An Occasional Paper of the NSF Advisory Committee for Environmental Research and Education
AC-ERE 1, May 2003

Environmental CyberInfrastructure (ECI)

Tools for the Study of Complex Environmental Systems

In January 2003, the NSF Advisory Committee on Environmental Research and Education (AC-ERE) published the report, *Complex Environmental Systems: Synthesis for Earth, Life, and Society in the 21st Century* (CES). This report identified cyberinfrastructure as a suite of critical enabling tools and research essential to the study of complex environmental systems. Cyberinfrastructure (CI) provides tools for storing, finding, analyzing, and synthesizing a diverse array of data. Through modeling frameworks, development tools, and hardware, CI supports integrated models of multi-component systems. In addition, CI supports the synthesis of observational data and models, provides mechanisms for collaboration, and offers new opportunities for formal and informal education in environmental science and engineering. Crucial to the effective use of CI are trained personnel and interdisciplinary education. Optimizing the contribution of CI to environmental science and engineering in general and to research on complex environmental systems in particular requires making choices and setting priorities. With this Occasional Paper, the AC-ERE attempts to describe some of the issues and provide more focused advice on environmental cyberinfrastructure (ECI) to NSF.

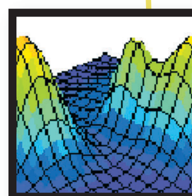
Growing Need for CI Tools

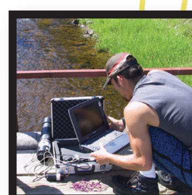
Hardware and software cybertools have been used in environmental science and engineering for decades. An early example is the computer simulation of atmospheric circulation. A more recent example is the numerical simulation of detailed models of the global climate system, which remains limited by available computer resources.

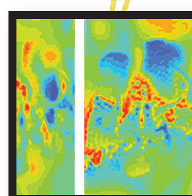
Over the years, NSF has responded to the need for cybertools in many ways and at many levels—mainly driven by basic scientific research. Support has been provided using a variety of programmatic and funding mechanisms. Major activities include general-purpose supercomputer centers, disciplinary computational science centers, developmental research on the middleware and hardware for computational grids, data grids, and networking.

Other research has furthered the development of collaboratory software, visualization tools, data-mining and data management techniques, and the CI components of sensor networks, including environmental observing and analysis systems. Individual investigators and small research groups have benefited from funding for collaborations involving environmental and computational scientists and support for mid-size computational infrastructure. Accomplishments from these investments have been impressive and the current level of NSF investment in these areas is substantial. Nevertheless, recent community assessments have underscored the need for further investment in these and other revolutionary technologies.

Now that some of these cybertools are maturing, the demand for their functionality is growing and there is a need for resources to support more widespread use of these systems as research tools. To address this need, NSF has begun to direct attention to integrating research on and deployment of cybertools across the Foundation. The AC-ERE explicitly supports this emerging focus on domain needs and expresses here some of the particular drivers in environmental research and education (ERE).







Suggested Citation: Futrell, J. and the AC-ERE, *Environmental Cyberinfrastructure (ECI): Tools for the Study of Complex Environmental Systems*, AC-ERE 1, May 2003.

ECI Workshop Describes Trends

In October 2002, participants at an ECI Workshop in Boulder, Colorado (www.ncar.ucar.edu/cyber) identified trends that result in a need for substantially enhanced ECI. Participants drew on their own experience as well as on the CI literature. Among the trends cited was growth in multidisciplinary research on complex, multi-component natural systems. Others include:

- Huge growth in volume of data collected
- Deployment of distributed, intelligent observational and sensor networks
- Rapid expansion in the capability of computer hardware
- Development of software for knowledge representation, discovery, and manipulation
- Development of tools for electronic publication, digital libraries, and virtual collaboration
- Growth of numerical models as tools for investigating processes
- Availability of more sophisticated mathematical and statistical tools, and
- Growing ability to blend models and data.

Taken together, the confluence of these trends represent a rapid evolution in the role of cyberinfrastructure in science and engineering research.

Workshop Recommendations to NSF

ECI Workshop participants identified specific priorities that need to be addressed:

- Data issues, including interoperability of databases, long-term archiving and curation of environmental data, and further development of tools for visualization, analysis, and mining of data
- Training of people in domain science and engineering cyberinfrastructure methods and in collaboration across fields
- Software engineering support
- Network access, including increased capacity and access to data and models
- Need for additional compute cycles in both midsize computational infrastructure and enhanced high-end computing
- Extensive deployment of collaboratory tools
- Ability to exploit ECI tools in education.

The Workshop recommended that NSF address these needs by establishing a crosscutting program to support the development and deployment of ECI for research and education. This approach would help foster collaborative approaches to cyberinfrastructure development, and enhance cooperation between general-purpose NSF supercomputing centers and domain-specific IT centers in environmental science and engineering.

The Workshop supported establishment of a formal collaborative activity to foster interoperability between data systems and models, and to ensure access to data, results, model components, methods, and best practices about implementation and management. Additional recommendations were to conduct workshops on educational and research aspects of ECI, to expand the AccessGrid[1], or similar technology, to encompass multiple nodes at all major universities with environmental programs, and to ensure that NSF-supported ECI activities are coordinated with other federal agencies.

