

Making Research Cyberinfrastructure a Strategic Choice

Growing demands for research computing capabilities call for partnerships to build a centralized research cyberinfrastructure

By **Thomas J. Hacker** and **Bradley C. Wheeler**

The commoditization of low-cost hardware has enabled even modest-sized laboratories and research projects to own their own “supercomputers.” We argue that this local solution undermines rather than amplifies the research potential of scholars. CIOs, provosts, and research technologists should consider carefully an overall strategy to provision sustainable cyberinfrastructure in support of research activities and not reach for false economies from the commoditization of advanced computing hardware.

This article examines the forces behind the proliferation of supercom-

puting clusters and storage systems, highlights the relationship between visible and hidden costs, and explores tradeoffs between decentralized and centralized approaches for providing information technology infrastructure and support for the research enterprise. We present a strategy based on a campus cyberinfrastructure that strikes a suitable balance between efficiencies of scale and local customization.

Cyberinfrastructure combines computing systems, data storage, visualization systems, advanced instrumentation, and research communities, all linked by a high-speed network across



campus and to the outside world. Careful coordination among these building blocks is essential to enhance institutional research competitiveness and to maximize return on information technology investments.

Trends in Research Cyberinfrastructure

The traditional scientific paradigm of theory and experiment—the dominant approach to inquiry for centuries—is now changing fundamentally. The ability to conduct detailed simulations of physical systems over a wide range of spatial scales and time frames has added a powerful

new tool to the arsenal of science. The power of high-performance computing, applied to simulation and coupled with advances in storage and database technology, has made the laboratory-scale supercomputer indispensable research equipment. These new capabilities can bestow a significant competitive advantage to a research group and help a laboratory publish better papers in less time and win more grants.¹

Many trends and forces shape research cyberinfrastructure today in academic institutions:

- Rapid rate of commoditization of computation and storage

- Emergence of simulation in the sciences
- Increasing use of IT in the arts and humanities
- Escalating power and cooling requirements of computing systems
- Growing institutional demands for IT in an era of relatively flat levels of funding for capital improvements and research

Commoditization Trends Affecting Cyberinfrastructure

The concept of building cost-effective supercomputers using commodity parts was introduced in 1994.² From 1994 until today, predictable trends of technology improvement and commoditization have increased the power of off-the-shelf components available for cluster designers (see Table 1). These trends include Moore's law, Gilder's law, and storage density growth.³ Downward trends in technology unit prices for storage and memory have accelerated since 1998.⁴

Semiconductor memory prices have experienced a similar price reduction. Complementary to commoditization trends is the growing pervasiveness and reliability of the Linux operating system and of open-source cluster-management tools. Many vendors now offer cluster products that are relatively simple to install and operate.

The research community is actively exploiting these trends to develop laboratory-scale capabilities for simulation and analysis. The growing influence of cluster computing since 1994 is clearly demonstrated by its impact on the distribution of computer architectures in the Top500 supercomputer list.⁵ Large clusters have displaced all other systems to become the dominant architecture in use for supercomputing today. This trend illustrates how the forces of commoditization have come to dominate high-end computing.

Adoption of IT in the Arts and Humanities

In the arts and humanities, fundamental changes are taking place in the conduct of research and creative activities. Funding is increasing for digital

Table 1**Technology Capability Growth Trends**

Moore's Law	CPU transistor density (and clock speed) doubles every 18 months.
Gilder's Law	Network bandwidth increases by a factor of four every three years.
Storage Growth	Storage density increases by a factor of 100 every 10 years. In 1998, disk drives cost about \$0.08/Mbyte. Today, a 500 Gbyte disk drive goes for \$286, or \$0.0006/Mbyte—less than 1% of 1998 prices.

content creation, synthesis of new content from existing digital works, and digitization of traditional works. A recent report from the American Council of Learned Societies on cyberinfrastructure for the humanities⁶ highlights these trends. The report describes significant unsolved “grand-challenge” problems of using information technology and cyberinfrastructure to reintegrate the fragmented cultural record. Addressing these grand-challenge problems will require institutional commitments to the long-term curation and preservation of digital assets and to providing open Internet access to unique institutional collections.

The digitization project by Google offers one example of this paradigm shift in the arts and humanities. The Google project aims to provide universal access to millions of volumes from research university libraries. As electronic collections grow in scale and size, new forms of creative expression and scholarship will become possible, further increasing demands for information technology infrastructure and support.

