EDUCAUSE Center for Applied Research

Research Bulletin

Volume 2006, Issue 3

January 31, 2006

IT and the Changing Face of Research in Higher Education

Stephen L. Daigle, California State University

Brian D. Voss, Louisiana State University

4772 Walnut Street, Suite 206 ◆ Boulder, Colorado 80301-2538 ◆ www.educause.edu/ecar/

Overview

Historically, scholarly research in higher education has been characterized by an individualistic and discipline-centric focus involving minimal collaboration. Increasingly, contemporary academic research emphasizes collaborative teams working across disciplinary and, more frequently, institutional boundaries. Academic research is also becoming more dependent on computing and networking infrastructures and related enabling technologies, fostering scientific study that is independent of space and time.

As a consequence, higher education policymakers and IT planners find themselves in the midst of a profound change in how scholarly research is conducted on campus. They must begin, if they have not already done so, to build campus IT infrastructures that both encourage and enable 21st-century scientific research. This requires understanding and appreciating the nature of contemporary research and its implications for acquiring and using diverse enabling technologies and associated computing resources at the state, regional, national, and even international levels.

In many respects, the changing face of research in higher education is something that chief information officers (CIOs) and IT planners at every major research university recognize: a classic elephant-in-the-room situation. But becoming part of, accessing, and using various elements of the national cyberinfrastructure is not a straightforward process. This process requires insight into the nature of current academic research and how it is driven, funded, and accomplished. We must also must recognize that computer-aided research has now joined with theory and observation as one of the three primary accepted methodologies in the scientific pursuit of new knowledge.

This research bulletin provides policymakers with an introduction to the evolving national cyberinfrastructure, discusses the role that new networking and enabling technologies play in integrating institutional research capabilities, and describes the impact these technologies may have on today’s higher education research and on campus IT planning.

Highlights of the Changing Face of Research

In the context of computing requirements, higher education researchers have always been major consumers of processing power, high-capacity data storage and archiving facilities, and high bandwidth for data transmission. Until 1985, IT resources for scientific research were located in university computing centers; after 1985, the National Science Foundation (NSF) funded additional geographically dispersed supercomputing centers to support very high-end research and began laying the foundations for high-speed networks to connect them to their university research clients.[1] Cummings and Kiesler noted one important IT trend that is changing the face of research:

> Today, dispersed collaborations are more feasible because communication technologies allow scientists to exchange news, data, reports, equipment, instruments, and other resources.[2]

Networks, networking architectures, and various emerging technologies that support these new research environments increasingly are considered fundamental components of contemporary academic research.

As scientific inquiry probes deeper and research problems become more complex, they require diverse disciplinary teams comprising members who, for both personal and professional reasons, work in various research locations around the country and the world. Collectively, these distributed research teams seek to use as many IT resources as possible, including

- a host of grid-mediated computational resources, including large parallel processors, Linux clusters, and specialty processor systems;
- transmission, processing, storage, and archiving of very large data sets stored at diverse locations and joined using advanced middleware software;
- widely distributed input-sensor instrumentation devices;
- a diverse range of advanced scientific visualization and interactive collaboration spaces; and
- dedicated high-bandwidth research and education networks.

Increasingly, traditional limits to research, such as distance, time, and institutional resources, are dissolving. An egalitarian ethos is spreading throughout the world's high-end research community of scholars—one grounded in using and participating in the national cyberinfrastructure.

The Evolving National Cyberinfrastructure

Since the publication of *Revolutionizing Science and Engineering Through Cyberinfrastructure: Report of the National Science Foundation Blue Ribbon Advisory Panel on Cyberinfrastructure*, also known as the Atkins Report,[3] the vision of an evolving, broadly based, ubiquitous cyberinfrastructure of interrelated information technologies has continued to grow and mature. The term *cyberinfrastructure* connotes more than just hardware and software, more than bigger computer boxes and faster networks connecting them. The term encompasses new "research environments" in which disciplinary experts, on interdisciplinary teams, supported by specialized computational support staff, have global, instantaneous access to enormous computing and networking resources. In terms of contemporary scientific research, these IT resources are viewed as integral rather than ancillary to today's research agendas. Researchers can now tap into a broad array of cyberinfrastructure components, and they can benefit from opportunities that emerge from a national commitment to the expanded social role of university research and education.

