7 things you should know about...
Cyberinfrastructure

Scenario

When Kara started her junior year in college, her biology professor invited her to participate in a multi-institutional, nationwide research initiative to study the effects of groundwater pollutants, specifically looking for connections with certain forms of cancer. Supported through a consortium of government agencies and higher education institutions, the project was designed to collect data from a range of sources and look for patterns and correlations that could guide federal environmental regulation. A complex, distributed network of remote sensors, archived data, historical records, and analytical tools would be available to the collaborators—federal policy officials and academics from disciplines including medicine, agriculture, and climatology.

Previous studies had shown a strong correlation between certain pollutants in the drinking water of children and later onset of several forms of cancer. Several hundred physicians were recruited from across the United States to submit patient data along with histories about where those patients grew up. Agricultural organizations provided enormous amounts of data about fertilizers, pesticides, soilborne pathogens, and crop and insect diseases. Soil scientists and climatologists analyzed factors that affect the distribution of contaminants in the soil and water. For her part, Kara and her biology advisor correlated historical concentrations of pollutants in groundwater and the incidence of cancer among the people studied. On an ongoing basis, they submitted findings to the project, where all other participants could see them and provide feedback.

Even though Kara attended a relatively small undergraduate institution, she had access to data and analytical tools at research universities and government labs. She and her professor collaborated with a distributed network of professional scientists and graduate students in a broad range of academic disciplines. By the time she began her senior year, Kara had had the opportunity to participate in a large-scale project that she believed would produce important results and recommendations. She had also been able to work closely with biologists, chemists, policymakers, and individuals with a wide range of skills and interests. The experience helped her understand the interconnectedness of environmental and social systems, and it helped her decide how she wanted to focus her senior-year studies and plan for graduate school.

1 What is it?

Cyberinfrastructure (CI) is not a new technology, per se, or merely a better, faster Internet. CI merges technology, data, and human resources into a seamless whole. While processors, storage devices, sensors, and other physical assets are part of CI, it is more than connecting people with advanced networks and sophisticated applications running on powerful computer systems—it is involving those people as participants in the generation of knowledge, giving them the opportunity to share expertise, tools, and facilities. Arden Bement, director of the NSF, described CI as "the engine of change" for the "second revolution in information technology." The heart of that engine, he says, are communities that support peer-to-peer collaboration and new modes of research and education.[1]

Computation has become an essential tool for research. Technology is essential in understanding complex systems such as biometrics, the environment, and the brain. CI—which is known as e-research, e-science, and e-infrastructure in Europe, Australia, and Asia—brings together high-performance computing, remote sensors, large data sets, middleware, and sophisticated applications (modeling, simulation, visualization). Beyond the technology, an essential part of CI is the way it allows distributed teams to turn "flops, bytes, and bits into scientific breakthroughs."[2]

2 Who's doing it?

Faculty, students (K–12 and higher education), staff, and everyday "citizen users" participate in CI. Although CI is often assumed to focus only on science, it also applies to economics, social sciences, and humanities. Researchers use CI to investigate complex, multidisciplinary problems. In the Mixed Apparatus for Radar Investigation of Atmospheric Cosmic-Rays of High Ionization project, high school students build devices that detect cosmic rays in the atmosphere and connect those devices to the larger CI. Not only are the students helping collect valuable data, they become part of the virtual organization focused on this research.

Other projects, such as the National Ecological Observatory Network (NEON), which bills itself as a "continental-scale research platform," recruit lay persons to collect scientific data. NEON consists of a network of instruments deployed across the United States to measure ecological variables—including air and water quality, soil characteristics, and climatic conditions—and monitor changes in numerous plants and animal species. NEON also relies on observational data collected from individuals around the coun-

more ⇨

EDUCAUSE Learning Initiative

www.educause.edu/eli

try. Along with data feeds from observatories and remote sensors, this "citizen" information allows scientists to better understand the ecological impact of climate change, shifts in land use, and the effects of non-native species.