Costs of Cyberinfrastructure for Research

The expansion of power and cooling requirements for modern computers are well known. Providing adequate facilities for current and future needs is one of the largest problems facing academic computing centers today.

Unlike hardware costs, environmental and staff costs to operate a research cyberinfrastructure are not driven by

the commodity market and represent large recurring expenses. In an era of flat budgets, this situation makes it difficult even for central IT providers to provide adequate facilities or professional staff to support the demand for computational clusters and research computing. These problems are compounded by the last decade of growth in digital and Web-based administrative and instructional services, which has put a strain on physical facilities and staff resources in central IT organizations.

The scarcity of central IT support and facilities for research cyberinfrastructure represents a gap between institution-wide needs and the capacity to deliver services at current funding levels. This capability gap puts the research community at a competitive disadvantage and drives individual researchers to meet their needs through the development of in-house research computing. Few researchers and scholars want to be in the business of developing their own cyberinfrastructure; they are simply seeking to remedy the lack of the cyberinfrastructure they need to support their work.⁷

It is sensible to leverage commoditization trends to broaden access to research cyberinfrastructure. Universities may promote or tolerate the trends of decentralization, but should understand all the costs involved in operating decentralized research computing. Some costs, such as capital expenditures for the initial purchase of equipment, are simple to quantify. Other costs, such as floor space to house equipment and depreciation, are less obvious and can

represent significant hidden costs to the institution.

Case Study: Cost Factors for High-Performance Computing

To understand the tradeoffs between decentralized and centralized research computing, we can break down some of the costs for operating a computational platform, using a supercomputer as an example. Cost factors include:

- Equipment costs—costs for initial acquisition, software licenses, maintenance, and upgrades over the useful lifetime of the equipment.
- Staff costs—operations, systems administration, consulting, and administrative support costs.
- Space and environmental costs—data center space, power, cooling, and security.
- Underutilization and downtime costs—operating over-provisioned resources and loss of resources due to downtime.

Patel described a comprehensive model for calculating the costs of operating a data center.⁸ To compare operational costs for centralized and distributed research computing, we ask “Is it less expensive to provide operational costs (space, power, cooling, staff, and so forth) in one central location, or is it cheaper to support many smaller distributed locations?”

Comparing equipment acquisition costs in these two scenarios must take into account significant savings possible through the coordinated purchase of one very large system, compared with many smaller independent purchases. In our analysis, we assume that a large central purchase costs less than the uncoordinated purchase of a number of systems.

Patel described the true total cost of equipment ownership as the sum of the costs for space, power, cooling, and operation. We consider each in turn.

Space, Environmental, and Utility Costs. The costs for providing space depend on how efficiently the space is used (amount of unit resources per square foot of space) and on facility construction costs. Modern data centers can provide

highly efficient and dense cooling and conditioned power at a lower unit cost than laboratory-scale computer rooms. This makes it feasible to host computer equipment in a central data center at a much higher density than a laboratory computer room. Furthermore, operating many small computer rooms that have over-engineered air-conditioning and electrical systems can result in greater aggregate underutilized capacity than a central data center.

In terms of cooling, there is a sizeable difference in cost per BTU between small and large computer room air-conditioning systems. Using data from the 2006 RSMeans cost estimation guide,⁹ installing a small 6-ton unit costs \$4,583 per ton versus \$1,973 per ton for a 23-ton cooling unit (commonly used in large data centers).

A recent development is the return of water cooling, which more effectively removes heat from modern computing equipment. Provisioning water cooling in a large central facility can use chilled water from a utility or a large chilling plant.

Comparing space, environmental, and electrical costs for an equal amount of computing power, we believe that a central data center is less expensive to provision and operate than several smaller decentralized computer rooms.

Operational Costs. Operational costs include personnel, depreciation, and software and licensing costs.

In a central data center, a coterie of qualified professional staff is leveraged across many systems. Although individual staff salaries exceed the costs for graduate students, the staff costs per unit of resource are fairly low.

In the decentralized case, graduate assistants (GAs) often provide support as an added, part-time responsibility. This decentralized staffing model has several inherent drawbacks. First, the GA's primary job is to perform research, teach, and work on completing the requirements for a degree, not to provide systems administration and applications consulting for their group. Second, compared with professional staff, GAs are generally less effective systems

administrators. They are hampered by a lesser degree of training and expertise and must distribute their efforts over a smaller number of computers housed in the laboratory in which they work. Third, the average tenure of a GA at a university is (or ideally should be) less than the term of a professional staff member. The lack of continuity and retention add transition costs for training new graduate students to take over support functions for the laboratory computational resources.

