

## CI4EI: How Can Cyberinfrastructure Help the Earth Institute Save the World?

The Earth Institute (EI) has successfully used the Internet to reach a wide audience in the scientific and policy communities and the public at large. The EI has also employed a range of information technologies to facilitate its work, ranging from real-time data collection, analysis, and visualization from ships around the world to development of large community databases to support international science.

Unfortunately, the EI does not at present have a coherent strategy for developing and using cyberinfrastructure (CI) to carry out its mission. Efforts to use the Internet to improve outreach and impact are scattered and lack the critical mass to have a significant, long-term effect. Efforts to bring new technologies into the EI's research agenda remain focused on "small science" rather than "big science," with the exception of a few traditional areas of research such as LDEO's new ship. The EI and Columbia University as a whole have not had much of a track record in large scale computing, and have not therefore benefited from the large CI investments made in the past by the National Science Foundation (NSF) and other agencies. However, the National Science Board and NSF have recently articulated a new CI strategic vision that does embrace a much broader concept of CI that emphasizes the importance of data and data management and virtual networks and organizations. These are areas where the EI could in fact have a strong competitive advantage.

More specifically, version 7.1 of the draft *Cyberinfrastructure Vision for 21<sup>st</sup> Century Discovery* (<http://www.nsf.gov/od/oci/ci-v7.pdf>) highlights NSF's new commitment to "Provide the science and engineering communities with access to world-class CI tools and services, including those focused on: high performance computing; data, data analysis and visualization; cyber-services and virtual organizations; and, learning and workforce development." Notably, although high performance computing remains a high priority, the report emphasizes additional goals including:

- 1) Support the continued development, expansion, hardening and maintenance of end-to-end software systems – user interfaces, workflow engines, science and engineering applications, data management, analysis and visualization tools, collaborative tools, and other software integrated into complete science and engineering systems via middleware – to bring the full power of a national cyberinfrastructure to communities of scientists and engineers.
- 2) Support the development of the computing professionals, interdisciplinary teams, enabling policies and procedures, and new organizational structures such as virtual organizations, needed to achieve the scientific breakthroughs made possible by advanced CI, paying particular attention to the opportunities to broaden the participation of underrepresented groups.
- 3) Support state-of-the-art innovation in data management and distribution systems, including digital libraries and educational environments that are expected to contribute to many of the scientific breakthroughs of the 21st century.
- 4) Support the development and maintenance of the increasingly sophisticated applications needed to achieve the scientific goals of research and education communities.

## For Columbia University Use Only

- 5) Provide a framework that will sustain reliable, stable resources and services while enabling the integration of new technologies and research developments with a minimum of disruption to users.

In a recent virtual presentation at Lamont during the ESIP Federation meeting hosted by CIESIN, Margaret Leinen, the Assistant Director for Geosciences, pointed out that NSF is spending on the order of \$500 million in FY06 on CI, of which one quarter was through the Office of CyberInfrastructure (OCI), now headed by Dan Atkins. The OCI budget is expected to grow from \$127 million in FY06 to \$182 million in FY07 (requested). She also mentioned the establishment of a new digital data interagency working group under the auspices of the National Science and Technology Council, which has already started to meet.

Internationally, there is also strong interest in CI-related initiatives under the rubrics of grid computing and e-science. The UK, the European Union, UNESCO, and other organizations and countries have established or are discussing new programs and investments. A recent visitor from the Chinese Academy of Sciences indicated that China is planning to invest on the order of US\$250 million per year in its computing and data management infrastructure over the next few years. Related activities include: the Global Earth Observing System of Systems (GEOSS), which is focusing on improving the interoperability and usability of Earth observation data from around the world; the International Polar Year (IPY), which is developing an associated Data and Information Service; and the electronic Geophysical Year (eGY), which is attempting to develop virtual observatories and other infrastructure for the international geoscience community.

It is therefore vital to the EI's research enterprise to begin discussing how it can better plan for and utilize CI to remain at the state-of-the-art in science. Equally important is discussion of how CI can facilitate more effective interdisciplinary research and education, and the transition from science to applications and decision making.

### *Planned Cross-Cutting Activity*

This cross-cutting activity is aimed primarily at *starting* the conversation and collaboration that will be needed to develop an effective EI-wide CI strategy and to launch the new partnerships that will be needed to compete successfully for CI support. We do not believe that a CI strategy developed "top down" will ultimately be successful, unless coupled with significant grassroots input from the broad EI and Columbia-wide community. Therefore, with the EI's support, we have initiated an initial effort to jumpstart discussion of both strategic and technical issues, to develop an involved and informed community, and to provide the basis for strategic planning.

To launch this activity, CIESIN invites all EI centers, EI-affiliated departments, the Columbia libraries, and affiliated organizations and programs (e.g., GISS, Conservation Medicine, and CERC) to nominate one representative to serve on an overall steering committee and to suggest other participants in the effort.<sup>1</sup> Specific tasks for this steering committee are to identify EI-wide needs for CI support, inventory relevant Columbia activities and resources, review various opportunities for funding, suggest needed internal and external partnerships, and encourage proposals to specific RFPs or funding sources. This steering committee will also designate a

---

<sup>1</sup> It may be appropriate for larger groups such as LDEO to nominate more than one representative.

## **For Columbia University Use Only**

person or subcommittee to interface with the Development office on potential funding issues and opportunities, including development of sources of matching funds for external grants and provision of inputs on facility needs. To ensure the effectiveness of the committee, CIESIN will chair the committee and provide a rapporteur for all meetings, who will ensure preparation and dissemination of meeting notes and action items. CIESIN will seek volunteers to host a wiki, web site, and other tools to facilitate discussion and information exchange, or will provide these itself if necessary.

The steering committee will also oversee the development of a larger, more informal community to encourage scientific and technical discussions, interactions, and “brainstorming” across the EI and the University at large. It will begin organizing seminars, training sessions, and other activities to encourage greater awareness of CI issues, educate the broader community about CI-related advances, encourage and coordinate training on new CI tools and techniques, obtain inputs on EI needs for CI, and in general try to establish a working CI community on campus. One specific set of activities will be a set of seminars with external speakers, probably two in the spring semester of 2007 and two in the fall. The steering committee will solicit suggestions for speakers and coordinate the invitations.

Two major outputs of the activity will be:

1. Identification of CI needs across the EI that could significantly help the EI carry out its mission, including hardware, software, personnel, training, data, tools, and partnerships;
2. Identification and evaluation of CI opportunities, e.g., major agency, foundation, and industry programs and potential sponsors.

This activity will not attempt to choose among strategic approaches, but will help identify and evaluate key needs, alternatives and opportunities that would inform strategic planning. We also expect the activity to lead to new partnerships and proposals for funding and to help the EI position itself when NSF begins to implement its new CI strategy.