern data workflow tools by:

- Continuing to enhance data server technologies such as OPeNDAP and the THREDDS Data Server that facilitate remote data access
- Building tools and workflows that are geared toward data-proximate analysis and visualization tasks, reducing the need to transport large datasets between storage and analysis locations
- Building tools and workflows that promote long-term reproducibility of scientific analysis, by making it easier for others to replicate the analysis both now and in the future

Supporting Legacy Workflows

Unidata will continue to support existing scientific workflows by:

- Providing ongoing technical support for existing packages and processes that are currently in heavy use in the geoscience community
- Working to recast legacy workflows to use cloud-computing resources
- Building workflows for archiving software to enhance long-term reproducibility of scientific results



Supporting People

Since its inception, Unidata has focused on supporting community members in their scientific endeavors. Our training and technical support activities aim to ensure that researchers and educators can build the technical skills needed to advance their data-intensive scientific inquiries. By working to expand the Unidata community, we hope to increase access to the tools and techniques we feel are enabling scientific discovery. Community-based standards play a large role in this endeavor; experience shows that processes that arise from practicing scientists doing real work are most likely to benefit a broad spectrum of participants.

Direct participation in Unidata's governance by members of the academic geoscience community grounds the program in the realities of teaching and doing science. In turn, this high level of involvement by researchers and educators allows Unidata to advocate effectively with science agencies, standards organizations, and other entities to benefit the broad community.

Providing Technical Support

Direct interaction between users and software developers at the UPC benefits all phases of the development process. Over more than 30 years, community members have come to rely on the expert assistance of UPC staff. Unidata is committed to maintaining the highest level of technical support through a variety of forums including:

- Speedy replies to e-mail support questions
- Community e-mail lists
- Open Source development forums like GitHub and StackExchange
- One-on-one interactions and in-person visits to community sites

Building Community Capacity with Cloud Workflows

Community members have indicated both a growing interest in using cloud-based workflows and concern about the best ways to benefit from the new technology. Unidata hopes to guide the community by:

- Providing demonstration workflows that leverage data-proximate tools
- Providing training on how to configure and use cloud-based resources

Unidata also serves in the role of convener, bringing stakeholders at many levels together to discuss data-related issues, community needs, and large-scale trends in the cyberinfrastructure landscape. Using venues like scientific meetings and data-focused conferences, we seek to bring people together to address issues faced by all.

Building Capacity through Open Source Software Development

Software developed by the UPC has always been freely available. Unidata aims to increase community participation in this development by:

- Encouraging community involvement in software development, testing, and documentation through mechanisms like Github
- Participating in Free and Open Source Software (FOSS) initiatives and advocacy efforts

Building Community Cyber-Literacy

Unidata's long-standing software training and cyber-literacy programs continue to evolve. For the benefit of our user community, and with an eye toward the development of the next generation of geoscientists, we continue to:

- Provide software training workshops at the UPC and at member institutions
- Enhance our distance learning and online training materials
- Host meetings and workshops to discuss and study cyberinfrastructure topics

Unidata's Vision for the Future

Transdisciplinary data synthesis, facilitated through increased interoperability of heterogeneous, multidisciplinary, and multiscale datasets, is essential to solving grand challenge environmental science problems like climate change and prediction of natural hazards and extreme events. Unidata's vision is to serve as a catalyst to accelerate geoscientists' ability to characterize, understand, and predict the complex Earth System by enabling data-centric scientific workflows.

Unidata envisions a future in which researchers spend less time finding, accessing, and processing data, allowing them to focus on analysis and interpretation — and publication of their results. As we continue to refine our data services, we look to enable ever-greater use of emerging data collections across scientific disciplines. We hope that our ongoing efforts to promote open science by educating our community on the use and value of shared standards for data discovery and interoperability will bear fruit in the form of well-documented and broadly shared datasets that can be explored and reused in a range of Earth System Science contexts.

We look to emerging web-based scientific workflows to make data-intensive science both easier and more reproducible. We will continue to showcase techniques that combine robust access to well-documented datasets with easy-to-use tools that are available across geoscience disciplines. But we will also work to foster adoption of new technologies like Jupyter notebooks and other data-proximate computational and analytic methods enabled through "Software as a Service" techniques. In this way, Unidata hopes to transform the conduct of geoscience research and education.

Unidata's systematic strategies and concrete efforts aim to advance the data, cyber, and geospatial literacy of students who comprise the next generation workforce. By enabling authentic data-enabled inquiries across the geoscience domains, and training on contemporary methods in data analysis and visualization, we hope to build capacity in the geoscience departments of our universities and colleges, helping them to produce a highly-skilled workforce with well-honed abilities to conduct data-intensive basic and applied geoscience.

The vision outlined here and the goals set for the upcoming years put Unidata in alignment with the National Science Foundation's vision of

> *A Nation that creates and exploits new concepts in science and engineering and provides global leadership in research and education*

as expressed in the NSF's Strategic Plan. We are confident that accomplishing the goals Unidata is setting for itself will result in a geoscience community that is better able to use and learn from the growing volume and increasing complexity of Earth System data becoming available. We hope that our efforts will result in wider adoption of workflows that promote transparency, reusability, and long-term stability of scientific knowledge.

References

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.

National Academies of Sciences, Engineering, and Medicine. 2017. *Envisioning the Data Science Discipline: The Undergraduate Perspective: Interim Report*. Washington, DC: The National Academies Press. DOI: 10.17226/24886.

National Science Foundation, 2014 (a): *Investing in Science, Engineering, and Education for the Nation's Future.* Accessed 29 March 2017. [Available online at https://www.nsf.gov/publications/pub_summ.jsp?ods_key=nsf14043]

National Science Foundation, 2014 (b): *Dynamic Earth: GEO Imperatives & Frontiers 2015-2020*. Accessed 5 April 2017. [Available online at https://www.nsf.gov/geo/acgeo/geovision/start.jsp]

Wilkinson, M. D. et al., 2016: *The FAIR Guiding Principles for scientific data management and stewardship*. Sci. Data 3:160018 DOI: 10.1038/sdata.2016.18.

Thanks to Our Community

The Unidata Program could not serve its community effectively without ongoing input from a broad spectrum of geoscience researchers and educators. Our governing committees serve as an important interface between the Program Center and this wide community, and we thank them and all who provide ideas and information for their ongoing efforts.

In particular, we appreciate the input provided by members of our Strategic Advisory Committee, who guided the creation of this plan.

Thanks to Our Funders

The Unidata Program is funded primarily by the National Science Foundation, under Award AGS-1344155.

Specific projects undertaken by the Program also receive supplemental funding from numerous U.S. Federal Agencies, including the National Science Foundation, the National Oceanic and Atmospheric Administration, the National Aeronautics and Space Administration, and the United States Geological Survey.

www.unidata.ucar.edu

Unidata is one of the University Corporation for Atmospheric Research's Community Programs, and is funded primarily by the National Science Foundation

Unidata Program Center, November 2017