An EarthCube Design Process: Unidata's Perspective

October 2011

Ethan Davis, Douglas Dirks,
Linda Miller, and Tom Yoksas

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

Mohan Ramamurthy, Director

1 Unidata's Experience

For over a quarter of a century, UCAR's Unidata Program Center has served a diverse, worldwide geoscience community of researchers and educators. Unidata works to advance data services and tools that help realize the community's vision of *Geoscience at the Speed of Thought*.

In an era of increasing data complexity and volume and need for multidisciplinary integration, the Unidata community has developed and profited from a rich set of data services and tools. The Unidata Program Center leads this community effort by:

- Exploring new technologies, technological standards, and tools that affect the geosciences community
- Supporting tools developed by others and developing tools to meet unmet needs
- Advocating for free and open access to geosciences data on behalf of the community by serving as liaison to data providers
- Providing solutions in new and creative ways
- Working for open standards, interoperability, and open-source approaches

This whitepaper presents Unidata's perspective on the EarthCube project design process, based on our experience providing community-centered cyberinfrastructure.

2 Unidata's EarthCube Vision

Historically, researchers have been forced to spend more time searching for, retrieving, and reformatting data than they spend doing research or sharing knowledge. Unidata's focus has been on bringing the atmospheric science community together to develop and share techniques and resources that remove data discovery, access, and format roadblocks. In the lessons we have learned from these efforts, we see building blocks for a wider-scale EarthCube project that brings together all areas of the Geosciences.

In outlining some of our guiding principles, organizational ideas, and lessons learned, we hope to contribute to shaping a community-driven data infrastructure for the larger EarthCube effort. There are many other important aspects of EarthCube — knowledge management, publication, scientific collaboration processes, peer review mechanisms, *etc.* — that lie outside of our direct experience; we trust that others with more involvement in those areas will provide ideas to complement ours.

2.1 EarthCube Principles

EarthCube should be guided by the following principles:

- **Community-driven, community-based governance**
  EarthCube community members and staff — from across the spectrum of geoscience disciplines — must be encouraged to interact with each other at all levels of the organization.

- **Flexible and evolving designs based on the best ideas**
  The EarthCube community should respond quickly and proactively to changes in the geoscience and technological landscapes.

- **Scalable and loosely-federated organization**
  The EarthCube organization must recognize community members' individual requirements and processes, while providing a framework to bring them together.

- **Preference for existing technologies**
  EarthCube should have a preference for incorporating (and improving) technologies and processes that are already in widespread use.

- **Use of open standards**
  EarthCube should display a strong preference for open standards, formats, and processes, which allow disparate systems to interact.

- **Strong strategic partnerships and collaborations**
  EarthCube should strive to bring members (and other stakeholders) into partnerships in order to maximize development and avoid duplication of effort.

- **National center to provide community leadership**
  EarthCube needs a central, unifying entity with long-term funding to coordinate and bring focus to the efforts of member groups.

We provide additional details on our selection of these principles in section 3, below.

2.2 EarthCube Organization

Unidata envisions the data management component of EarthCube as a conduit that makes geoscience data widely and easily available via standard access and processing mechanisms. We see the project as being composed of a long-lived network of community-governed centers focusing on different aspects of the ongoing endeavor. This network should be composed of:

- A national center that coordinates the efforts of discipline-specific centers

- A small number of existing and new discipline-specific centers that focus on different portions of the EarthCube knowledge management system

Long-term community and sponsor commitment to a sustainable EarthCube network is essential to the success of the project.

The national center should:

- **Lead the EarthCube community effort**
  We have found that a central organizing role is essential in building an effective community effort. A national EarthCube center would serve the larger community by acting as a clearinghouse and information source for EarthCube funding solicitations, by facilitating creation of collaborative proposals, organizing conference sessions and community workshops, building connections within the community, and providing administrative support.

