Unidata Reimagined: New Approaches to Community Data Services

For nearly four decades, the Unidata Program has been a cornerstone data facility for the Earth Systems Sciences (ESS), serving a diverse community of education and research institutions with the common goal of sharing data and tools for scientific analysis. Over this time the program has seen dramatic changes in the technological, scientific, educational, and public policy landscape; these shifts have and are continuing to transform the ways our community members conduct their research and educate new generations of scientists.

The pace of change affecting the ESS community is, if anything, accelerating. The volume of environmental data available to scientists is exploding, with sensors of ever-higher resolution in place on land, in the air and oceans, and in orbit. Numerical prediction models continue to increase both their spatial and temporal resolution, further swelling the quantity of information available to scientists. Helping our community gain access to the data they need to answer questions about the earth has always been at the core of Unidata's mission, but expanding data volumes require new approaches to this undertaking. Where once it was most efficient for scientists to download datasets to their local computers, today's enormous data volumes argue for placing data and computing processes side-by-side in cloud-like environments, transferring only analysis-ready results to researchers' local computers. To keep up, Unidata must offer technology and training that help community members adopt new data workflows.

As the scientific data environment has changed, so have the software tools scientists use to analyze and visualize that data. Legacy tools are often unsuitable for use with cutting edge computational algorithms at scale. Proprietary systems have given way to a robust Open Source scientific software ecosystem. Cloud computing systems allow researchers to scale computational resources quickly and efficiently, but these systems require new tools and workflows suited to remote operation. Modern interpreted programming languages like Python and R lower barriers to creating software tools, democratizing development processes and making it possible for a wider range of participants to contribute to the scientific software ecosystem. And the recent explosion of machine learning (ML) capabilities, as well as the emergence of large language models and generative artificial intelligence (AI) tools like ChatGPT, have already caused seismic shifts in the scientific data landscape, highlighting the need for analysis- and AI-ready datasets. These advances in technology are outpacing changes in education and pedagogy, and the gap keeps increasing — creating significant challenges for educators and learners. Making these tools widely accessible and helping our community adapt to their use in ways that are both efficient and ethical will require both technical skill and thoughtful consideration of societal imperatives.

Changes in societal ideas about the best ways to conduct science are also important drivers for Unidata's evolution. Issues of representation in the ESS community imbue in us a sense of urgency for the task of broadening the reach of Unidata's programs to better serve historically underrepresented groups. To our work toward implementation of FAIR data principles (Findable, Accessible, Interoperable, and Reusable) we add a commitment to the CARE principles for indigenous data governance (Collective benefit, Authority to control, Responsibility, and Ethics). To paraphrase the White House Office of Science and Technology Policy's Year of Open Science announcement, we want Unidata's activities to "improve research infrastructure, broaden research participation for emerging scholars, and expand opportunities for public engagement." The National Science Foundation (NSF) approaches these social challenges by applying the lens of *convergence research*, focusing on addressing complex problems by integrating knowledge, methods, and expertise from across many scientific, engineering, and social science disciplines. To Unidata's

historical strength in bringing disparate types of *data* together, we must add support mechanisms for disparate *communities* joining forces to solve problems.

Studying and preventing future disasters like 2023's wildfires on Maui or catastrophic flooding in Libya will require this type of convergence ESS approach. Diverse groups of scientists, public officials, planners, and other citizens will need robust access to data from the atmosphere, oceans, terrestrial ecosystems, human land use, and myriad other sources. Unidata's goal to extend its impact to new communities of practice through the creation of *community hubs* seeks to bring these elements together, using technology to facilitate the gathering of diverse communities to share knowledge, data, and tools, creating a whole that is more useful than the sum of its parts.

While the Unidata Program has a long history of successfully serving as a cyberinfrastructure facility, evolving to meet the needs of the ESS community, these dramatic changes in the present day technological, scientific, and social landscape require us to reimagine how the program can best fulfill its mission. This proposal provides a description of how Unidata plans to serve its community going forward so that our users continue to be successful in their research and education endeavors.

Who We Are and What We Do

Unidata is a community cyberinfrastructure facility for the Earth Systems 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 hub for activities related to Unidata's mission, is managed by the University Corporation for Atmospheric Research (UCAR), a consortium currently comprising 127 North American colleges and universities providing science in service to society.

Unidata exists to engage and serve researchers and educators who are advancing the frontiers of ESS; we hope to support their efforts by creating opportunities for community members from many backgrounds and disciplines to share data, knowledge, methods, and expertise. As part of this effort, we strive to provide well-integrated data services and tools that address the entire geoscientific 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, some of the UPC's most important activities include:

- Acquiring, distributing, and providing remote access to real-time ESS data.
- Developing software for accessing, managing, analyzing, visualizing, and effectively using a wide variety of ESS data.
- Providing learning experiences, training, and technical support to community members to help them be more productive in their scientific activities.

From the outset, Unidata has been governed by its community. Representatives from universities and colleges populate standing committees that help set policies for the program and provide first-hand feedback from users of program software and services. Non-voting representatives from government agencies 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 priorities as the technological landscape and community needs change.

While Unidata's primary mission of serving universities engaged in ESS education and research has remained unchanged through the years, the evolution and broad usefulness of its products and services have greatly enlarged its user base. Today, the Unidata community includes users from all sectors in over 200 top-level Internet domain countries and territories, including more than 1800 academic institutions.

Unidata's most recent strategic plan, developed in 2023 with the aid of its governing committees and community, leads us to envision Unidata as a community space for gathering data, tools, knowledge, and people to advance ESS. That plan reaffirms Unidata's mission *To transform the Earth System Sciences community, research, and education by providing innovative data services, tools, and expertise*, and identifies four strategic goals to guide our actions:

- **Provide Data and Tools**: ensure fair and equitable access to ESS data as well as tools to analyze and visualize that data.
- **Reduce Barriers to Participation**: help to build a larger, more inclusive community of ESS practitioners.
- **Foster Community Action**: engage community members to work collectively to improve ESS teaching and research.
- **Provide Innovative Technical Solutions**: guide the ESS community toward innovative technical solutions.

These objectives and our plans to achieve them will help us focus our efforts in areas we believe are most vital to our community today and in the coming years; they underlie everything you'll read in this proposal. But our plans for the future are rooted in Unidata's history of community engagement and service as well as technical innovation, building on our strengths — we are only in a position to be successful in our planned endeavors thanks to the solid foundation we have created. This section briefly describes the interlocking set of activities that form the core of Unidata's work; the Plan of Action that follows describes how we will adjust and augment these activities to address Unidata's strategic goals for the empowerment of its broad community.

Goal: Provide Data and Tools

One of Unidata's first mandates, and a historical focus of our activities, involves ensuring that ESS researchers, educators, and students have access to the data they need. For many years, the UPC has coordinated the Internet Data Distribution system (IDD), in which hundreds of universities cooperate to disseminate near real-time Earth observations via the Internet. This ongoing effort highlights Unidata's partnerships with the National Weather Service (NWS) and other data providers to gain access to the observations and model output needed by community members to advance their science and teaching. The data "pushed" to member sites via the IDD include weather radar, satellite imagery, forecast model output, and others — more than 30 distinct data feed types in all. As data volumes have increased, and use of the IDD mechanism by community members has risen to record levels, Unidata has worked to lead the community towards use of remote data access methods. Our goal is to make it easy for community members to shift their workflows towards remote data access and data-proximate analysis and visualization technologies, taking advantage of cloud-based resources and reducing the need to transfer large volumes of data. Unidata's long-term engagement with the National Oceanic and Atmospheric Administration (NOAA), the National Aeronautics and Space Administration (NASA), the World Meteorological Organization (WMO), and other agencies as they develop their own cloud-based data access and use infrastructures has helped ensure that our community members have the tools they need to work in these new environments.

Unidata develops, distributes, and supports a suite of freely available software tools for managing, analyzing, and visualizing ESS data.