Seamless access to these resources via advanced and dedicated networks is allowing previously unimaginable capabilities to be brought to bear in the advancement of nearly all disciplines—medicine, the sciences, the arts, and even the humanities. One key aspect of the evolving cyberinfrastructure is the capability of regional, national, and global networks to create virtually unlimited connectivity between researchers and the resources they need to pursue their work—the new research environments.

One example of this connectivity is the Electronic Visualization Laboratory (EVL) at the University of Illinois at Chicago. In November 2004, EVL acquired a 3,200 mile, 10 gigabit-per-second (Gbps) wavelength on the National LambdaRail (NLR) network infrastructure running from Chicago to San Diego. Tom DeFanti, EVL's director, explains that this dedicated wavelength, called CAVEwave,

> provides researchers with a deterministic network, with guaranteed bandwidth, schedulable times, and known latency characteristics, in order to understand requirements for the real-time visualization, analysis, and correlation of terabytes and petabytes of data from multiple storage sites. [We have] all this bandwidth, which supplements our existing network infrastructure, for less than the cost of a 32-node cluster at each end![4]

Another excellent example of how networked resources are being used both to advance science and to serve social needs is NASA's finite-volume General Circulation Model (fvGCM) project. Running on NASA's 10,240-processor supercomputer system (Columbia) and developed by a diverse team of researchers at several NASA facilities, the fvGCM produces real-time numerical weather predictions targeted at improving hurricane tracking and intensity forecasts.[5] The team accurately predicted landfall and storm intensity with an advance warning of three to five days based on simulation results for several of the 2004-season hurricanes (Frances, Ivan, and Jeanne). This and similar climate and earth research projects are under way, bringing to bear resources from universities, government agencies, and labs (NASA, NOAA), and high-performance computing centers and laboratories around the United States and the world.

Network Lambdas

Many have suggested that the future of such compute- and data-intensive research lies in the creation of optical networking "clear channels," known as lambdas, between campuses, throughout states and regions, across the nation, and around the globe. These lambdas can be dedicated to an individual campus researcher working globally in collaborative research efforts with diverse and distributed partners.

These developments require renewed attention to and emphasis on networking architectures and the associated campus infrastructure. Researchers need (and increasingly expect) campus-based access to the cross-continent, high-speed arteries being built by organizations such as NLR and Internet2. Some states have their own optical fiber infrastructure projects (referred to as regional optical networks, or RONs) either completed, under construction, or in advanced planning phases. But the "last mile" question looms: when those access nodes are driven to the campus border, what is the process to get them to the resources and researchers on campus so they can be connected to the national and international high-speed networks supporting advanced research?

Science Drives the Applications Layer

New applications to meet scientific needs form yet another layer of the cyberinfrastructure—a layer pushed and pulled by contemporary science that demands improvements in critical research-support technologies. Modern compute- and data-intensive research drives the development and dispersion of these applications to support an enormous variety of scholarly endeavors. As research into physical, geopolitical, sociological, and humanistic problems becomes more complex, demands on the computing and network infrastructures increase.

Often, the processing and storage resources of a single supercomputer center are not sufficient, and accessing it may be impractical. Yet geographically dispersed researchers collaborating on a project still need to share enormous data sets, programming codes, and other materials supporting their research. By assembling grids of large computers linked by advanced networks (optical, dark fiber, and so forth), computational power can be generated, transmitted, and distributed much like a power grid shares electricity among consumers.

As Internet2 President Douglas Van Houweling has noted, however, data-intensive research also can be considered disruptive, since it can consume as much shared bandwidth as available and “break the network.” For example, such research may include

- real-time access by physicists to particle collisions at CERN, Fermilab, and elsewhere that require 6–7 gigabit throughput;
- access to pathology tissue banks for telemedicine, requiring gigabit speeds per simultaneous user; and
- access to data from distributed radio telescopes, microscopes, and other high-performance instruments.[6]

Still, opportunities for conducting such science over very high-speed optical networks are increasing. Projects such as Internet2’s Hybrid Optical and Packet Infrastructure (HOPI) are helping to create a national network test-bed infrastructure that can serve as a model for the next generation of research network architectures to succeed Abilene and other existing very high-speed networks. These hybrid networks will enable researchers and scientists from around the world to experiment with dynamically provisioned bandwidth, new transport protocols, and circuit-switched environments.