3 How does it work?

CI depends on a technical infrastructure that knits together high-speed networks with high-performance, high-availability, and high-reliability computational resources. Management systems control the usage, performance, and availability of computing systems, while security systems protect these assets. Data can be large aggregations of previously collected data (from radio astronomy, for example) or live feeds from remote sensors located anywhere in the world. Many of the systems are housed in different locations, and experiments typically run on "virtual machines" in which spare cycles from dozens or hundreds of computers are used for a single task. Data sets can be distributed as well, with neurological data coming from one university and physiological data from another.

4 Why is it significant?

According to the NSF (http://www.nsf.gov/od/oci/ci_v5.pdf), CI provides an opportunity for a new kind of scholarly inquiry and education, empowering communities of researchers to innovate and revolutionize what they do, how they do it, and who participates. CI has been linked with the notion of "open notebook science," in which data are posted online as they are collected, facilitating real-time, distributed science. By providing access to research before it is published, this data sharing has the potential to accelerate scholarship. Data that are collected, archived, and analyzed are on a scale that was previously unimaginable; CI tackles this mountain of information, allowing researchers to answer questions that could hardly be asked a decade ago. It also allows those previously unable to join in leading-edge scientific research to participate to learn by doing, not just by listening. A project at the Guana Tolomato Matanzas National Estuarine Research Reserve, for example, provides public, online access to grid-enabled simulators that model water quality. Disciplinary and geographic barriers are being removed as new tools bring together scientists to conduct research never before possible. For example, CI may transform the study of language by making data about all the world's languages accessible to researchers. Linguists could map out the structures of the world's languages in great detail. This linguistic "cognomen" could transform the linguistic sciences in a way similar to how the human genome project has transformed the biological sciences.[3]

5 What are the downsides?

Because of the scale of many current and proposed CI projects, some believe that the opportunities are limited to large, well-funded research institutions. Individuals at smaller institutions assume they lack the infrastructure to participate, even though many can contribute to and benefit from such projects. A focus on the physical assets necessary for CI fails to adequately characterize the importance of people. Moreover, as with any large-scale, distributed-technology project, security concerns arise for data, instruments, and applications and must be adequately addressed to promote participation.

6 Where is it going?

CI is in a relatively early stage of development. Research is under way to improve its capacity, reliability, and management. While its origins are in research and science, CI is increasingly being applied in the humanities and arts, as well as in education. As the awareness of educational applications grows, more students will have opportunities to learn through participation. CI is becoming a global initiative. As the NSF's Bement notes, many large research facilities are too costly for individual institutions (or nations) to develop. CI enables increased international collaboration in terms of virtual organizations, sharing of data, and access to tools. With CI, these tools become community resources, on a national and global scale.

7 What are the implications for teaching and learning?

CI opens new doors for faculty to involve students in active, hands-on learning. CI helps close the gap between smaller institutions and larger campuses, providing the opportunity to engage with data and systems in ways not previously possible. As participants in virtual teams, students may begin with relatively minor roles, but as they gain experience and knowledge, they can take on additional responsibility. CI offers students earlier access to the "tools of the trade" than would be possible otherwise, and as these tools are themselves improved through CI, students can participate in that development. In one project, for example, students interested in nanotechnology can run simulations that mirror what researchers do and can construct their own experiments.

Access to instruments, data sets, and experiments provides opportunities for faculty to model disciplinary thinking for students, who can participate in the interpretation of data and ask new questions. Many students are motivated by participation in communities focused on complex problems, such as the environment. These communities of interest, along with real data, computational resources, and visualization tools, are highly engaging, allowing students to learn by doing. Learners are challenged to develop new literacies, new ways of thinking, new communities, and new kinds of knowledge.

---

1 Arden Bement, "Cyberinfrastructure: The Second Revolution," *Chronicle of Higher Education*, January 3, 2007, 53(18) B5.

2 NSF's Cyberinfrastructure vision for 21st century discovery, <http://www.nsf.gov/od/oci/ci_v5.pdf>.

3 White paper: SBE cyberinfrastructure portfolio, lessons, challenge and priorities. From ACCI meeting, October 31, 2006.

EDUCAUSE Learning Initiative
advancing learning through IT innovation
www.educause.edu/eli

August 2007