- NetCDF (network Common Data Form), a collection of data access libraries that provide a machine-independent data format that is self-describing, portable, scalable, appendable, shareable, and archivable. NetCDF has been adopted widely by the ESS community, and is especially popular among climate and ocean modelers. For example, model output datasets for the Sixth Assessment Report of the Intergovernmental Panel on Climate Change must be

submitted in netCDF format, using the associated Climate and Forecast (CF) metadata conventions. Data from the latest generation GOES series satellites is also delivered in netCDF format. 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 ArcGIS, MATLAB, and IDL.

- MetPy and Siphon libraries for working with ESS data within the scientific Python ecosystem have also gained widespread acceptance; more than 45 open source packages depend on the MetPy project, and some 250 theses and peer-reviewed publications mention MetPy and Siphon. Unidata's MetPy developers continue to engage with the Pangeo project, a grass-roots effort to develop a community stack of tools serving the atmospheric, oceanic, land, and climate science.
- The Advanced Weather Interactive Processing System (AWIPS) is a weather forecasting, display, and analysis package developed by the NWS and NCEP for use in Weather Forecast Offices (WFOs) and National Centers for Environmental Information. Because many university meteorology programs are eager to use the same tools used by NWS forecasters, Unidata community interest in AWIPS is high. As designed for the WFOs, however, AWIPS is a very resource-intensive system; installing and maintaining the NWS version is beyond the capabilities of many universities. As a result, Unidata has reconfigured and repackaged AWIPS for non-operational use, allowing universities access to the system for teaching and research with significantly reduced equipment requirements. By operating AWIPS Environmental Data EXchange (EDEX) servers in the NSF Jetstream cloud on behalf of the university community, Unidata has lowered barriers to educational use of the system even further.
- Unidata's THREDDS Data Server (TDS) allows for remote access to collections of scientific data via electronic networks. Data published on a TDS are accessible through a variety of remote data access protocols including OPeNDAP, DAP4, OGC Web Map Service (WMS) and Web Coverage Service (WCS), NetCDF Subset Service (NCSS), and HTTP. 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 technology that underlies the TDS: Unidata's Common Data Model (CDM). NOAA's ERDDAP, NASA's Panoply, and CMAS' VERDI are examples.

These software tools, along with other Unidata-developed and supported packages including the Integrated Data Viewer (IDV) and Local Data Manager (LDM), form a robust platform for creating, accessing, and using ESS data.

Goal: Reduce Barriers to Participation

Unidata has a variety of programs and activities whose goal is to make sure Program offerings, services, and expertise are available to communities who can benefit from them. We actively pursue opportunities to interact with students and educators from minority-serving institutions (MSIs); recent collaborations have included Historically Black Colleges and Universities (HBCUs) like Howard University and Morgan State University and Tribal Colleges and Universities (TCUs) such as Navajo Technical University, Southwestern Indian Polytechnic Institute, and Haskell Indian Nations University. We participate in the NCAR Rising Voices program to foster collaborations of Indigenous and Earth (atmospheric, social, biological, ecological) sciences, and are working to expand our connection with other underserved communities.

The Program also strives to alleviate resource barriers for community members. Spurred by COVID-19 travel restrictions, we have focused efforts to provide training without the need to travel, improving our hybrid and fully-remote interaction capabilities and working to build asynchronous learning resources. The Unidata Science Gateway makes robust cloud-computing resources available

to community members at no cost to them, lowering their costs and IT infrastructure needs. And through our Community Equipment Award Grants program, we seek to make it easier for institutions to participate in our community mission to share data, tools, and techniques.

Goal: Foster Community Action

Partnering with U.S. tribal institutions, Unidata has worked to promote indigenous data sovereignty both through education and on-the-ground projects including the NSF-funded *Sovereign Network System for Environmental Monitoring, Data and Information Exchange, and Collaboration* (NSF 2021). Through interaction with project participants and other indigenous and non-indigenous groups, Unidata has expanded its own approach to the idea of Open data to include the CARE principles for Indigenous Data Governance: Collective benefit, Authority to control, Responsibility, and Ethics. We look for ways that Unidata technologies and programs can support communities in their quest to share data while protecting their own cultural interests.

In addition, Unidata undertakes a wide variety of activities aimed at fostering a shared vision for and community ownership of the Unidata Program itself, 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. Thanks to these strong community partnerships, the Program is well positioned to represent community interests in collaborations with standards bodies and organizations that work to foster Open and Convergence ESS approaches. As standards-based solutions have become increasingly important to the conduct of international science and interoperability, 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.

Goal: Provide Innovative Technical Solutions

In addition to creating and supporting widely used software packages, Unidata development teams constantly search for ways to expand community members' ability to conduct their research and educational activities. When the COVID-19 pandemic forced universities to limit in-person classroom learning, Unidata was able to offer remotely accessible resources through the Unidata Science Gateway. As part of the Sovereign Network project mentioned above, UPC Science Gateway development staff have been working closely with the project partners to build capacity for environmental modeling for Tribal Nations. Project partners can now run the Weather Research and Forecasting (WRF) model over tribal lands on the NSF Jetstream2 cloud through the use of a containerized version of WRF.

Unidata continues to focus on ways that modern software practices can benefit ESS practitioners. Unidata software developers work closely with groups like the Zarr community to provide cloud-computing ready access to large N-dimensional typed arrays, or Project Pythia to provide educational resources for the ESS community. Program center staff are at work enhancing our software packages to better incorporate AI/ML workflows, to present ESS datasets in ways that are amenable to efficient use in ML models, and to build skill with AI/ML techniques among ESS students, educators, and researchers.

Unidata Snapshot

The following tables provide a snapshot of the Unidata program in October 2023.

Table 1: Data Services	2023
Host machines on the IDD network[1]	533 (190 unique sites)
Approximate volume of data ingested into the IDD	1.75 TB/day
Volume of data pushed to IDD sites	90 TB/day
Volume of data pulled via remote access protocols	2.3 TB/Day
Uptime of UPC data and support infrastructure	99.98%

Table 2: Software Package Downloads	2018-2023
AWIPS	40,417
IDV	35,506
Local Data Manager	10,495
McIDAS	263
MetPy[2]	965,316
netCDF-C Libraries (includes FORTRAN, C++ support)[3]	919,337
netCDF-Java Libraries (Common Data Model)	36,246
Siphon[2]	379,659
THREDDS Data Server	13,875
UDUNITS	56,038

Table 3: Workshops	2018-2023
Training Workshop Participants (in person, virtual, and hybrid)	1124
Users Workshop Participants	91

Table 4: Miscellaneous	2023
Number of countries and territories where Unidata software and services are used	222
Number of academic institutions participating (U.S.)	481
Number of academic institutions participating (worldwide)	1,883
Number of citations of Unidata software and data services, 2018-2022	17,422
Number of EPSCoR institutions participating	117
Number of HBCU institutions participating	9
Number of HACU institutions participating	66
Number of TCU institutions participating	8
Number of organizations participating worldwide	2,382
Number of summer interns since program inception (2013)	26
User support e-mail transactions, 2018-2023	26,883
Number of Community Equipment Awards, 2018-2023	17
Average staff FTEs at the UPC, 2018-2023	27

[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] MetPy and Siphon are downloaded via Python package managers rather than UPC download infrastructure.

[3] 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.

Results of Prior Support Under NSF 1901712

A great majority of activities at the UPC are sponsored under the five-year core-funding award "Unidata: Next-generation Data Services and Workflows to Advance Geoscience Research and Education" (NSF 1901712). The period of performance for that award is 1 May 2019 to 30 April 2024.

The proposal for that funding award grouped the Unidata program's activities into the following focus areas identified in Unidata's 2017 Strategic Plan:

1. Managing Geoscience Data
2. Providing Useful Tools
3. Supporting People

During the period of performance for NSF 1901712, Unidata has met many of the objectives set forth in that proposal, while also pivoting quickly to help the community meet the challenges of the COVID-19 pandemic and the associated needs for remote teaching and learning capabilities. Unidata has increased the volume and variety of data made available to the community; made its software offerings more robust and accessible; enhanced its online, hybrid, and in-person learning offerings; and taken steps to ensure its ability to provide guidance and technological leadership in emerging fields like Artificial Intelligence and Machine Learning. The program has worked to build connections with historically underrepresented communities including Tribal Colleges and Universities and other MSIs across the United States, empowering faculty, students, and researchers to be more productive and enhancing their ability to advance science and learning.