- **Coordinate the efforts of discipline-specific centers**
  Discipline-specific centers will bring their own needs and perspectives to the EarthCube community. A national center will be well-placed to coordinate efforts across centers, provide technological gap analysis, to promote the use of common tools, infrastructure, and standards, and to establish community-based task forces when necessary.

- **Establish strategic partnerships**
  A national center can provide the contacts (and prestige) to help members establish partnerships with external entities in the geoscience enterprise. Partnerships with U.S. government sponsored agencies, international organizations, and private industry will all enhance the value of the EarthCube network.

- **Advocate for the entire EarthCube community**
  A central organization will be in the ideal position to effectively represent the EarthCube community to national agencies and other high-level organizations. A national center would have significant leverage when negotiating for community-wide access to data sources and securing funding for initiatives that are beyond the scope of a single discipline-specific center.

The national center's primary role should be to support the discipline-specific centers in their front-line work on the EarthCube project.

The discipline-specific centers should:

- **Interact directly with and support their respective communities**
  Discipline-specific centers work directly with their own communities to identify unmet needs and propose solutions that solve the problem in the context of the larger EarthCube community.

- **Advocate for their communities**
  Discipline-specific centers have intimate knowledge of their members' work processes and needs. Bringing members ideas and concerns to the larger EarthCube effort will help ensure that tools and processes work across geoscience disciplines.

- **Coordinate efforts of their community members**
  Individual community members may prefer to focus their efforts on smaller projects. The discipline-specific centers can organize these efforts to advance larger EarthCube goals.

- **Provide technology solutions for their communities**
  Discipline-specific centers are well-placed to provide software, technology, and technical support services for their own communities.

2.3 What has worked for Unidata

Our ideas about both national and discipline-specific EarthCube centers are informed by our experience in leading community-building and cyberinfrastructure efforts for the atmospheric science community. Our successes have grown out of the following activities:

- Providing access to and support for existing software tools
- Identifying gaps in capabilities and developing new tools to fill those gaps
- Providing members with direct access to real-time and near-term retrospective data
- Identifying new data sources and advocating on behalf of the community for access to those data
- Developing a community around use of tools and data through workshops and training
- Connecting community members via meetings, workshops, and online interactions
- Providing informed and authoritative support to the community in its use of data and tools
- Providing cyberinfrastructure leadership to guide members in their adoption and use of emerging standards
- Establishing effective strategic partnerships with agencies and private industry
- Periodically re-evaluating our activities through the lens of community governance

Our experience has not been without challenges. Some at the forefront of our thinking now are:

- How best to scale efforts to ever-broadening communities
- How to balance the need to develop new solutions with the need to provide excellent support for existing solutions
- How to do all of these things in an era of resource constraints

3 System Design Principles

We have already listed the general principles Unidata believes should guide the EarthCube project. This section provides a brief rationale for each.

Community-driven, community-based governance

Unidata's experience has been that by collaborating closely with our community, we are able to recognize and respond quickly to community members' changing needs. Community members provide program guidance, technical input, and design ideas, and provide services through their own institutions. Thematic community workshops and meetings provide

members with the opportunity to share ideas and form ongoing professional relationships. Engineers at the Unidata Program Center not only collaborate with community members to develop software, but also serve as the technical support staff for software users. This direct experience with users' technical questions (and frustrations) allows Program Center staff to recognize and correct problems quickly.

Flexible and evolving designs based on the best ideas

Unidata's experience has been that community-based standards — standards developed "from the ground up" and based on processes that are actually in use — are much more likely to garner widespread support and acceptance. We have seen the value of this fluid, community-based approach through our involvement with the netCDF data format, the OPeNDAP data access protocol, the Climate and Forecast metadata conventions, and other software components, interfaces, and protocols. Software and systems built with the involvement of an existing community of users have proven their resilience in the face of long-term change.