Based on these factors, we believe that personnel costs for decentralized research computing support greatly exceed costs for a central data center. Not only are the obvious costs higher, but the redirection of productive graduate student energies into providing support represents a hidden drain on the vitality of the institutional research enterprise. It makes better sense for graduate students to focus on activities in which they are most productive—research—rather than on activities that could be provided more effectively by professional staff.

Under Use and Downtime Costs. Two hidden costs were not quantified by Patel: under use and downtime. Under use occurs when a computational cluster is not fully utilized. If a system sits idle, it delivers no productive work while consuming resources and depreciating in value. Unused time is much less likely on a central shared cluster, which should be adequately provisioned to balance capacity and demand to avoid under use or over subscription. Downtime occurs when the system is unavailable due to hardware or software failures or when the lack of a timely security patch forces a system shutdown. Downtime is much more likely in a small laboratory situation in which researchers have limited time available to keep up with security patches. Inadequate cooling and power systems can also increase the probability of system hardware failure.

Although the purely decentralized model potentially provides shorter wait times for resource access, the hidden costs and decreased research productivity borne by the institution from under

use and downtime can be enormous. For example, at electric rates of \$0.08 per kilowatt-hour, a 1-teraflop (TF) system consuming 75 kilowatts of electricity will generate an annual utility bill of \$52,416. If 20 of these 1-TF systems are distributed over campus, the total annual utility bill will reach \$1,048,320. If the total achieved availability and use of these systems reach only 85 percent, then \$157,248 in annual utility costs will be wasted powering systems during the 15 percent of the time they sit idle. If a smaller 18-TF system with 95 percent availability (essentially providing the same number of delivered cycles as the 20 TF system) is supplied by the central IT organization, the university can achieve a power savings of \$104,832 per year. The savings can be used to hire professional staff or purchase additional equipment.

As research computing scales up in both power and pervasiveness within the institution, the cost differential between centralized and decentralized approaches will continue to increase. Based on our analysis of the true costs of equipment ownership, we believe the purely decentralized approach to research computing is not cost effective. Moreover, the decentralized approach has significant hidden costs that can hinder institutional research efforts.

The costs described in this section are incurred to support the research activities of the institution. By nature, universities and research organizations tend to favor local or disciplinary specialization that favors decentralization. The activities and infrastructure within research laboratories are driven by research projects conducted in those labs. The costs of operating this infrastructure are borne by the institution regardless of the existence of a coordinated strategic approach for acquiring and operating this infrastructure.

Acknowledging this situation, we believe it's important to develop a purposeful strategy for guiding and shaping the flow of computational resources into the institution. The strategy should attempt to rationalize investments, eliminate redundancies, and minimize operational costs. If it is possible to

reduce costs by even 5 percent, the payoff can easily justify efforts to develop and put into place a campus strategy for campus cyberinfrastructure.

A Purposeful Strategy for Campus Cyberinfrastructure

The trends and forces we have described are a major part of the impetus toward decentralized research computing. The challenge to IT organizations is to formulate a strategy to respond to these changes. Realistically, a completely decentralized or centralized model for research computing won't work. Innovation, autonomy, and discovery happen at the edges, in laboratories and studios where scholars and researchers work. At the same time, economies of scale and scope can only be realized centrally, where it is possible to leverage large-scale systems and professional staff.

A central tension separates these two models. Several questions must be considered to design an effective solution:

- What balance between the two makes the most financial sense for the institution and optimizes research productivity?
- How can institutions best leverage central resources and staff to provide a base infrastructure for research that allows individuals at the edge to focus on building on the central core to add value for their discipline?
- What impacts does a campus strategy for cyberinfrastructure have on faculty, students, and staff?