Key Achievements 2018-2023

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

- **Data Delivery**. The volume of observational data and model output delivered to Unidata community members and institutions in near real-time continues to grow. As of mid-2023, Unidata's IDD clusters deliver roughly 90 Terabytes per day to downstream systems, a significant increase from the roughly 35 Terabytes per day at the beginning of the current five-year period. The volume of data served via remote access methods (TDS, ADDE, and standard web servers) now averages approximately 2.3 Terabytes per day.
- **Unidata Science Gateway**. Leveraging the strengths of the cloud computing environment to enhance universities' access to ESS data and tools has been one of Unidata's highest priorities during the period of this award. Efforts in this area have been concentrated in the Unidata Science Gateway program, which leverages resources made available by NSF's Jetstream and Jetstream 2 projects to make data and Unidata software packages more readily available to the university community.
  The Unidata Science Gateway was a key element of Unidata's response to the COVID-19 pandemic. When universities reacted to restrictions on in-person instruction by moving classes online, Unidata offered to set up custom ESS-focused cloud computing environments for courses and workshops. Instructors are able to request Science Gateway resources configured as a JupyterHub server, with access to software environments and data streams pre-configured. To date, this Science Gateway academic support program has enabled courses and workshops at 22 different universities, accommodating more than 1200 students.
- **Outreach to Underserved Communities**. During the award period, Unidata has been collaborating with the Southwestern Indian Polytechnic Institute (SIPI) and Navajo Technical

University (NTU) on NSF grant 2131301, "A Sovereign Network System for Environmental Monitoring, Data and Information Exchange, and Collaboration among Tribal Colleges and Universities." Through this pilot project, the collaborators have installed environmental monitoring equipment, stood up systems for data storage and sharing, and built relationships with other Tribal Colleges and Universities (TCUs) with the goal of expanding the network. In 2023, Nebraska Indian Community College joined the sovereign network, and other TCUs have expressed interest in participating. Unidata's participation in the project has opened doorways in other areas as well. The project team (including Unidata staff) is participating in an NSF-funded working group on data sovereignty (Earth Data Relations). The NSF project was also awarded an NCAR Collaborative Opportunities for Research Engagement award; funds from that grant will be used to put on an American Indian Higher Education Consortium (AIHEC) partners workshop alongside a data workshop for the sovereign network.

- **Educational Program Enhancements**. Supplemental NSF funding supported the addition of a dedicated Educational Designer to the Unidata staff during the award period; as a result, Unidata has re-evaluated its technical training efforts. New initiatives include implementation of an eLearning system for publicly-available asynchronous learning materials and use of the Kirkpatrick Model for Training Evaluation to assess the impact of learning experiences. Unidata now provides both synchronous opportunities (webinars, instructor-led training, workshops, train-the-trainer sessions) and asynchronous offerings (eLearning modules, eLearning courses, tutorials & reference guides, videos, microlearning simulations, self-assessments and needs assessments) as well as blended learning experience types to fit multi-phased approaches that deliver effective learning for varied learners and multi-year objectives.
- **AI/ML Initiatives**. During the award period, supplemental funding from the NSF allowed the program to hire an engineer to focus on artificial intelligence/machine learning topics, with the end goal of reducing the "time to machine learning." Unidata's software development teams have begun work to enhance our existing packages to better support ML workflows — for example by adding support for cloud-native storage formats like Zarr — and are engaging with the Unidata scientific community to understand the "pain points" encountered as they adopt AI/ML techniques in their own research and teaching. An initial focus has been to provide "scaffolding" for ESS educators and students, so they better understand where machine learning techniques can be applied to their investigations. In support of this effort, Unidata's AI/ML and educational design staff were awarded an NSF CyberTraining program pilot grant titled "Machine Learning Foundations and Applications in the Earth Systems Sciences." Work on the pilot project, which will be conducted in association with Metropolitan State University of Denver, is beginning in the third quarter of 2023.
- **Internships and Mentoring**. In the years 2019-2023, the UPC has hosted a total of fifteen summer student interns, who built skills in software development and science communication. Unidata summer interns have presented their work at AMS annual meetings and have served as student ambassadors for the Unidata program. Most recently, two former student interns returned as presenters at the 2023 Unidata Users Workshop. Former interns have joined the Unidata Users Committee as both academic representatives and graduate student representatives, and one former intern has joined Unidata's technical staff as a software engineer.
- **Support for Cloud-Native Data Formats**. Recently, the Unidata has extended the netCDF-c library to provide access to data in object storage systems from commercial and non-commercial cloud service providers. The netCDF-C library accomplishes this by providing a mapping from a subset of the full netCDF Enhanced data model (aka netCDF-4) to a variant of the Zarr data model that already has mappings to key-value pair cloud storage systems. The netCDF version of this storage format is called NCZarr.
- **Community Workshops**. Originally scheduled for 2021, but delayed due to COVID-19 meeting restrictions, Unidata organized and hosted a Unidata Users Workshops in June 2023. The

workshop, titled “Storytelling with Earth System Science Data: Challenges and Opportunities for Effective, Ethical, and Reproducible Science,” brought community members together to learn new tools and techniques to analyze data from multiple disciplines, make use of AI/ML technologies to better understand data, and implement CARE and FAIR practices.

- **Regional Workshops**. Unidata staff conduct workshops focused on building skills with Unidata software packages in the context of the Earth Systems Sciences. While COVID-19 travel restrictions severely limited our ability to conduct in-person workshops for several years, by holding a mix of in-person and virtual software events we were able to reach more than 700 attendees at some 20 different workshops over the award period. Most recently, we organized an in-person training event serving students from the University of Maryland Baltimore County, Howard University, Morgan State University, and Johns Hopkins University.
- **Equipment Awards**. During 2019-2023, Unidata provided equipment grants to 17 different 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. In keeping with Unidata’s intention to encourage universities to experiment with modern workflows, during this period the UPC has entertained proposals for grants to fund purchase of cloud-computing resources in addition to traditional computing hardware.
- **MetPy and Siphon**. Program Center staff lead development of the open source MetPy project, which has been downloaded over 965,000 times. MetPy has an active and growing community of users and developers; according to GitHub, 425 repositories and 47 distinct packages depend on MetPy. Similarly, the Siphon project, a Python library to facilitate access to remote data sets, has been downloaded over 379,000 times. The MetPy and Siphon projects have greatly benefited from community participation; individuals outside the UPC have submitted almost 400 bug fixes and feature contributions.
- **AWIPS**. Unidata has refactored and repackaged the National Weather Service’s Advanced Weather Interactive Processing System (AWIPS) into a configuration suitable for use at universities and other organizations with limited computing resources. In addition to making the Unidata AWIPS distribution available, the UPC operates a cloud-based EDEX data server in the NSF Jetstream cloud (as part of the Science Gateway project), allowing universities provide courses using the AWIPS client software (CAVE) without the need to expend resources to build and configure a local data server.

Additional Highlights 2018-2023

In addition to the above-listed achievements, the following sections provide a brief summary of some of the other accomplishments in the strategic goal areas outlined in NSF 1901712.

Managing Geoscience Data

Objective	Accomplishments
Making Geoscience Data Accessible	Unidata has worked extensively with NOAA to ensure community access to GOES-series satellite data streams. Unidata’s IDD data servers have an uptime of 99.96%, and a network topology that helps ensure the flow of data even in the event of an outage. Unidata now operates an IDD cluster at the NCAR Wyoming Supercomputing Center

Making Geoscience Data Usable	Unidata is deeply involved in the development of the Climate and Forecast (CF) metadata conventions for netCDF. UPC staff serve on the CF Governance Panel and assist in the organization of annual CF workshops. Because relaying data from the NWS NOAAPort system is an important component of Unidata's data services, UPC staff created technology to merge broadcasts from geographically distributed sites into a single, high-quality stream that is distributed via the IDD.
Enhancing Community Access to Data	Unidata has begun a partnership with the UCAR COSMIC Program to provide Radio Occultation data on the IDD. Unidata distributes value-added Level 2 products created at Texas Tech University from the current GOES satellites' Geostationary Lightning Mapper (GLM) instrument.