AC-ERE Advice to NSF on ECI Activities

The AC-ERE has reviewed the ECI Workshop proceedings and was briefed by experts from the ECI community in October 2002 (see www.nsf.gov/geo/ere/ereweb/minutes.cfm). These experts described similar needs and supported similar actions. The AC-ERE concurs with the recommendations described above.

[1]A network of computer and multimedia resources to support group-to-group interactions, such as large-scale, distributed meetings.

As NSF undertakes implementation of Workshop recommendations, the AC-ERE suggests that the following features be incorporated:

- A high degree of community participation in planning, developing, and improving ECI using means such as the AC-ERE task group on Environmental Infrastructure, workshops, and national meetings
- A stable yet flexible ECI design that will rapidly accommodate technological changes
- Inclusion of support for both innovative research and basic infrastructure enhancement
- A plan maximizing access for the research community, resource managers, educators, and the general public
- Increased interoperability of data management systems, data analysis tools, and models
- Integrated and transparent archiving of digital data ensuring its retrievability
- Opportunities for collaborative research that involves all disciplines, including computer science and mathematics
- Best practices in cyber-security
- Intellectual property policy that ensures full access to data, tools, and models
- Open systems and the free and open exchange of data, software and results.

The AC-ERE looks forward to being a full partner in addressing the critical needs for ECI in the future.

AC-ERE Comments on Implementation

The following five areas are critical in the early implementation of an ECI strategy.

Data Management. Environmental researchers are often constrained by the limited capacity available to integrate and analyze many kinds of data, which often span a wide range of space and/or time scales, and to synthesize a high volume of data in near real-time.

The ERE community needs to move toward a single logical confederation of data systems. There are a number of ways to accomplish this goal, all of which would be enhanced by partnerships with other agencies and steering committees that include members of the academic community, resource managers, and non-academic users.

The NSF should develop an appropriate framework, such as an environmental data systems confederation, to (a) track research on data management techniques in the academic and commercial arenas, (b) study the lessons learned and technology developed in ERE data system efforts, (c) develop and publish "best practices" methods that improve interoperability of ERE data systems and integration with the larger data universe, and (d) assist data systems that adopt these practices. A complementary activity might provide support to ERE data management projects for software engineers who would be the "glue" between such projects.

Another suggestion is to consider support for broad, long-term data management activities. These structures should provide archiving services for data systems established by individuals, small groups of investigators, or finite lifetime research projects. They might also be appropriate test beds for new data and knowledge analysis software. Development of an inter-agency permanent digital data-archiving capability should be pursued.

IT Resources. Many systems required for ECI are too costly to be supported by individual grants, but fall short of the size of supercomputer centers. This problem might be addressed by a specific program for computational infrastructure. Some examples of needs that fall in this category are mid-size computational hardware and operations personnel, departmental networking upgrades, and visualization and collaboration hardware.

In addition, there should be a small number of environmental synthesis centers for ERE that would include high-end computing capabilities. Researchers at the centers would collaborate with the external ERE community. The centers could offer synthesis and research workshops, a long-term visitor program, a postdoctoral program, and maintain substantial high-end computing resources for use by both external and internal researchers.

Computational Environmental Research. It is desirable to foster partnerships among environmental computational science specialists, the computer science community, and the NSF Directorate of Computer Information Science and Engineering (CISE). One possible mechanism is to establish a long-term interdisciplinary program to fund collaborative research in computational environmental research to encourage collaborations between the environmental and computer science communities.

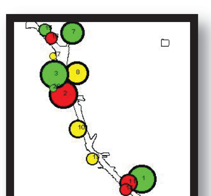
Another computational issue is to support research-enabling software infrastructure projects, such as development of common modeling frameworks and models that can be used by a broad community.

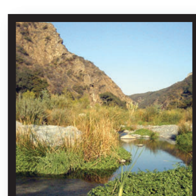
Collaborative Environmental Research. Deploying new tools that facilitate collaboration among scientists and engineers working in environmental areas is also essential to the advancement of environmental research. This will require increased support for collaboratory software and hardware, as well as for community workshops and other activities that encourage collaborative approaches to investigation of environmental research frontiers.

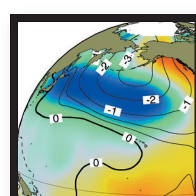
Education and Training. Enhancement of ECI ultimately depends on the development of a capable workforce. Programs for interdisciplinary graduate training that integrate computer science with environmental science and engineering are a particularly attractive option for developing expertise in a relatively short timeframe. Another approach is to emphasize projects that use advanced IT tools in formal and informal environmental research and education. As one example, digital library activities in ERE have already begun to prove valuable and should continue to be expanded. Inclusion of education and outreach components in large ECI activities is strongly encouraged.

Summary

Increased ECI capacity is essential to achieve the goals of environmental synthesis: asking integrated interdisciplinary research questions; creating diverse teams; assessing complex systems across spatial and temporal scales; and providing broad access to knowledge and know how. NSF should take the lead in providing researchers across the full range of disciplines—social, biological, and physical sciences and engineering—with the critical cybertools they need to advance environmental research and education in the 21st Century.







www.nsf.gov/ere