We argue that the right approach to answering these questions is to create an institutional *cyberinfrastructure* that synthesizes centrally supported research computing infrastructure and local discipline specific applications, instruments, and digital assets. As noted above, cyberinfrastructure combines high-performance computing systems, massive data storage, visualization systems, advanced instrumentation, and research communities, all linked by a high-speed network across campus and to the outside world. These cyberinfrastructure building blocks are essential to support the research and creative

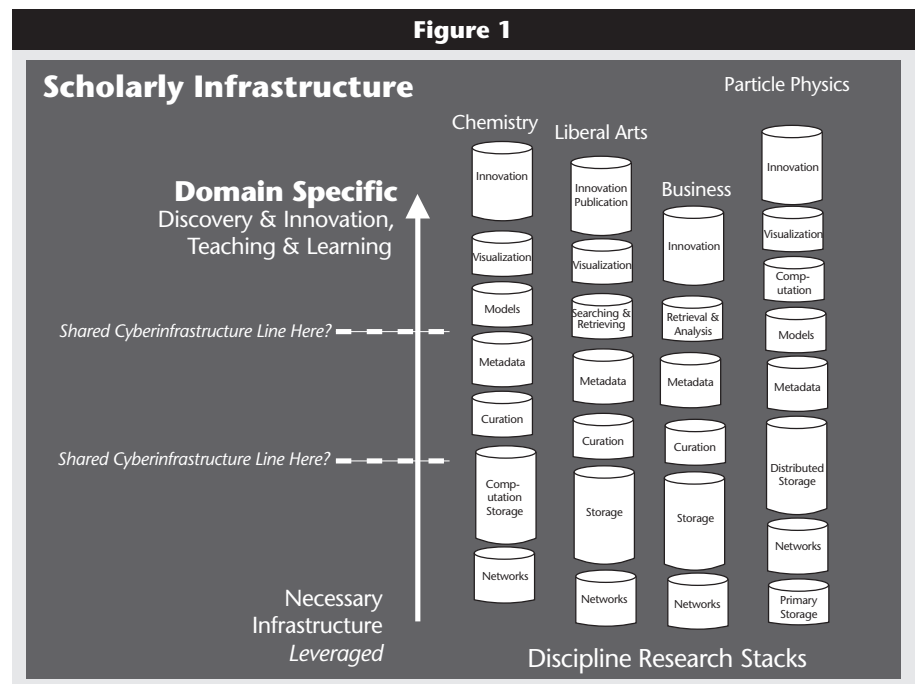
activities of scholarly communities. Only through careful coordination can they be linked to attain the greatest institutional competitive advantage. Ideally, a campus cyberinfrastructure is an ongoing partnership among the campus research community and central IT organization that is built on a foundation of accountability, funding, planning, and responsiveness to the needs of the community.

Specific needs for research computing depend on the prevalence and diffusion of computer use within a discipline. In the arts and humanities, for example, information technology only recently has begun to play a broad and significant role.¹⁰ In contrast, science and engineering have a tradition of computer use spanning half a century. Figure 1 illustrates a continuum from shared infrastructure at the bottom of the figure (Networks) up through layers of progressively more specialized components that support domain-specific activities. The transition from shared cyberinfrastructure to discipline-facing technologies operated by researchers depends on the specific needs and requirements of the domain. For example, business faculty may require a well-defined set of common statistics and authoring tools. In contrast, the particle physics commu-

nity may need to directly attach scientific equipment computing and storage systems using specialized software. The transition from shared cyberinfrastructure to laboratory-operated systems will be much lower in this figure for physicists than for business faculty. Central IT providers must be sensitive to these disciplinary differences and willing to work alongside the research community to develop specific cyberinfrastructure solutions for each discipline.

Campus Cyberinfrastructure Goals

We believe that a campus cyberinfrastructure strategy must achieve several specific goals to succeed. First, it should empower scholarly communities by reducing the amount of effort required to administer, learn, and use resources, which frees the community to take risks, explore, innovate, and perform research. To meet this goal, institutions should seek to eliminate redundant efforts across campus. They must break down silos and centralize activities that central IT organizations can most effectively provide. By reducing redundancies, local IT providers can focus energies on adding value to the core infrastructure for the research community.



To encourage resource sharing and develop centers of expertise and excellence at local levels, institutions should establish discipline-specific local cyberinfrastructure initiatives. Once a functional campus cyberinfrastructure initiative and local cyberinfrastructure initiatives are established, the next logical step is to broaden external engagement with discipline-specific research communities to create a national discipline-oriented cyberinfrastructure. An example of this approach is the U.S. Atlas project, which brings together a collaborative community of physicists to search for the Higgs boson.

Second, a campus cyberinfrastructure strategy must develop a central research computing infrastructure through consensus and compromise among university administrators and researchers. To reduce the motivation for units to develop redundant services, the central IT organization must carefully plan and fund infrastructure improvements to meet current and projected needs. Cost savings realized from centralizing base-level services should be captured and reinvested back into expanding basic shared IT facilities and infrastructure, which are essential for the ultimate success of a campus cyberinfrastructure strategy.