Providing Useful Tools

Objective	Accomplishments
Promoting 21st Century Scientific Workflows	The MetPy project created a "declarative interface" for the package to give users of the GEMPAK visualization package an easy pathway to convert their workflows to use Python.
Creating Modern Data Management Workflows	Building on enhancements to the netCDF libraries, the TDS implemented prototype read support for the modern Zarr object data storage format.
Supporting Legacy Workflows	UPC staff worked with community members to transition the GEMPAK visualization package to a community support model, freeing Unidata resources for other projects while preserving access to the software for those who still rely on it.

Supporting People

Objective	Accomplishments
Providing Technical Support	Unidata provides technical support for its software and data services to community members at no charge. Over the period covered by NSF 1901712, UCP staff fielded nearly 27,000 support queries.
Building Capacity through Open Source Software Development	Unidata maintains publicly available source code repositories for its software packages on the GitHub platform, and actively solicits contributions of code, issue reports, and documentation from community members. While community participation is greatest for software written in the Python programming language, there is also significant participation for packages written in Java and C. UPC staff work with community submitters to resolve issues and incorporate code contributions.

	UPC engineers contribute code and expertise to a variety of "upstream" software packages that underlie Unidata's offerings.
Building Community Cyber-Literacy	Unidata has established an eLearning site to provide asynchronous learning experiences related to Unidata and other ESS software. Learners can access the resources at no charge, and instructors can incorporate modules into their own learning management systems. Since 2019, UPC staff have offered introductory-level workshops on using Python in the atmospheric sciences at the AMS Annual Meeting's Student Conference. To date, over 350 students have participated in these free workshops.

Intellectual Merit: Reimagining Unidata

In the Unidata Program's nearly 40 years of service to the ESS community, advances in the environmental data landscape, including monitoring, processing, analysis, visualization, and application, have profoundly changed the way researchers, educators, and learners interact with data and with each other. Over this history our sphere of activity has expanded significantly, from early efforts to ensure that university atmospheric science departments had access to weather data in near real time to supporting a broad community of ESS researchers and educators with a wide palette of data and analysis tools. The success of our efforts is evidenced by the wide adoption of Unidata tools, techniques, and services across the ESS landscape, with academia, government agencies, and industry all benefiting from our efforts.

It would have been natural to suppose that we had already found the most effective ways to serve our community. But decades of experience have taught our program to greet change as an opportunity for improvement, recognizing that the conduct of research and education in ESS is continually evolving to meet the changing landscape. For instance, when the COVID-19 pandemic spurred the academic community to shift to remote learning, Unidata moved quickly to provide cloud-computing based resources for university instructors, but also started down a path of revamping its existing in-person training methods to better support asynchronous, remote, and hybrid learning opportunities. As societal imperatives shifted toward the urgent and critical need to lower barriers to participation by underserved populations in ESS education and research activities, Unidata focused on expanding its collaborations with MSIs, but also began looking for ways to support communities in creating their own self-governing data and knowledge networks. The ever-increasing volume of ESS data validates our efforts to help shift the community's data workflows toward cloud-based data access, but it has also spurred us to begin partnerships with community members working to evolve existing data formats for efficient and robust cloud-native operation. And the rapid increase in the capability of ML techniques is driving Unidata's efforts to both make it easier to use ESS datasets with these new algorithms and to increase the capacity of today's university students to understand and use the rapidly advancing technology of Artificial Intelligence.

As part of our strategic planning process, these and other factors discussed below have led Unidata to look closely at its approaches to serving the broad ESS community. While our community depends on Unidata's facilities for making ESS data broadly accessible, and on our support for widely-used software packages for data management, analysis, and visualization, we have identified ways in which we can re-envision our connections with the community to make our offerings more useful, effective, and relevant to ever-changing needs of the ESS community. This section outlines the primary drivers for this reimagining, and touches on ways we plan to shift our focus for even greater community impact.

Motivating Concepts for a Reimagined Unidata

As part of Unidata's strategic planning process (Unidata 2023), we looked closely at the data challenges facing today's ESS community, in the context of our rapidly changing technological environment. The following are the key concepts that have motivated the selection of our strategic goals and the course of action we are mapping for the coming years.

Importance of Convergence Science Approaches

A *convergence science* (NSF 2023) approach to ESS seeks to use techniques of interdisciplinary collaboration to address "Grand Challenge" questions in areas such as climate change, ocean health, natural disaster resilience, and effective collection and use of Earth observation data. Global collaborations that bring together diverse types of expertise in natural, physical, social, biological, and information sciences, intermingling knowledge, theories, methods, and data have the potential to inspire innovative courses of action with respect to these complex problems.

Unidata can foster convergence science approaches by supporting communities in the creation of holistic solutions to the need for ESS data. Researchers need robust data services that enable them to access the ESS data they need from different domains in ways that complement their workflows and without transferring massive datasets. Unidata's existing cyberinfrastructure, coupled with planned enhancements to cloud-native data storage and access mechanisms, move scientists a significant way towards this goal. Researchers need freely available analysis and visualization tools that are intuitive, easy to use, and well-integrated with their existing systems. Unidata software tools are designed to interoperate with others in use across a wide range of ESS disciplines. Furthermore, Unidata's ongoing Science Gateway project seeks to support convergence science approaches by making these tools and data streams available (alongside other community-supported resources) to diverse scientific communities in ways that are easily accessible and simple to use.

Importance of Collaborative Community Approaches

Communities of Practice are groups of people who share a common goal or interest, and who seek to interact with one another in order to learn and improve. Expanding on this idea, collaborative community approaches bring practitioners together to share resources and act in concert in addition to learning from one another. As communities expand — geographically, across scientific disciplines, and in total numbers — it becomes increasingly important to build infrastructure that supports efficient ongoing collaboration.

Building on its experience creating cyberinfrastructure resources for small ESS communities through the Unidata Science Gateway, Unidata sees the creation of *Community Hubs* as a path toward providing useful organizational tools to disparate communities of practice. In this vision, community hubs will take advantage of the technological underpinnings of the Unidata Science Gateway, but also serve as loci for both scientific and social interaction among members. The nascent Sovereign Network Project (mentioned on page 8) is one example of how a community hub might evolve; other possibilities include community hubs organized around regional ESS phenomena, specific scientific investigations, or even tool creation. By combining learning facilities, access to a wide variety of ESS data, software tools for analysis and visualization, and a platform for communication between participants, the hubs will serve to meet the needs of specific individual communities. In providing a space for people to meet and participate in community activities, community hubs begin to resemble Collective Impact networks (De Moor 2018), in which a cooperative network of community members, organizations, and institutions advance equity by learning together, aligning, and integrating their actions to achieve population and systems level change.

It is important to note that Unidata's conception of community hubs is anchored in ideas of self-governance and local control. Through a series of "listening workshops" we hope to begin identifying self-defined communities of interest that want to take advantage of Unidata resources and expertise to increase their own impact. While the Unidata program will provide guidance, tools, and (potentially, through the Unidata Community Equipment Awards program) hardware, it will not serve as a controlling or governing body. Decisions about content and access will be made by the communities themselves, based on their needs and specific interests. Beyond providing resources, Unidata hopes to play a role in linking the hubs into a community network, allowing them to share information and resources with each other as they find appropriate.

Enabling the establishment of community hubs and fostering their development is a key part of Unidata's strategy moving forward. One of the first steps down this path will be to identify and recruit *Unidata Community Hub Champions*, much like the NSF ACCESS program's Campus Champions (Campus Champions 2023), who help spread familiarity with the ACCESS program by providing information about services and opportunities. By distributing knowledge about Unidata resources (and, eventually, those from established community hubs), the community hub champions will work to empower researchers, educators, students, and other community members to use hub resources efficiently and productively to advance their community goals.