For example, NSF’s OptlPuter project is an “envisioned infrastructure that will tightly couple computational resources over parallel optical networks using the Internet protocol communication mechanism.”[7] OptIPuter is a five-year research program led by the University of California, San Diego, and the University of Illinois at Chicago with partners at San Diego State University; the University of Southern California’s Information Sciences Institute; the University of California, Irvine; Northwestern University; the San Diego Supercomputing Center; and other TeraGrid sites. The initial deployment consists of 6 Gigabit Ethernet waves providing dedicated network paths in support of data-intensive scientific research. The project has two application drivers—bioscience and geoscience—where scientists are generating multigigabytes of three-dimensional data objects residing on distributed archives that can be correlated, analyzed, and visualized in an interactive and near real-time environment. The OptIPuter represents a revolution in architecture in the sense of being a virtual parallel computer where the individual processors, memory, and peripherals are all widely distributed but connected through standard Internet protocol delivered over multiple dedicated waves. In effect, optical networks replace computers in this model, creating a world of supernetworks.

While these examples are focused on the hard sciences, the technology enables advancement across a wide variety of other disciplines. Choreographers use visualization and simulation techniques to model and teach dance; orchestral performers conduct master classes across great distances; stage designers use these same tools to model and envision lighting for theater productions; and literature scholars use algorithms to conduct content analyses of texts long believed to have been “mined out.” Geographers are experiencing an intellectual renaissance, having incorporated remote sensing, global positioning systems (GPS), and other data-intensive techniques into their research practice.[8] And both scientific discovery and social policy are being served through projects such as the Network for Earthquake Engineering Simulation and the Human Genome Project. As converged technologies become more ubiquitous, on-demand, embedded, and mobile, the opportunities for conducting such research will extend to virtually every discipline, moving far beyond the confines of traditional research labs into what George Gilder calls a “network-centric culture.”[9]

Networks: Connecting the Pieces

In his Supercomputing 2004 keynote address, Tom West, chief executive officer of National LambdaRail, Inc., referred to network planners as plumbers—people who are trying to push ever greater amounts of information, at faster speeds, through some sort of pipe.[10] Increasingly, those “pipes” are fiber-optic cables using pulses of light—wavelengths—as the transmission source. Theoretically, there is almost no limit to the amount of data, voice, and video information that can be delivered over a fiber-optic network. For example, a single fiber cable using Dense Wavelength Division Multiplexing (DWDM) can be divided to serve multiple purposes, all with different bandwidth capacities and transmission speeds. In addition, large installations of dark fiber around the nation have unused potential in terms of capacity for meeting the ever-increasing needs for data transmission and interconnectivity.

The bottom line is that the flexibility of fiber-optic and networking media such as dark fiber can yield important dividends. For example, part of a network can be dedicated to routine operational functions of a university; other parts can be allocated to researchers studying the network itself, and still others can be set aside for high-end scientific applications employing very large data sets. Finally, the network can be partitioned for on-demand, real-time collaborative research at multiple locations.

Still, with all of the developments in network technology, requirements of leading-edge applications (such as those associated with “grand challenge” science) will require thoughtful partitioning of this resource. Grid-enabled computing, an emerging network-centric component, is now viewed as an inevitable next step in further enabling contemporary research.

Grid Computing

Grid computing is essentially a form of networking that seeks to harness the diverse elements of major, national supercomputing centers and join them into a network-mediated virtual supercomputing center (a metacenter). Grid computing brings together geographically dispersed resources such as databases, scientific instruments, computer hardware and software, sensors, storage devices, visualization facilities, and researchers themselves, often in real-time interaction. It requires coordinated sharing and brokering of resources, enhances collaborative problem solving, and enables the formation of dynamic virtual communities of scholars.

The strength of grid computing lies in its ability to allow users to invoke resources (compute engines, instruments, display devices, collaborative spaces, and so forth) from any location on the grid at any time. This “when-and-where-needed” capability of the grid facilitates investigations into today’s very complex scientific issues such as the human genome, brain science, climate change, and high-energy physics. A grid is not a new concept; the principles behind it are the same as those used for delivering utilities such as water, electricity, and phone service, and scientists and researchers ought to be able to access grid resources as seamlessly as these more commonly understood grid services.