The final goal is realignment of existing, disjointed research-computing efforts into a harmonized campus-wide cyberinfrastructure. A crucial aspect of building a consolidated campus cyberinfrastructure is developing a common set of middleware, applications, infrastructure, and standards that are compatible with emerging cyberinfrastructure platforms at other institutions. Adopting a common platform makes it possible to build bridges from campus cyberinfrastructure to regional and national cyberinfrastructure initiatives. If a campus adopts the use of X.509 certificates for authentication and authorization, for example, the campus cyberinfrastructure can easily interoperate with other national cyberinfrastructure initiatives that use X.509.

Another concrete example of this comes from Indiana University's participation in the Sakai project. Several years ago, a strategic decision was made to

transition away from several incompatible learning management systems (LMS) to a common LMS based on Sakai. The adoption of a common LMS has made it possible to partner with other institutions using Sakai and to win external funding for collaborative projects that build on the Sakai framework.

An important factor to consider is how these goals will affect how people work. For faculty, graduate students, and researchers, the desired outcome is to increase research productivity by freeing time now spent running low-value activities in their own IT shops and by improving the effectiveness of infrastructure available for their use. For IT staff, as a result of greater coordination and reduction of replicated services, more time should be available to develop and deploy new services that add value to the underlying IT infrastructure.

Building a Campus Cyberinfrastructure

Building a campus cyberinfrastructure for research is not only a technical process but also a political, strategic, and tactical undertaking. It suffers from a "which came first, the chicken or the egg?" causality dilemma. Developing political support for making big investments in central systems to start the process of building cyberinfrastructure relies on the perceived trustworthiness of the central IT shop. A dilemma arises when the central IT shop suffers from the lack of funding necessary to provide very high levels of reliability to the campus, which is a necessary first step in building trust.

As we described in the section on cost factors, the institution is already making investments in centralized or decentralized computing. We believe the institution must be willing to risk starting the process by making significant strategic investments in core computing. This section describes some steps that could be taken in building a research cyberinfrastructure. These activities are not linear; rather, they are simply areas to consider and address.

The first activity in forging a common cyberinfrastructure is to identify common elements of campus infrastructure

that can be centralized. These common elements include computer networks, storage resources, software licenses, centrally managed data centers, backup systems, and computational resources. Many broadly used applications (such as Mathematica or SPSS) could be centrally sponsored and site licensed to keep costs down and guarantee consistent support.

The second activity is to adopt and create common standards for middleware, which is the software that lies between infrastructure and applications. The functions of middleware include authentication, authorization, and accounting systems; distributed file systems; Web portals (such as the Open Grid Collaboration Environment portal¹¹); and grid computing software, such as Globus,¹² PBSPro,¹³ and Condor.¹⁴

The middleware needs of disciplines can vary. One set of disciplines may be actively engaged in developing new middleware tools that require complete access to and control over the middleware layer for development and testing. Other disciplines might not develop new middleware, but may rely entirely on centrally supported middleware systems and services (such as Kerberos). Central IT organizations need to collaborate with these disciplines and learn to accommodate a wide range of support needs. Finding the best balance among openness, security, privacy, and stability may be the most difficult step in building common middleware.

The third activity is to identify and develop a cyberinfrastructure application layer, which relies on coordinated infrastructure and middleware layers. In many respects, this is the "face of the anvil" on which research communities carry out innovation and creative work. Finding the best balance between local and campus cyberinfrastructure depends on the characteristics of the discipline. For example, anthropologists may need significant training and central support to build new metadata models for capturing and archiving field data. Chemists, on the other hand, may only require basic infrastructure to run scientific codes used by a small research community.

One effective way to balance the tension between centralization and localization is to develop a cost-sharing model for funding specialized applications used by a small fraction of the research community. Researchers developing new applications and tools need well-supported development environments, mathematical libraries, secure authorization and authentication frameworks, source code management systems, debugging tools, and training materials. Providing stable and secure development environments for multiple platforms and programming languages frees the research community from the necessity of provisioning their own environment. This allows them to focus on creating new intellectual value in which the university has a vested interest.

The fourth activity is to focus on the social aspects of campus cyberinfrastructure. Scholarly communities form the topmost layer, which is the locus of innovation and research. Cyberinfrastructure frees members of these communities from constraints of physical location and time by facilitating collaborative activities across projects and disciplines. An example of this layer is the Open Science Grid, an open collaboration of researchers, developers, and resource providers who are building a grid computing infrastructure to support the needs of the science community.