Importance of Broadening Participation, Increasing Diversity and Inclusivity

The lack of diversity in ESS creates barriers to progress and innovation. According to NSF 01-53, *Strategy for Developing a Program for Opportunities for Enhancing Diversity in the Geosciences* (NSF 2001), demographics data confirm underrepresentation of women, African-Americans, Hispanic-Americans, Native-Americans/Native-Alaskans and persons with disabilities in science and engineering in general, and in the geosciences in particular. Further, it states that the ESS community is not availing itself of the talents and skills of a large and increasing segment of society — individuals from underrepresented groups — and that there is a need to take steps to actively address this problem. As society faces grand challenges to live responsibly, sustainably, and ethically, increased diversity has clear benefits for scientific advancement. The scale and complexity of grand challenges require that Earth Systems scientists use a range of approaches to broaden participation and inclusivity by engaging with new groups with diverse interests, life experiences, and perspectives, finding unique approaches to solve societally relevant problems. For these reasons, in its Dear Colleague letter on the GOLD-EN program, NSF affirms: "Diversity is a vital priority for the geosciences community because it promotes innovation, strengthens the community's ability to tackle complex geoscience research problems, and engenders widespread public Earth and environmental science literacy."

Unidata's approach to this work is to itself participate in as many scientific contexts as possible, expanding awareness of Unidata's services in MSIs, including HBCUs, TCUs, and communities identified as part of the above-mentioned community hubs program. We have learned that strong community ties grow best when backed up with long-term commitments and participation, and through active partnerships to expand learning opportunities, data access, and knowledge sharing.

Importance of STEM Education, Training, and Workforce Development

To push the boundaries of scientific frontiers and drive innovation requires a wide pool of workers who are well-versed in science, technology, engineering, and math (STEM) disciplines. Specifically, enhancing students' scientific, data, computational, and geospatial literacy will be critically important to the next generation workforce, who will be asked to solve some of the most complex problems of our time. The current and future needs for skills associated with strong STEM literacy are no less urgent in ESS than in other scientific fields, and a focus on convergence science approaches suggests the importance of STEM skills for society at large. And in thinking about this need, we note the Pew

Research Center analysis that Black and Hispanic workers are currently underrepresented in STEM jobs relative to their shares in the U.S. workforce as a whole (Pew 2021).

Unidata has decades of experience in supporting STEM education by providing data, tools, and training to member universities. Our vision now is focused on expanding access to Unidata's resources to additional communities; we plan targeted outreach to MSIs to offer training services, provide technical assistance, and aid in the creation of community hubs. Improvements to the way we organize and plan learning experiences, as well as work to provide additional remotely accessible, synchronous and asynchronous learning opportunities guide us in this area.

Importance of Open and Reproducible Science Practices

A host of U.S. Federal science agencies have designated 2023 as a Year of Open Science, with the goal of advancing adoption of open, equitable, and secure science practices. The declaration acknowledges that the results of scientific research have not always been available to others, and that stated findings have not always been reproducible. It responds to a 2022 White House Office of Science and Technology Policy (OSTP) mandate that agencies undertake new plans to ensure that peer-reviewed publications and data arising from federally funded research be made immediately and freely available upon date of publication.

Unidata looks to contribute to this push for open and accessible scientific data through enhanced access to its suite of data management technologies, including the THREDDS Data Server, improved cloud-based data storage mechanisms such as the NCZarr project, and ongoing access to data via the IDD network. We continue to actively advocate for strong community-based standards like the CF metadata conventions for netCDF, which enhance reproducibility of scientific findings though accessible and understandable metadata. All Unidata-created software is itself Open Source, available to be used, inspected, and improved by anyone.

Importance of FAIR, CARE, and Ethical Principles

FAIR data principles (Wilkinson et al. 2016) suggest that scientific data should be Findable, Accessible, Interoperable, and Reusable, making it easier to share and use data, software, and platforms. CARE principles (Carroll et al. 2020, Carroll 2021, GIDA 2023), developed by the International Indigenous Data Sovereignty Interest Group (a part of the Research Data Alliance), complement FAIR principles to ensure that data are used for Collective Benefit, with Authority to Control, Responsibility, and Ethics. CARE principles articulate Indigenous Peoples' sovereignty over data about their communities in the context of open data and open science.

Unidata's data management technologies have long supported FAIR principles, by promoting the role of descriptive metadata, use of remote data access mechanisms, standard data formats, and open access. As promoters of a central role for data in the Earth Systems Sciences, we feel a responsibility to help socialize CARE principles as well. We will do this through technical mechanisms within our data management software when possible, by working for adoption of metadata standards that support effective use of the principles, and through educational efforts to explain the historical uses of Indigenous environmental data that led to their creation.

Impacts of a Rapidly Changing Technological Environment

The technological environment in which research and education are conducted is rapidly evolving. While advances in science and technology have always had synergistic and reinforcing effects on each other, the pace of change in recent years has led to even more profound impacts for the ESS community. Advances in information technology have increased research output through high-performance computing, cloud-based data storage, high-speed networks, and data-proximate analysis

techniques. New ways of communicating and sharing data have increased collaboration among scientists worldwide. Scientists who have come of age in the era of mobile computing and ubiquitous networks have different expectations about access to data and what types of technology should be required to use that data. The disruption of academic institutions and workplaces by COVID-19 restrictions on physical gathering led to speedier adoption of tools for remote work and online learning. And the advent of widely available generative AI tools like ChatGPT, Bard and others have spurred a rush to understand the implications of machine learning tools to create transformative learning opportunities for students, accelerate scientific discoveries and scholarly research, and advance science and society.

Unidata acts as a trusted technology counselor for our community, investigating how innovations can improve or streamline scientific workflows and sharing what we learn to help speed adoption of useful techniques. Our academic partners value this role for Unidata; our advance scrutiny allows them to work toward adoption of new technologies with confidence that their effort will not be wasted in blind alleys.

Impact of Growing Data Volumes

The volume of Earth observation data and predictive model output continues to grow at a phenomenal rate, as new instruments, platforms, and extremely high-resolution and ensemble prediction systems are deployed. One indicator of the ESS community's response: the amount of data requested by universities participating in the IDD network has increased nearly sevenfold in the past decade alone, to a current volume of over 90 Terabytes each day — even as the community begins to shift to cloud-based data access and data-proximate analysis workflows.

Unidata strives to provide innovative solutions to the rise in data volumes. We are working to increase the efficiency of our remote-access data systems, and to develop processes to create datasets that are "analysis ready" or "AI ready" in the sense that ML algorithms and other scientific applications can readily digest the valuable information in enormous datasets without the need for correspondingly enormous data transfers or additional processing. We have already made significant progress in cloud-optimized access to existing data stored in the netCDF format, paving the way for cloud-based analysis of data ranging from real-time GOES satellite imagery to climate reanalysis datasets.

The Upshot: How Unidata is Adapting

The Unidata Program has a long history of providing data services to an ever-growing community. Over time, our scope of activity has expanded from providing real-time observational data to established university atmospheric science departments to working for the broad accessibility of all types of data to researchers, educators, and students across ESS. As we re-imagine Unidata's role in this complex ecosystem in the context of the motivating concepts discussed above, we have tried to focus on moving the ESS community forward through collaboration, inclusiveness, education, and technological advancement. The initiatives described in our Plan of Action will help us focus our efforts on adapting to the changing needs of the ESS community and society at large.

Broader Impacts

Service to the broad ESS community is at the core of all of Unidata's activities. As shown in Table 4 (page 6), Unidata's software and services are in use by over 2200 organizations including more than 1800 academic institutions in 222 countries and territories. Unidata technologies benefit research labs, weather services around the world, national and international agencies, and scientific projects. Across a wide range of disciplines, scientists and service providers use Unidata's scientific software

libraries to conduct their research and share data with others; Unidata technologies are built into more than 20 commercial and 100 open source software packages.