Grid technology is based on high bandwidth, together with an acceptance of common access and use standards, that make a “computing utility” feasible; however, given the number of resources and the potentially large number of people involved, grid computing introduces a host of security and policy issues far beyond those encountered on a single campus. Trust, risk, and cost are still major considerations for researchers using grid-based resources, particularly as the concepts, technologies, and management issues associated with grids continue to evolve.

In an attempt to probe these overarching issues, NSF has funded an ambitious national project combining fiber-optic networks with grid computing environments: the TeraGrid. This $98-million, multiyear effort seeks to build and deploy the world’s largest, fastest, distributed infrastructure for open scientific research.[11] Currently, eight partner sites form the TeraGrid, which provides

- more than 40 teraflops of computing power (consisting of commodity cluster technology as well as specialty computational systems);
- storage facilities capable of managing more than 2 petabytes of spinning data (disk); and
- specialized data analysis and visualization resources, all interconnected at 10–30 gigabits/second via a dedicated national network.

The geographically dispersed partners represent a diverse cross section of scientific institutions. The TeraGrid seeks to revolutionize collaborative science by giving researchers the bandwidth they need to access a grid of advanced IT resources for performing large-scale data analyses across geographic locations. Among the first wave of researchers to use the TeraGrid are scientists studying the evolution of the universe, contaminated groundwater cleanup, seismic simulations, and biomolecular dynamics.

Merging the capabilities of advanced networks and grid computing requires yet a third key element to building the infrastructure for 21st-century science: middleware.

Middleware

The NSF's Middleware Initiative, or NMI, was launched in 2001 with $12 million in grants to be distributed over three years. Well before the end of the funding, NMI projects permeated U.S. universities and linked American scientists and researchers with software developers and scientists in the United Kingdom, Europe, and Asia.[12]

Middleware is best described as specialized software that connects two or more otherwise separate applications across the Internet or local area networks. Specifically, the term refers to an evolving layer of software services that resides between the network and more traditional software applications for managing computing activities such as security, authentication, user access, and information exchange. Middleware is viewed as critical for managing grid-based technology resources spread across multiple locations by transparently handling the underlying details for sharing resources in an integrated fashion, allowing researchers to focus on the substance of their work.

An example of such middleware, often referred to as the Common System Software Stack (CSSS), seeks to remove the onerous tasks associated with managing directory services, user authentication and authorization, network connections, financial accountability and control, security, and all of the software needed for sharing computation, storage, and transmission resources.

One incarnation of a CSSS—the Globus Toolkit—supports computer grids that allow users to securely share computational power, databases, and tools across networks worldwide while maintaining local autonomy. The toolkit includes software for security, information infrastructure, resource management, data management, communication, fault detection, and portability.[13]

What It Means to Higher Education

The evolving cyberinfrastructure consisting of high-capacity dedicated networks, middleware, and grid-based resources is altering both the kind and quality of academic research in higher education. This infrastructure is stimulating collaborative research endeavors that draw on diverse and dispersed partners across national and international communities. As funding competition increases from government and private sources, so does the need to broaden and strengthen research alliances beyond the campus boundary. Institutional research organizations are witnessing a decline in the success rate for grant submissions; the rate of increase in the number of quality proposals exceeds the capacity for funding resources by a growing amount. As a result, successful funding rates for NSF programs have dropped dramatically in the past four years—from

33 percent in 2000 to 24 percent in 2004; the success rate for competitive, merit reviewed processes was just 22 percent. Worse still, the rate of decline has accelerated, as some key research directorates such as Computer and Information Science and Engineering were able to fund just 16 percent of the proposals received.[14] Leveraging an institution's ability to meaningfully participate in the cyberinfrastructure may be a way (perhaps the only way) to create partnerships capable of successfully obtaining the funding necessary to advance higher education research agendas.