Scalable and loosely-federated organization

As Unidata's community has grown beyond our original atmospheric science base, users from other geoscience domains have begun to contribute to and help shape our data-handling and analysis tools. These collaborations, while fruitful, have increased the technical support load placed on Program Center staff. With a national center taking the leadership role in coordinating the activities of the individual domain-focused centers, the individual centers would be able to focus on their own areas of expertise while benefiting from cross-pollination with other domain groups.

Preference for existing technologies

Unidata's practice of adopting and adapting visualization and analysis software (including McIDAS, which shares features with Unidata's IDV, and GEMPAK, which is now transitioning to AWIPS II) for use in the Unidata community has been very successful. We have found it to be easier to urge adoption of new features to existing applications and workflows than to introduce new technologies. Similarly, netCDF and HDF developers have merged their ideas and development efforts together to create data formats useful to a global community. Working to extend and improve existing, proven technologies provides a real-world laboratory for new ideas, and reduces the risk of building software no one wants.

Use of open standards

Unidata has been actively involved in the development of "bottom-up" standards based on real practice in the field. Incorporating open standards and processes into the EarthCube system where appropriate will enhance interoperability, encourage widespread participation, and speed development of additional solutions.

Strong strategic partnerships and collaborations

Unidata's experience in building and maintaining strategic partnerships has been unconditionally positive. Partnerships with agencies such as NOAA's National Weather Service and Earth System Research Laboratory have led to the sharing of data and software tools with the atmospheric science education and research community. Other agency partnerships have led to cooperation on the creation of data conventions and protocols, for the benefit of the entire community. Partnerships with commercial entities including Vaisala

and WSI have made data generated by private industry available to Unidata's community members at no cost.

National center to provide community leadership

The Unidata Program Center's experience addressing the cyberinfrastructure needs of the atmospheric science community leads us to believe that technical leadership is most efficiently exercised by groups with a direct working relationship with their own community members. Despite our appreciation of the merits of "local" leadership at the science-domain level, Unidata believes strongly that it is important to establish a national center to provide long-term stability to the EarthCube project. A national center would play a coordinating role between domain-focused centers and between individual community members and the participating groups.

4 Cyberinfrastructure Principles

Unidata's experience in providing technology solutions for interoperability has highlighted a consistent evolutionary pattern. Simple solutions with simple implementations that satisfy a subset of user requirements transform over time into more complete solutions with more complex implementations. This kind of continuous improvement results in software that is backward compatible for previous uses but faster, more comprehensive, and more capable for new uses. We have found that fostering the evolution of existing working solutions in an incremental fashion results in a flourishing of the fittest technologies.

Problems in the real world are rarely amenable to a single universal solution, but a number of distinct solutions can often be integrated behind a simple interface, providing the experience of uniformity rather than complexity. This is a basic pattern for interoperability solutions — moving the complexity of dealing with diverse representations into the infrastructure, so users interact with what appears to be a single uniform system. The fact that the system is actually implemented as an extensible framework of plug-ins or agents, dealing with multiple cases that must be handled separately, is transparent to the user. Ideally, the user can ignore the infrastructure that delivers data or services entirely, focusing instead on the knowledge to be gained through analysis of the data itself.

Unidata's experience and thoughts on cyberinfrastructure design principles and technology solutions are addressed more completely in a separate EarthCube white paper. Please read *Technology Solutions for Scientific Data Interoperability: Unidata's Perspective* for additional insights into our experience.

5 Governance Principles

While the technical hurdles facing the EarthCube endeavor are enormous, the biggest challenges may be sociological. Broad adoption of and participation in EarthCube will require busy researchers and educators to adopt new methods, augmenting or replacing their current workflows. EarthCube's governance model must recognize and address this natural

resistance to change. It must also work to bridge the cultural differences between disparate disciplines and between developer and user communities.

Unidata has grappled with these issues on a smaller scale for more than 25 years. Our experience with a community-based governance system suggests that strong community participation provides essential feedback mechanisms and helps ensure that the program remains flexible. We recognize the following key principles of community governance:

- **Balanced representation by active community members**
  Governing committee members s are drawn from a wide range of institutions and programs. Individuals who are already active on behalf of their own local communities have proved to be the most effective committee members. Active and informed committee members exhibit a high level of accountability and responsiveness to the community at large.