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 are compelling measures of Unidata's impact on research; as we begin implementing the community hubs program, we expect the impact of Unidata's programs on research activities to increase further

As a baseline, over the five-year period 2018-2022 a search of American Meteorological Society journals retrieved 364 papers referring to Unidata software or data services; 862 papers referring to Unidata software or data services were published in American Geophysical Union journals during the same period. Many of these papers refer to Unidata software or data services without mentioning the Unidata Program itself; this correlates with anecdotal evidence of widespread use of Unidata products (especially netCDF) beyond the communities traditionally served by Unidata.

A review of citations reported by the Google Scholar search engine revealed that between 2018 and 2022, Unidata software and data services were cited 17244 times in the full range of scholarly literature encompassed by the search engine. Publications mentioning Unidata software include *Geoscience Data Journal*, *Environmental Engineering and Management Journal*, *Pure and Applied Chemistry*, *Ecological Informatics*, *Ocean Engineering*, *Computational Molecular Science*, and *Agronomy*.

Impact on Education

Unidata software and data services are currently in use at over 480 U.S. colleges and universities and more than three times that many in other countries; the list includes universities in all 28 EPSCoR jurisdictions, including many institutions that have a large number of students from underrepresented communities. We expect the initiatives described in this proposal, along with existing programs providing equipment grants, local training workshops, and data services, to continue expanding our impact, especially among MSIs.

Unidata has recently begun the process of refining its own teaching methods to better focus on student preparedness and specific learning goals. We expect these changes in our approach to continue to bear fruit; the local organizer of a 2023 workshop at the University of Maryland Baltimore County for students from UMBC, Howard University, Morgan State University, and Johns Hopkins University had this to say:

> *Thanks again for arranging the Python course; my staff/students here mentioned how useful it was. I also want to ask you things about this type of site visit teaching/training plan. It's sort of a flip classroom thing but with the mobility to reach out to more people. Would you mind sharing some details (e.g., logistics, human resources, costs, etc.)? I am planning some other activities like this for my partner institutions.*

Unidata has also begun to shift towards training techniques that make more efficient use of staff time, including "Train the Trainer" style workshops. A recent series of events held in partnership with Colorado State University focusing on exploratory data analysis included Unidata-led sessions for undergraduates followed by a separate session for teaching assistants. Said one of the participants in the second portion:

> *The resources that were shared with all of us seem really valuable, and I will reference them in my own teaching immediately. The training was a great way to get familiar with them in an environment with experts on the material, and to get some elaboration on them and help with any unclear parts.*

Additionally, Unidata programs to create asynchronous learning resources available online to any learner will continue to expand access for those unable to travel to attend workshops.

Sovereign Network Project

Unidata is an active participant in NSF grant 2131301, "A Sovereign Network System for Environmental Monitoring, Data and Information Exchange, and Collaboration among Tribal Colleges and Universities." Working closely with partners at NTU and SIPI during the pilot phase of the project, Unidata provides technical expertise, software support, and engineering effort to create a proof-of-concept network that is viable for sites with limited internet bandwidth. Through talks and presentations at community events, the project team has generated significant interest among other TCUs, and hopes to include all 37 members AIHEC in the Sovereign Network within five years.

In corollary efforts:

- Unidata staff take part in the Earth Data Relations Working Group funded by NSF incubation grant 2220614 "Responsible Research and Engagement Practices for Indigenous Data Governance in Earth Science Institutions." The aim of this project is to build on CARE Principles by developing a framework for implementing CARE in ESS open data repositories.
- Unidata staff collaborate with members of Oglala Lakota College and the Environmental Data Science Innovation & Inclusion Lab (ESIIL), focusing on data access, visualization, and analysis of ESS data. This collaboration is also exploring ways to integrate their efforts into the Sovereign Network Project.

Unidata Summer Internships

The ongoing Unidata Summer Internship program offers graduate students and upper-level undergraduates the opportunity to work closely with Unidata's professional staff on projects that interest them and benefit the wider Unidata community. Interns interact with software developers, instructional designers, technical writers, and other mentors within the program, building skill with modern software development processes, learning about data-enabled science, and generally getting to know what it's like to work in a community-focused organization supporting ESS research and education. Between the program's inception in 2013 and the summer of 2023, a total of 26 students have taken part in Unidata's internship program, including 7 from MSIs. Former interns have joined national laboratories and state departments of natural resources, become NASA FINESST fellows and assistant professors, joined Unidata's governing committees as both academic representatives and graduate student representatives, and even joined Unidata's technical staff as software developers.

SOARS

Unidata continues to collaborate with UCAR's Significant Opportunities in Atmospheric Research and Science (SOARS) program, providing program support and mentorship for SOARS protégés. Since 2018, SOARS has been making use of JupyterHub resources on the Unidata Science Gateway for its Computation and Data Workshops. In addition, Unidata supports individual SOARS protégés with tailored Science Gateway resources allowing them to accomplish their summer work without needing to be physically present in Boulder.

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 institutions to join the Unidata community and to allow existing members to continue and enhance their participation. During the past five years, Unidata made 17 awards to institutions ranging from major U.S. research universities to community colleges. Projects funded in the past five years include:

- Introducing an AWIPS EDEX data server and multiple AWIPS CAVE client workstations for the Jackson State University Meteorology Program.
- Valparaiso University's installation of a JupyterHub server to enhance the use of Python in undergraduate ESS coursework.
- Hardware upgrades for the College of DuPage's widely used NEXLAB server.
- A joint project at the University of North Carolina's Charlotte and Pembroke campuses to create a remote collaborative geoscience learning environment.

Going forward, Unidata will allocate a minimum of 20% of Equipment Awards program funds to support projects at MSIs. A complete list of 113 past projects funded at 66 colleges and universities is available online on Unidata's web site.

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. Within NOAA alone, at least 26 distinct divisions or groups are known to be using Unidata technologies to access, distribute, or archive data.

The letters of commitment to collaborate from some of Unidata's partners, provided in Appendix A, provide a sampling of the array of organizations that benefit from Unidata's work.

Unidata's open source software projects are built upon and are part of a broader ecosystem of scientific tools, especially scientific Python. Unidata staff often contribute problem reports, and sometimes code fixes and maintenance, to projects used by Unidata tools. Between 2018 and 2023, Unidata developers have made over 300 contributions to non-Unidata open source projects via GitHub. This contributes to the overall health of these projects and the broader scientific software ecosystem spanning many scientific domains.

Synergistic Activities

Unidata's mission to serve the broad ESS community leads us to undertake activities and participate in projects that are funded through non-core awards. In choosing activities funded by other awards, we select opportunities that allow us to contribute in ways that are aligned with our mission and synergistic with our ongoing core-funded activities. We believe that these projects and the resulting collaborations are essential to maintaining a healthy program. In addition to the intrinsic merit of the projects, non-core funding brings modest additional resources to the UPC, relieving some of the pressure on Unidata's budget. Both the Unidata Strategic Advisory Committee and NSF have encouraged Unidata to participate in synergistic opportunities as appropriate.

Currently, the UPC is participating in the following projects that are not funded by NSF 1901712:

- A NASA-funded activity in partnership with Navajo Technical University to create innovative use of NASA's Astrobee microgravity robot on the International Space Station.
- A NOAA-funded activity for "Establishment of a GOES-S Downlink and Remote Access Facility by COMET and Unidata for Earth-system Training, Research and Education."
- An NSF-funded activity in partnership with Southwestern Indian Polytechnic Institute to create "A Sovereign Network for Environmental Monitoring, Data and Information among Tribal Colleges and Universities."
- An NSF CyberTraining program funded pilot program "Machine Learning Foundations and Applications in the Earth Systems Sciences."
- An NSF-funded activity in partnership with University of Illinois and several other institutions for the "HDR Institute: Geospatial Understanding through an Integrative Discovery Environment."
- An NSF-funded (OAC Software Institutes) activity "Scaling MetPy to Big Data Workflows in Meteorology and Climate Science."
- An NSF EarthCube program funded activity in partnership with NCAR CISL for "Project Pythia: A Community Learning Resource for Geoscientists."
- An NSF Jetstream2 program funded activity in partnership with the University of Indiana: "Accelerating science and engineering on-demand."