The need to participate in the cyberinfrastructure is not limited to the sciences. Scholars in the humanities and social sciences also need remote access to large data sets, instruments, and archives. A need for collaboration and the penetration of cyberinfrastructure development is felt pervasively throughout higher education, not only in research but also in the fulfillment of the institution's teaching, learning, and service missions. Taken together, these factors contribute to the democratization of research, along with a potential blurring of lines between amateur versus professional and commercial versus public. Traditional boundaries between academic disciplines and political entities are softening.

Within the framework of IT support for research, IT planners are increasingly focused on campus researchers who are collaborating with others around the world. The areas requiring IT support services, which until now have been limited primarily to campus-based resources or access to specific supercomputers via dedicated networks, require expertise in portals and common software stacks, as well as broader involvement in regional, national, and international high-performance computing and networking initiatives. Key issues involve network architectures for the campus last mile, as well as the overall campus IT environment required for balancing resource distribution and network engineering. Some important governance issues are

- the location of data storage and computation resources on campus (more easily centralized);
- where to locate the instruments and visualization resources (less easily centralized);
- campus network architectures and protocols required to link this equipment to researchers' desktops; and
- issues of distributing campus-based resources across networks over which they must be accessed and operated.

Two primary models for addressing the new last-mile problem for high-performance computing exist. In one model, computational resources are centralized in a single place on campus (typically a computing center), to which is attached the main interface from the outside research and education network. Through the computer center, campus resources then can be attached to, and researchers can access, the resources and services associated with the national cyberinfrastructure. This model works well when the campus network is robust and when strategic, policy, and political factors have been well articulated and well integrated.

A second, dispersed model offers a different approach when certain factors are not well aligned—for example, where a strong, central, strategic IT environment does not exist, or a culture of distributed research and IT does exist. A strong case can be made for trying to overcome culture-based situations where IT resources are dispersed rather than centralized. Since the research agenda is shifting from a local to a national perspective, there are strategic issues concerning how resources are configured and where they are located. For institutions participating in national and even international collaboration, competitive pressures may be sufficient to overcome a culture of unrestrained distribution and decentralization. This simplifies the last-mile issue by decreasing the number of on-campus networks capable of very high performance (lambda level) and focuses attention on a main access point to the national cyberinfrastructure from the central computing center or hub.

Beyond these two models, a third approach to the campus network may be required to accommodate new devices such as scientific and medical instrumentation and visualization facilities, which, by their very nature, cannot be easily centralized. Data-gathering instruments such as weather stations, seismic stations, and telescopes, as well as visualization devices such as CAVEwaves and display walls, typically must be located in special places on campus. In order to transmit the vast amount of data they collect, higher-speed networks are required. When these conditions exist, either alone or in combination, a more distributed high-performance network infrastructure must also be present on campus to support that environment.

For example, the Indiana University (IU) Bloomington campus network was designed with the idea that the majority of computer processing power, storage, and archiving would be located in a central area in support of a strong and strategic centralized IT infrastructure and a research culture that is well served by this model. Last-mile pressures therefore were not a major concern, and the focus was on building the "on-ramp" to the national cyberinfrastructure. But IU, like most other research institutions, faces a challenge as an increasing number of high-performance devices and data-collection instruments are distributed across campus. As high-performance computing evolves, instruments will be collecting data from distributed locations and large data-storage silos. Network planners must therefore revisit existing architecture designs, since more places on campus may need higher capacity networks than would typically be deployed.

Contemporary research agendas in higher education are dictating an institutional response that embraces emerging network capabilities along with a growing portfolio of enabling technologies (middleware, grids, visualization and analysis tools, and so forth). As networks become more robust, in terms of capacity and capability, IT planners and university administrators must facilitate links to what is rapidly becoming a ubiquitous component of the national cyberinfrastructure. Access to grid-enabled resources of all types is fundamentally tied to networking infrastructures that, in turn, allow scientists and researchers to collaboratively pursue multidisciplinary and interinstitutional research projects. These are the hallmarks of 21st-century academic research that form the benchmarks for IT planners in higher education today.