- **Tight feedback loops on program projects and priorities**
  Active governing committees provide invaluable feedback on the progress of current projects and initiatives, while helping to align current work with long-term priorities. Quick feedback allows the program to be nimble, adjusting priorities as circumstances change and community needs evolve.

- **Direct lines of communication between community, governing bodies, and staff**
  Free communication between representatives of the community and those working to implement the program's projects enhances community cohesion and a sense of shared responsibility for the program's success. Direct connections between program staff and individual community members are equally important for a full understanding of how individual projects are faring in practice.

- **A consensus-oriented and pragmatic approach to decision-making**
  Open, transparent, and inclusive deliberations are a key to building community participation and fostering community adoption of new products and processes. Through their representatives, community members must have real say in program decisions.

Unidata's experience and thoughts on community-based governance are addressed more completely in a separate EarthCube white paper. Please read *Unidata Governance: A Quarter Century of Experience* for additional insights into our experience.

6 Sustainability Principles

In considering how a community-driven EarthCube endeavor might thrive and grow within the global scientific ecosystem, our thoughts are strongly influenced by Unidata's history of more than 25 years. To be successful in the long term, the EarthCube endeavor must:

- **Have strong community buy-in**
  Involvement by a wide range of community members brings a continuing flow of fresh ideas and talent. Strong community participation also serves to help balance the competing needs of disparate sub-communities. The more value individual members

receive from the larger EarthCube community, the stronger their long-term commitment to the community project.

- **Have strong, long-term commitment from funding agencies**
  The EarthCube endeavor will not be finished in five years, or ten, or twenty. While the project may evolve toward the use of alternate funding sources, it is imperative that EarthCube participants be able to embark on initiatives with relatively long time horizons with confidence that successful efforts will receive the support needed to continue.

- **Build on existing organizational infrastructures**
  Creating a new organization is costly. Where possible, locate new centers in existing organizations, allowing the centers to focus on their scientific, technical, and community-building missions rather than on administrative start-up details. Such an approach would allow EarthCube to begin producing gains for community members more quickly. The early phases of the EarthCube endeavor will be critical in bringing communities into the effort, fueling long-term participation.

- **Build mechanisms for international collaboration and cooperation**
  Geoscience is a global endeavor. In order to be maximally relevant to scientific communities in the U.S., EarthCube must be designed to work closely with other similar communities around the world. Global cooperation increases the quality and quantity of resources available to all community projects, strengthening the entire endeavor.

- **Build mechanisms for public- and private-sector collaboration and cooperation**
  Geoscience is not a purely public endeavor. Bringing private-sector ideas, resources, and talent into the mix — while maintaining openness — will add vigor to the EarthCube project.

- **Work to secure additional funding from a variety of sources**
  The National Science Foundation will clearly be the primary sponsor and funding source for EarthCube in the beginning. As the project grows and evolves, however, there must be mechanisms in place that allow for the addition of funding from a variety of public and private sources.

The national and discipline-specific centers should strive to become indispensable to their own communities by doing an excellent job of:

- **Supporting their communities**
  Each center must excel in the task of supporting its own constituents with appropriate discipline-specific data, tools, and techniques.

- **Building their communities**
  Each center must strive to bring community members together, creating strong professional, scientific, and social bonds. Strong individual commitments between community members lead to a stronger, more resilient program.

- **Advocating for their communities**
  Each center must work both to represent their own community's requirements in the larger EarthCube context, but also to bring the larger perspective of the entire EarthCube community to their "home" community members.

7 Definition of Success

The key to the success of the EarthCube endeavor, as we see it:

In much the same way that the Internet has grown from a research network into a household utility, EarthCube must aim from the beginning to evolve into a necessary piece of the world's scientific infrastructure.