Achieving these objectives is not necessarily a sequential process. Formulating a response to the factual trends shaping the course of research computing requires making a set of choices that carry costs and risks: the time required to build community consensus among campus constituencies; the need for leadership awareness and attention to research computing and accompanying costs; the extra effort required by IT staff to collect information for activity-based costing, balanced scorecard, and annual surveys; and the extra diligence required to proactively plan and build cyberinfrastructure (along with the risks of unforeseen change) rather than reacting to specific problems and crises as they arise. Choices that work for one institution may not be effective at others. The

ultimate success of a cyberinfrastructure plan depends on organizational context and the application of leadership skills to develop a strategy and plan.

Engaging the campus community on all these levels while building campus and local cyberinfrastructure is an effective way to seek rough consensus and establish accountability between the research community and central IT organization. By working together rather than independently, the university community has the best chance of creating a working and sustainable infrastructure and support model for research computing.

Campus Cyberinfrastructure at Indiana University

Indiana University is a confederation of two large main campuses and six regional campuses serving more than 90,000 students. The main campuses are in Bloomington and Indianapolis. The Bloomington campus portfolio includes physics, chemistry, biological sciences, informatics, law, business, and arts and humanities. The Indianapolis campus provides undergraduate and graduate programs from Indiana University and Purdue University and includes the IU Schools of Medicine and Dentistry. The six regional campuses provide undergraduate and master's level programs for Indiana residents across the state.

In the mid-1990s, the IT infrastructure of Indiana University spread across eight campuses, with very little sharing of infrastructure or staff expertise. Each campus had a CIO or dean of IT who was responsible for academic and (at some campuses) administrative computing for his or her respective campus. Clearly, a major institutional intervention was required to achieve system-wide efficiency and optimal performance. In 1996, a strategic vision developed for Indiana University included a "university-wide information system that will support communication among campuses..."

In 1998, IU developed a comprehensive five-year IT strategic plan (ITSP)¹⁵ that involved nearly 200 faculty, administrators, students, and staff working

together in four chartered task forces. The task forces identified critical action items and steps to address existing deficiencies in the IU IT environment. The final ITSP described 68 specific action items and established the basis for planning, redeploying existing funding and resources, and seeking new funds.

Using the ITSP as both a plan and a proposal, IU approached the Indiana Legislature to seek additional funding to make it a reality. The legislature responded by providing a small increase to IU's budget over a period of five years (the lifetime of the ITSP) specifically targeted to building IU's effectiveness and reputation through leveraging IT to enhance teaching, research, economic development, and public service.

The ITSP included a section focused on research computing support across all IU campuses. Within this section, seven specific action items were identified, one for each research computing strategic area:

- **Collaboration.** Explore and deploy advanced and experimental collaborative technologies within the university's production information technology environment, first as prototypes and then, if successful, more broadly.
- **Computational Resources.** Plan to continually upgrade and replace high-performance computing facilities to keep them at a level that satisfies the increasing demand for computational power.
- **Visualization and Information Discovery.** Provide facilities and support for computationally and data-intensive research, for nontraditional areas such as the arts and humanities, as well as for the more traditional areas of scientific computation.
- **Grid Computing.** Plan to evolve the university's high-performance computing and communications infrastructure so that it has the features to be compatible with and can participate in the emerging national computational grid.
- **Massive Data Storage.** Evaluate and acquire high-capacity storage systems capable of managing very large data volumes from research instruments,

remote sensors, and other data-gathering facilities.

- **Research Software Support.** Provide support for a wide range of research software including database systems, text-based and text-markup tools, scientific text processing systems, and software for statistical analysis.
- **Research Initiatives in IT.** Participate with faculty on major research initiatives involving IT where appropriate and of institutional advantage.

Building IU's comprehensive cyberinfrastructure began with a comprehensive strategic plan and funding. The institution took the risk of developing core computing capabilities to support research across all IU campuses. This leads back to our central thesis: by taking the steps of assessing all the costs, developing a plan to coordinate activities, securing funding, and building political support, IU solved the chicken and egg dilemma.

Putting a cyberinfrastructure in place is one part of the solution. Building a sustainable cyberinfrastructure requires additional elements to make the vision a reality. The first element involves using the IT strategic plan as a living document. The second necessary element is accountability.