Key Questions to Ask

- As currently configured, is our IT organization properly aligned and integrated with institutional policies, priorities, mission, and culture, with respect to enabling research?
- To what extent do our campus network and IT environment enhance institutional prestige and competitiveness for recruiting and retaining leading researchers and scholars?
- Do our campus computer network and related research resources enable high-end research across multiple disciplines?
- How well do our IT and sponsored research functions work with researchers to resolve the issues of centralization and distributed architectures for key resources, such as computation and storage? What benefits, if any, might accrue from pooling computation, storage, instrumentation, and visualization resources?
- Are we planning and building the campus infrastructure ahead of emerging technologies (wireless, VoIP, IP multicast, 10 Gigabit Ethernet, converged services) and researcher demand? To what extent do we treat the network as another utility in campus budgets?

Where to Learn More

- F. Berman, G. Fox, and A. J. G. Hey, eds., *Grid Computing: Making the Global Infrastructure a Reality* (Hoboken, N.J.: Wiley Publishers, 2003).
- I. Foster and C. Kesselman, eds., *The Grid 2: Blueprint for a New Computing Infrastructure* (San Francisco: Morgan Kaufmann Publishers, 2003).
- W. Gentzsch, “The Grid: Defining the Future of the Net,” MCNC Grid Computing and Network Services, <http://www.mcnc.org/index.cfm?fuseaction=page&filename=gridvsinternet.html>.
- B. D. Voss, “The Impact of Globalization on Campus Networks” (Boulder, Colo.: EDUCAUSE Center for Applied Research, Research Bulletin, Issue 4, 2004), <http://www.educause.edu/LibraryDetailPage/666?ID=ERB0404>.

Endnotes

1. L. Smarr, “Providing Your Faculty Global Access to the Instruments of Scientific Discovery,” presentation at the 2004 ECAR Symposium, Coronado, California, November 16, 2004, <http:www.educause.edu/LibraryDetailPage/666?ID=ECR0406>.
2. J. N. Cummings and S. Kiesler, “Collaborative Research Across Disciplinary and Organizational Boundaries,” *Social Studies of Science*, Vol. 35, No. 5, 2005, pp. 703–722, <http://sss.sagepub.com/cgi/content/refs/35/5/703>.

3. D. Atkins et al., *Revolutionizing Science and Engineering Through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure* (Arlington, Va.: National Science Foundation, January 2003), <http://www.nsf.gov/cise/sci/reports/atkins.pdf>.
4. "End-to-End 10 Gbps Wavelength Inaugurates New Optical Networking Infrastructure," *National LambdaRail News*, November 2, 2004, <http://www.nlr.net/20041102-PR.html>.
5. See, for example, images from fall 2005 hurricanes Katrina and Wilma at <http://svs.gsfc.nasa.gov/search/mostrecent.html>.
6. J. A. Pirani et al., *Information Technology Networking in Higher Education: Campus Commodity and Competitive Differentiator* (Boulder, Colo.: EDUCAUSE Center for Applied Research, Research Study, Vol. 2, 2005), p. 136, <http://www.educause.edu/LibraryDetailPage/666?ID=ERS0502>.
7. OptIPuter, <http://www.optiputer.net/>.
8. C. A. Stewart et al., "High-Performance Computing: Delivering Valuable and Valued Services at Colleges and Universities," proceedings of SIGUCCS, Portland, Oregon, October 2001, <http://www.indiana.edu/~rac/stewart_siguccs.pdf>.
9. G. Gilder, *Telecosm* (New York: The Free Press/Simon and Schuster, Inc., 2000).
10. T. West, "Control of the Network Infrastructure: Vital to Advancing Research and Science and Bridging Communities," keynote address at SC2004, Pittsburgh, Pennsylvania, November 9, 2004, <http://www.nlr.net/presentations/SC2004_TWW_Text.pdf>.
11. TeraGrid, <http://www.teragrid.org/>.
12. NSF National Middleware Initiative, <http://www.nsf-middleware.org/>.
13. The Globus Alliance, <http://www.globus.org/>.
14. C. C. Boesz, "Management Challenges for NSF for FY 2006," NSF Office of Inspector General memorandum, National Science Foundation, October 14, 2005, <http://www.nsf.gov/oig/managementchallenges2006.pdf>.

About the Authors

Stephen L. Daigle (steved@calstate.edu) is Senior Research Associate in the division of Information Technology Services at California State University, Office of the Chancellor. Brian D. Voss (bvoss@lsu.edu) is Chief Information Officer for Louisiana State University and the A&M Campus in Baton Rouge, Louisiana.