**Shared Cyberinfrastructure for Environmental Observatories: Identifying Common Needs, Barriers, and Solutions**

Barbara Minsker and Randy Butler, National Center for Supercomputing Applications (NCSA), University of Illinois, Urbana, IL

The purpose of this white paper is to propose a process for identifying common needs for cyberinfrastructure (CI) across the major NSF environmental observatories (NEON, OOI, WATERS Network, LTER, AON, Earth Scope) and the barriers that could prevent construction of shared infrastructure to address those needs. While there are numerous other observatory projects, most notably in the physics and astronomy disciplines, we believe there is a broad overlap in the needs of the above mentioned environmental observatory (EO) projects that offers the greatest opportunity for cooperation. Several NSF-sponsored workshops in recent years have identified significant overlapping needs among these observatory initiatives, but to date the cyberinfrastructure developments have been proceeding in a piecemeal fashion through multiple projects sponsored by various directorates, with some ad hoc coordination among the projects. Each observatory has been undertaking its own requirements and evaluation process to determine the CI appropriate for its specific needs.

Experience has shown that if efforts are not coordinated, the observatory communities will continue to expend great resources developing their CI in isolation, often with development efforts overlapping other communities, as they discover the same hard fought lessons over and over again. While numerous CI tools and services already exist that can be leveraged, they represent different architectural approaches, typically targeted at a very narrow set of requirements, and are not well integrated with each other. This fragmented approach presents obvious drawbacks, including: 1) requiring each community to attain a sophisticated level of technology expertise and 2) requiring each community to develop software which they then must support. The developed solutions are typically unique to a particular community's requirements, limiting their general applicability, and are typically not interoperable with those developed by other communities.

Thus, without a more formal mechanism for coordination, it seems likely that the observatories will produce redundant and/or incompatible CI, an inefficient use of NSF resources that may not fully meet the environmental community's needs for integration (e.g., to access data from multiple observatories). To avoid this outcome, we propose the following process:

1) Invite a small group (about 20 participants) of CI and domain experts who are engaged in EO planning to a 2-day workshop. Use this white paper as a starting point for discussion. The the first day would focus on overviews of each EO and the major CI challenges they face, identifying the EOs common requirements and needs, and the timelines and deadlines they face in developing their CI. The second day would focus on barriers to implementing common CI and identifying potential solutions to address common CI requirements and overcome the barriers. The box on the next page gives a starting point for the discussion.

2) If there is agreement after the workshop that there are sufficient common requirements to move forward, an unsolicited proposal would be submitted to NSF from the meeting

participants to develop a more detailed blueprint for moving forward. This proposal could include:

a) In-depth assessment of common CI requirements across observatories and the state of CI development to support them.
b) Workshops to develop consensus on common requirements, barriers, and solutions.
c) Development of a cross-observatory CI blueprint through interaction with the workshop participants, NSF, and other organizations as needed. The blueprint would be a 5-year agenda, including what-if scenarios to address risks, plans for addressing those scenarios, and performance metrics.

**Draft Cross-Observatory CI Needs, Barriers, and Solutions**

**Common CI Needs**

a) Standards and inter-operability protocols for key interfaces between and among sensors and CI technologies. This will ease incorporation of new technologies and customization to meet individual observatory and researcher needs.
b) Power control and networking systems for distributed sensor systems.
c) Data storage and management systems that enable the entire environmental community to readily access and integrate data from multiple observatories and numerous other sources (Federal, state, and local agencies, major centers and projects, individual investigators, etc.).
d) Provenance and metadata systems that enable easier sharing and quality control of data and other resources generated from data (e.g., workflows), as well as knowledge networking to support researchers and educators in identifying available resources or collaborators in cyberspace.
e) Workflow systems that support easier integration of data, analytical and visualization tools, and models from diverse sources (including tools and models from commercial, government, and research sources), as well as sharing of these resources with other community members. These systems also need to support event-driven workflows triggered from streaming sensor data.
f) Collaboration and knowledge networking systems that support remote teams analyzing results from i), ii), and iii) above; managing the observatory infrastructure; identifying research needs and available resources and collaborators; etc.
g) High-performance computing to support computationally-intensive modeling, analyses, and visualization, including real-time and on-demand support.
h) Long-term support for a stable infrastructure that integrates all of the above (a *cyberenvironment*), evolves over time to meet new needs of the communities, and provides flexibility for intra-community and individual customization. The role of industry in providing long-term support also needs to be discussed.
i) Common security infrastructure to allow simple but secure authentication and authorization.

**Barriers to Cross-Observatory CI**

a) Each observatory is proceeding with separate funding on individual projects with different time tables. Team members with tight deliverables may be reluctant to rely on technology development performed by other groups, over which they have no control.
b) Supporting cross-project integration is a major undertaking that requires significant time and resources from programs that are already over-taxed.
c) Enabling flexibility to address observatory-specific and individual needs will require community standards and protocols, particularly at the interfaces between technologies to enable "plug-and-play" customization. Creating standards and protocols across such a large community could be a difficult and lengthy process that might slow down infrastructure development.
d) CI technologies are constantly evolving and keeping such a large community trained and supported will be an ongoing challenge.

**Potential Solutions to Overcome the Barriers**

a) Form an environmental observatory consortium, with representatives from all observatories and CI experts serving on the management team, to oversee cross-observatory CI development. The consortium would need sufficient resources to address the barriers noted above.
b) Form standards and protocol teams, perhaps through existing professional organizations, to address ongoing inter-operability requirements.