The central IT organization is a service organization that supports the institution. As such, it must be accountable to clients and customers as well as to university leadership. Accountability to university administration is accomplished through the use of four mechanisms:

- Activity-based costing
- Annual activity and performance reports on strategic plan progress
- Adhering to the strategic plan as a basis for yearly budget and planning activities
- Periodic comprehensive efficiency reviews that seek to reduce redundancies and retire obsolete services

Annual reports on cost and quality of services¹⁶ are open and available to the university community. Accountability to customers relies on the use of a comprehensive user satisfaction survey¹⁷ sent to more than 5,000 randomly selected staff, faculty, and stu-

dents across all eight IU campuses. Based on survey responses and individual comments, each unit reviews and makes any necessary changes to services it provides.

The survey results ensure that the central IT organization remains responsive to needs of the university community. Based on survey results, the research computing unit maintains an annual balanced scorecard¹⁸ that provides a comprehensive overview of efficiency and user satisfaction with research computing services. These quantitative tools allow IT leadership to monitor user satisfaction, ensure cost-effective service delivery, and retire outdated services that no longer serve user needs or are not cost-effective.

Feedback from the research community to the systems and services provided to meet research needs has been positive. Detailed comments from researchers from 16 years of survey results are publicly available on the Web.¹⁹ In 2006 alone, more than 430 detailed comments were received from the user community.

One tangible example of this process is a change made several years ago in campus e-mail service. Satisfaction with text-based e-mail was declining, and an investigation determined that the community had a growing unmet need for Web-based mail. In response, the central IT organization formulated a plan and one-time budget expenditure to establish a Web-based mail system. After successful deployment of the system, user satisfaction returned to the previous high levels.

With the firm foundation of reliable services and resources in place, IU is working to build the middleware, application, and collaborative technology cyberinfrastructure layers necessary to construct an excellent campus cyberinfrastructure.²⁰ IU's activities bridge IU campuses within the state and connect IU and national scholarly communities. The projects include Sakai, Kuali, Teragrid, and regional, national, and international networks, as well as working with communities such as the Global Grid Forum and the Open Science Grid.

Where Is Research Computing Going?

Research computing in the future will be shaped by current trends and forces, as well as by several emerging trends that will take hold over the next three years.

Commoditization trends will continue. With increasing globalization it is likely that commoditization will move down the value chain. One recent example of this is Sun Microsystem's announcement of the availability of a computing utility service over the Internet at a price of \$1 per CPU per hour. Development will be driven by the home market for computing and entertainment. New technologies developed for this market (such as the use of artificial intelligence for intelligent game agents) will continue to appear on the commodity market.

Web portals, Web services, and science gateways will likely reach maturity within the next few years. They have the potential to increase the collaborative power of cyberinfrastructure and broaden access to computing for researchers.

Another emerging force is the growing awareness of the significance of data. Data-centric computing seeks to capture, store, annotate, and curate not only the results of research but also all observations, experimental results, and intermediate work products for decades and potentially centuries. An additional trend is the developing need for central IT support in the arts and humanities.

A major force shaping research computing is the tide that ebbs and flows—federal research funding. Historian Roger Geiger²¹ has observed 10- to 12-year cycles in federal research funding, with peaks of rapid growth followed by periods of relative consolidation. If this trend persists, the current period of decline that began in 2004²² may be followed by a period of growth starting in the next few years. An encouraging sign is the recent State of the Union message, in which President Bush proposed doubling research funding for basic science research in the next 10 years. Laying the foundations of cyberinfrastructure now will help to prepare the institution

for potential future growth in the availability of research funds.

Conclusion

We believe the most effective response to the trends and forces in science and IT that are creating tremendous demand for research computing is to build partnerships among scholarly communities and central IT providers to develop campus and discipline-facing cyberinfrastructure capabilities. A successful cyberinfrastructure strategy will help prepare the institution for the coming globalization of the academy and research and for potential future growth in federal research funding. Advances in research and creative activity in the future will most likely come from global collaboration among scholars and scientists. Universities that learn to use cyberinfrastructure effectively to support the needs of their research community will gain a competitive advantage in the race to attract excellent scholars and win external funding to support research. *e*