Unidata undertakes non-core projects only after careful analysis of their merit and benefits to the ESS community, and endorsement by the Strategic Advisory Committee. We look for novel projects that provide us with opportunities to strengthen our core programs, build technical capacity, expand our reach to new communities, and gain expertise in ways that would not otherwise be possible with the resources of our core funding.

Strategy and Work Plan

As Unidata reimagines its role as a community facility supporting ESS research and education, we strive to balance our approach to our community's needs for data, tools, and services in the present with activities that will enhance their ability to meet future challenges. To be successful, we will need to maintain and enhance programs our community finds useful, deploying resources efficiently for new endeavors while remaining agile enough to respond to changes in the technological landscape.

This section describes our approach to each of the primary goals described in our strategic plan, along with some of the key actions we envision taking to make the plan a reality.

Providing Data and Tools

Our goal is to ensure fair and equitable access to ESS data, and to provide tools to analyze and visualize that data. To achieve this goal, we will:

- Provide performant, high quality data storage, access, and visualization tools
- Provide core cyberinfrastructure for community data access and use
- Build community capacity to access and use cloud-computing infrastructure
- Promote FAIR and open access to data, tools, and services
- Support convergence science approaches to ESS questions

Action: Provide Easy Access to ESS Data

Providing the community with access to ESS data remains a core activity for the Unidata program. Unidata's IDD now pushes 90 Terabytes of data to community users each day using the LDM. The continued growth of cloud computing has resulted in an increasing emphasis on publicly available hosted collections of data that can be easily accessed using protocols and formats that are optimized for the cloud. To ensure the community's continued access to data in ways that help ensure a smooth transition to modern workflows we will:

- Continue coordinating the IDD and supporting community participation in data distribution to facilitate the community's access to real-time data from operational weather centers. The data coming over the IDD will also continue to be hosted on THREDDS Data Servers hosted at the Unidata Program Center. In order to promote easier location and use of these data, we will expand the resources available that identify and describe data available from the IDD.
- Facilitate access to data collections that are stored in cloud-based object stores. This includes simplifying their access with Unidata tools and continuing to engage with NOAA, NASA, and other agencies and data providers as they continue migrating data collections to cloud environments and analysis-ready, cloud-optimized formats (as has previously been done for NWS NEXRAD datasets).

Action: Strengthen Support for Cloud-focused Storage of ESS Data

As the volume of ESS data continues to expand, improving ease of access to data stored remotely is imperative to support cross-disciplinary Convergence ESS, Open Science approaches, and adoption of FAIR principles. We will:

- Rearchitect the THREDDS Data Server to better support cloud-focused workflows and data-proximate analysis. The envisioned "microservices" architecture will make it much easier for community members to contribute to TDS development efforts. Early goals prioritize a "minimum viable product" version of the server, with catalog, OPeNDAP, and file download components running as independent microservices during the initial phase, with involvement by community members (running the new TDS version themselves or contributing code) following.
- Provide native support for Zarr-formatted data across the netCDF library family. Early goals include support for reading from and writing to Zarr data stores, both locally and remotely, across the entire family of netCDF implementations supported by Unidata. Once robust library support is available, publish resources using Unidata analysis tools (IDV, MetPy) to analyze and visualize data from remote Zarr stores. We will also investigate approaches like the Kerchunk Python library for gaining the benefits of the Zarr format without the need to convert existing data collections.

Reducing Barriers to Participation

Our goal is to build a larger, more inclusive ESS community. To achieve this goal, we will:

- Promote and integrate accessibility, equity, and ethical principles, e.g. FAIR/CARE
- Provide equitable access to high quality learning and research resources and services
- Actively engage with underrepresented communities to build partnerships and strong community relationships
- Broaden and support our community by active engagement, outreach, and advocacy
- Foster next-generation workflows and career skills

Action: Enhance the Unidata Science Gateway

The Unidata Science Gateway is a key component of Unidata's strategy to provide equitable access to leading edge ESS tools to a wide variety of communities. The gateway serves as a testbed and proving ground for Open Science technologies and educational techniques that can be implemented by individual community hubs to support their specific needs and goals. We will:

- Consolidate Unidata's educational offerings into a single educational resource, integrating Convergence Science datasets and techniques with materials that are made freely available to ESS educators and students. Initial efforts will focus on existing Unidata-developed technologies, with expansion over the project period to include both additional Unidata-led topics such as AI/ML-ready datasets and community-generated learning materials.
- Create a robust network catalog linking diverse ESS datasets directly to resources on the Unidata Science Gateway, enhancing Gateway users' ability to investigate cross-disciplinary questions. High-priority candidates for inclusion include coastal and other hydrologic datasets and social/GIS datasets such as land use or population.
- Expanding on current efforts as part of a pilot project funded under a separate NSF CyberTraining grant, incorporate low- and no-code resources and training materials into Unidata Science Gateway offerings. Work funded by the CyberTraining program is underway; integrations with the Unidata Science Gateway will begin in the second year of this project.

Action: Inaugurate Network of Community Hubs

Building on the ideas of the Unidata Science Gateway, we look to make it possible for self-governing communities of practice to create their own community hubs. In this endeavor, Unidata plays the role of organizer, convener, facilitator, and technology resource. We will:

- Collaborate with the existing Sovereign Network Project team (TCUs and partners) to begin addressing the technology needs of indigenous communities for ESS data governance, access, and communication. This effort will begin immediately and inform decisions about the organization of future community hubs and serve as a case model platform for interoperable, convergence science.
- Convene one to three "listening workshops" with participants from diverse ESS communities in the first year of this project. (This activity will proceed in parallel with the Sovereign Network Project.) Our goal is to develop a stakeholder gap analysis and summarize community needs, interests, and priority areas of support.
- Through the listening workshops, identify one to two additional communities with whom to partner in creating Community Convergence Science hubs. By the end of the period of performance, one to three independent but affiliated community hubs, in addition to the community associated with the Unidata Science Gateway itself, will be in operation.

Action: Reimagine our Learning and Development Services to Foster a Diverse and Inclusive Workforce

To improve accessibility and support the development of a transdisciplinary STEM workforce and next generation, we will work to develop and deliver community-needs based learning and development services that increase diversity, equity, inclusion, and support the broad community to advance ESS. We will:

- Broaden the scope of educational resources available directly from Unidata. As part of the Unidata Science Gateway project, we will focus on creating needs-based learning experiences

that span disciplines. By using a collaborative approach, we hope to deliver effective experiences for learners from varied communities and across starting skill levels.
- Offer educational resources and guidance to groups working to create community hubs, working in parallel with the Unidata Science Gateway efforts.
- Begin work by performing needs analysis, allowing us to create effective learning resources with tailored self-assessment tools.
- Incorporate measurable learning objectives, web accessibility compliance, and adherence to modern adult learning design techniques.

Action: Raise Awareness and Foster Use of FAIR and CARE Data Principles

To effectively value both people and purpose in ESS pursuits, Unidata will work to promote and implement FAIR Principles and CARE Principles for Indigenous Data Governance. By helping to socialize the importance of the FAIR and CARE principles, we hope to promote equitable and accessible participation and outcomes. To work as advocates for open and ethical ESS data practices, we will:

- Consult with our collaborators on the Sovereign Network project as well as partners within the Earth Data Relations working group (NSF Award #2220614) (NSF 2022a) to capture their insights and recommendations on sharing and integrating FAIR and CARE principles and facilitating indigenous data sovereignty and data governance.
- Develop guidance on use of Unidata tools alongside FAIR and CARE principles.
- Work within the CF metadata standards (CF 2023), Open Geospatial Consortium (OGC), and Earth Science Information Partners (ESIP) efforts to socialize use of metadata that supports CARE principles including Authority to Control.
- Work to ensure that all Unidata communications, both internal and public-facing, are inclusive and accessible. Initial steps will include implementation of Web Content Accessibility Guidelines (W3C 2023) for online materials.