Endnotes

1. U.S. Department of Energy, "The Challenge and Promise of Scientific Computing," 2003, <http://www.er.doe.gov/sub/Occasional_Papers/1-Occ-Scientific-Computation.PDF> (accessed December 1, 2006).
2. P. Goda and J. Warren, "I'm Not Going to Pay a Lot for This Supercomputer!" *Linux Journal*, January 1998, p. 45.
3. J. Gray and P. Shenoy, "Rules of Thumb in Data Engineering," in *Technical Report MS-TR-99-100* (Redmond, Wash.: Microsoft Research, 1999).
4. E. Grochowski and R. D. Halem, "Technological Impact of Magnetic Hard Disk Drives on Storage Systems," *IBM Systems Journal*, Vol. 42, No. 2, 2003, pp. 338-346.
5. Top500 Supercomputer Sites, <<http://www.top500.org>> (accessed November 17, 2006). Architecture distribution over time can be accessed at <<http://www.top500.org/lists/2006/11/overtime/Architectures>> (accessed December 1, 2006).
6. American Council of Learned Societies, "The Draft Report of the American Council of Learned Societies' Commission on Cyberinfrastructure for Humanities and Social Sciences 2005," American Council of Learned Societies, New York, pp. 1-64, <<http://www.acls.org/cyberinfrastructure/acls-ci-public.pdf>> (accessed December 1, 2006).
7. K. Klingenstein, K. Morooney, and S. Olshansky, "Final Report: A Workshop on Effective Approaches to Campus Research Computing Cyberinfrastructure," sponsored by the National Science Foundation, Pennsylvania State University, and Internet2, April 25-27, 2006, Arlington, Virginia, <<http://middleware.internet2.edu/crcc/docs/internet2-crcc-report-200607.html>> (accessed December 1, 2006).
8. C. Patel and A. Shah, "Cost Model for Planning, Development, and Operation of a Data Center in HPL-2005-107(R.1)" (Palo Alto, Calif.: Hewlett-Packard Internet Systems and Storage Laboratory, 2005).
9. RSMMeans, *Building Construction Cost Data 2006*, Vol. 64 (Kingston, Mass.: RSMMeans Construction Publisher, 2006).
10. American Council of Learned Societies, op. cit.
11. D. Gannon et al., "Grid Portals: A Scientist's Access Point for Grid Services (DRAFT 1)," GGF working draft Sept. 19, 2003 <<http://www.collab-ogce.org/nmi/index.jsp>> (accessed March 29, 2006).
12. I. Foster, "Globus Toolkit Version 4: Software for Service-Oriented Systems," in *IFIP International Conference on Network and Parallel Computing* (Berlin: Springer-Verlag, 2005), pp. 2-13.
13. "Altair Computing Portable Batch System," 1996, <<http://www.altair.com/software/pbspro.htm>> (accessed November 17, 2006).
14. D. Thain, T. Tannenbaum, and M. Livny, "Distributed Computing in Practice: The Condor Experience," *Concurrency and Computation: Practice and Experience*, Vol. 17, No. 2-4, pp. 323-356.
15. University Information Technology Committee, "Indiana University Information Technology Strategic Plan," 2001, <<http://www.indiana.edu/~ovpit/strategic/>> (accessed May 2006).
16. "Indiana University Information Technology Services Annual Report on Cost and Quality of Services," <http://www.iu.edu/~uits/business/report_on_cost_and_quality_of_services.html> (accessed April 2006).
17. "Indiana University Information Technology Services User Satisfaction Survey," <<http://www.indiana.edu/~uitssur/>> (accessed November 17, 2006); and C. Peebles et al., "Measuring Quality, Cost, and Value of IT Services," EDUCAUSE Annual Conference 2001, <<http://www.educause.edu/ir/library/pdf/EDU0154.pdf>> (accessed November 14, 2006).
18. "Indiana University Research and Academic Computing Balanced Scorecard," 2005, <http://www.indiana.edu/~rac/scorecard/2005/racscorecard_2005.html> (accessed November 17, 2006).
19. See <<http://www.indiana.edu/~uitssur/>> and Peebles, op. cit.
20. Klingenstein, Morooney, and Olshansky, op. cit.
21. R. Geiger, *Research and Relevant Knowledge: American Research Universities since World War II*, transaction series in higher education (New Brunswick, N.J.: Transaction Publishers, 2004), pp. xxi, 411.
22. *American Association for the Advancement of Science Guide to R&D Funding Data—Historical Data*, 2006, <<http://www.aaas.org/spp/rd/guihist.htm>> (accessed November 14, 2006).

Thomas J. Hacker (hacker@cs.purdue.edu) is Assistant Research Professor, Discovery Park Cyber Center, at Purdue University in West Lafayette, Indiana. Bradley C. Wheeler is the Chief Information Officer at Indiana University and an Associate Professor of Business.