Fostering Community Action

Our goal is to engage productively with community members, working collectively to improve ESS teaching and research. To achieve this goal, we will:

- Re-envision the Unidata Science Gateway as a demonstration Community Hub
- Marshal and build strategic partnerships with data providers on behalf of the community
- Support community adoption of new techniques and approaches to solve ESS problems
- Work collaboratively with a broad and expanding community of ESS researchers, educators, and students to create and share data, tools, resources, and knowledge

Action: Engage and Collaborate with Members from ESS Disciplines MSIs

To foster community action, it is imperative that we broaden our scope of participation, increasing the diversity and accessibility of the Unidata Program. Unidata sits at the intersection of multiple sectors and communities; by supporting an inclusive approach to ESS we can help ensure that the field is open and accessible to all. We will:

- Visit a minimum of 2-3 MSIs each year to introduce Unidata's offerings and hold listening sessions that allow us to perform gap analysis, identify opportunities to engage and build partnerships with new communities, and elevate underrepresented voices.
- Partner with at least one MSI each year to host a training workshop.

- Broaden participation in the Unidata Program through strategic engagement and partner-based projects representing a spectrum of ESS disciplines and R2 and MSI representatives. Initial steps will focus on increasing committee member representation, broadening the range of award recipients, and working to attract summer interns from diverse communities and backgrounds.
- Continue to have staff participate in key societies, workshops, working groups, and conferences to engage with participants from MSIs and other institutions to advance shared goals.

Action: Build a Community of "Unidata Champions"

Partnering with the community, we will work to build a network of local Unidata "Community Champions" committed to training others and actively participating in Unidata's mission. Working with these individuals will allow us to better assess local needs for training, workshops, and other engagement opportunities. The goal of this network is to increase community engagement and create community-driven science and solutions to weather, water, and climate challenges. We will:

Develop a "Train the Trainer" program. By 2027, we will identify one to two MSI or R2 institutions with whom we can collaborate to host annual workshops for trainers and local Unidata community champions. The goal will be to provide individuals with technical information about specific Unidata software, suggested techniques for software training, and more general information about Unidata programs and services suitable for sharing with local communities.

- Leverage community participants including governing committee members, Equipment Award recipients, and other Unidata collaborators as "prototype" community champions. With UPC support, these individuals will work to foster awareness of the Unidata Program and its tools and services, and to support mutual learning and knowledge transfer to address complex science and societal challenges using Unidata data, products, and services.

Action: Foster Open-Source Development Communities around Unidata Tools

We hope to develop a stronger sense of community ownership of Unidata's software tools. While we will continue to lead development efforts, we will work to create opportunities for community members to collaborate in building open source ESS software. We will:

- Actively promote community project contributions to Unidata software projects across its community communications.
- Work to expand community understanding that all types of contributions (e.g. bug reports, documentation, feature suggestions) are important to open source projects, not just code.
- Evaluate holding regular, scheduled virtual community development calls/meetings to expand the scope of contributions to Unidata software packages; we will work with our governing committees to determine which packages to target initially.
- Investigate the feasibility of holding community hybrid "hackathons" to encourage contributions to one or more Unidata software projects.

Providing Innovative Technical Solutions

Our goal is to provide cyberinfrastructure leadership for the ESS community, helping move it toward innovative technical solutions. To achieve this goal, we will:

- Identify, shape, and consolidate trends to advance community capabilities with next generation technology solutions
- Support development and use of sustainable modern workflows, tools, and infrastructure

- Harness community expertise and encourage collaborative contributions in development and use of modern and novel workflows

Action: Expand Resources for Applying AI/ML Tools to ESS Problems

Artificial Intelligence and Machine Learning tools represent a rapidly growing technological frontier that is at the forefront of solving scientific problems. It is critical that Unidata, as it continues providing cutting edge ways to work with enormous volumes of ESS data, provides efficient data access processes for AI/ML tools and techniques. We will:

- Work with the netCDF community to develop standards for AI/ML model serialization.
- Develop and promote data optimization techniques for ML workflows.
- Enhance our learning resources to demonstrate how to use Unidata tools for data access, analysis, and visualization together with best practice AI/ML toolkits to solve authentic ESS problems. Our resources will also examine questions and provide guidance concerning the ethical use and training of machine learning models.
- Create collections of representative training data for sample ESS workflows, making the datasets available via Unidata's standard data access mechanisms (e.g. TDS/Zarr and Siphon), including the Unidata Science Gateway.
- Enhance our suite of visualization and analysis tools to include common tools for the analysis of ML models, especially supporting Explainable AI (XAI) workflows, which are growing in importance as a way to understand the results of model training.
- Streamline the process of using AI/ML workflows within Unidata's data access, analysis, and visualization tools, making it easier to access the large volumes of data needed to train ML models. We will investigate ways to improve parallelization capabilities within netCDF libraries to increase throughput for datasets stored as Zarr, and ways to have the TDS directly provide pre-processed data that is ready for ML model training.

Action: Evolve Software Tools to Cloud-first, Browser-based Environments

Many recent developments in computing technology have centered around migration to remote, cloud-based workflows and hosted software tools. Much of this migration is driven by the availability of alternative computing platforms, like tablets and inexpensive laptops with limited computing capacity. Environments like Unidata's Science Gateway make it possible to leverage these platforms as a portal to full-powered scientific computing tools. To support the use of these environments, we will:

- Integrate Unidata's locally installed visualization and analysis tools, like the IDV and AWIPS CAVE, into the Unidata Science Gateway, enabling their use over browser-based network connections. We will investigate ways to more tightly integrate these tools into the Jupyter Notebook environment that is currently at the center of Science Gateway workflows.
- Investigate the use of technologies like Javascript, Web Assembly and "Python in the browser" tools (*e.g.* pyodide, pyscript, jupyterlite) together with Unidata tools like MetPy and netCDF to build browser-based analysis and visualization environments. This approach has the potential to reduce barriers to entry by providing the "no install" experience of a hosted environment running on the user's local computing system.

Action: Collaborate on Standards-based Solutions

Innovative software solutions succeed when they mesh well with existing tools and workflows. Unidata has a long history of collaborating with community-based technical standards groups to propose and implement standards that formalize practices that have proven themselves to be useful in

widely-used software packages. To continue advocating for the ESS community inside these technical organizations, we will:

- Actively engage with technical groups including the Zarr project, the OGC, the WMO, and the Climate and Forecast metadata conventions for netCDF (netCDF-CF).
- Work with technical groups and other OSS projects to implement support for community standards in various software tools and test interpretability between these and other tools.

Conclusion

Looking back on Unidata's history, we are proud that our software and data services have made important contributions to the growth of data-enabled Earth Systems Science, supporting crucial scientific efforts such as the Intergovernmental Panel on Climate Change assessment reports and the National Weather Service's dissemination of WSR -88D radar data. But beyond easily identifiable achievements, we value the close relationship that has developed between the UPC and our community, allowing program and practitioners to succeed together in the day-to-day conduct of research and education.

Looking forward, we are even more excited by the opportunity to play a role in the technological and social evolution of ESS as a field. Technology-driven modes of scientific exploration, revealed most notably by recent advances in AI/ML, are giving the sciences new momentum that grows more thrilling by the day. AI-enhanced tools have the potential to save researchers from mindless labor, opening up opportunities for human creativity and potentially leading to breakthroughs that might otherwise have required decades of painstaking effort. Unidata has navigated such changes for decades and empowered its users along the way; the plans outlined in this proposal will allow us to continue guiding our community towards the innovations that are most useful and productive for their own research and education undertakings.

But while we're eager to embrace innovations where they can benefit our community, we also appreciate that community's needs for data, tools, and support *right now*. It's crucial that Unidata sustain the services our community relies on day-to-day, even as we chart a path forward. We would be privileged to have the chance to do this work hand-in-hand with those we serve, and to gain the benefit of wider perspectives as we strive to increase diversity and equity, broadening the community of scientists to better include historically underrepresented groups. Our community hubs concept promises to bring new and existing groups together, along with new technologies and methodologies we have developed over our history, to create even larger impacts through collaborative action.

In its 2022-26 strategic plan (NSF 2022b), the National Science Foundation articulates its vision for

> *A nation that leads the world in science and engineering research and innovation, to the benefit of all, without barriers to participation.*

We feel that with new focus on the activities outlined here, Unidata is well positioned to make significant contributions toward achieving this vision.