



# Transformations OF THE Research ENTERPRISE

By Sandra Braman

At least since the time of Plato and Socrates, we've been aware that the technologies we use to create and share knowledge affect what it is that we know. Thus it is not surprising that over the last few decades, innovations in information technology have brought about fundamentally new approaches to research.<sup>1</sup> Nor is it surprising that, as a result, research methods, practices, and institutions are evolving into forms quite different from those that have dominated modern science. These changes are so significant that they are leading to theoretical innovations as well.

*Sandra Braman is Professor of Communication at the University of Wisconsin-Milwaukee and an EDUCAUSE Center for Applied Research (ECAR) Fellow. This article was produced in support of a forthcoming ECAR study on the engagement of IT in research.*

Colleges and universities making decisions about such matters as whether or not to centralize IT support for research, how far to go with the development of research problem-specific software, and how best to coordinate on-campus activities with those that take place elsewhere do so within a dynamic environment—one shaped not only by technological innovations but also by politics. Over the last couple of decades, computational science has come into its own as a discipline distinct from computer science; whereas the latter involves research on computing, computational science is the use of computing for research in the physical, life, and social sciences and in the humanities. This has transformed those who work with information technology from providers of services into full collaborative peers with researchers. In addition, the value of computational science to U.S. security, economic strength, and competitiveness in science and technology has been established. Thus, federal and state governments both enable (through funding) and constrain (through laws and regulations) research institutions and activities.

For these reasons, those involved in information technology in higher education need to be aware of the past, present, and future of IT engagement with research. This article reviews the political, economic, and intellectual developments that shaped the U.S. agenda for research and information technology, describes the contemporary research environment and identifies features that have IT implications, and highlights key computational, networking, and data challenges for tomorrow.

### The National Agenda for Research and Information Technology

In the United States, the national agenda for research and information technology developed out of a long appreciation of the value of scientific and technological innovation to society as a whole. The inclusion of intellectual property rights in the U.S. Constitution was one manifestation of the political value placed on this insight, and by the early nineteenth century, diverse government units—even the U.S. Post Office—were engaged in

scientific activity. A role for colleges and universities in creating and distributing knowledge in support of national goals was institutionalized via the Morrill, Hatch, and Smith-Lever Acts of Congress between 1862 and 1914, and college and university faculty were heavily involved in governmental activity during and following World War I. But it was not until after World War II and the launch of the National Science Foundation (NSF) in 1950 that the groundwork for today's developments was laid.

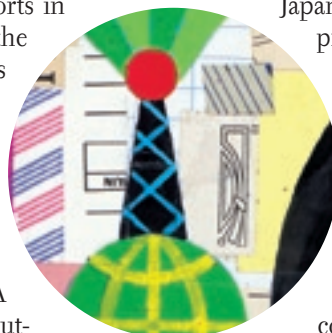
The NSF began its efforts in this area by supporting the development of campus computing centers. By the mid-1970s, however, it was still the case that only those researchers who worked with agencies such as the Department of Energy and NASA had access to supercomputers, limiting the scope of research for most faculty members. Those in experimental computer science and computation began to organize to communicate their concerns about the scholarly—and national—consequences of these restrictions, issuing the Feldman Report in 1979. They were soon followed by those in the physics community, who released the Press Report in 1981, and those in the physical and biological sciences, who produced the Lax Report the next year.<sup>2</sup> Taken together, this suite of reports argued for the revitalization of experimental computer science to generate products of tangible use in research (including the development of new types of computer systems to solve otherwise intractable problems); highlighted the importance of computation for the intellectual, economic, and military strength of the United States; and pointed to the need for more supercomputers and more access to those supercomputers for researchers based in academia.

Political support for the achievement of these goals received a tremendous

stimulus when the Japanese government launched its \$500 million “Fifth Generation Computer Project” in 1981. Intended to leapfrog existing technologies, this move catalyzed both the U.S. government and the U.S. scientific community. Federal funding agencies such as the NSF, the National Security Agency (NSA), and the Department of Energy believed that supporting academic access to supercomputing would serve three goals: (1) it would help protect the domestic supercomputing industry in face of the

Japanese challenge; (2) it would provide increased access to these machines for computer and computational science research; and (3) it would provide the experimental facilities needed to attract the best computer scientists to academic (rather than corporate) work, where they would then train the next generation of computer scientists. A growing number of scientists from across the disciplinary spectrum began to refer to themselves as “computational” scientists, having realized that computer simulations and other new types of interactions with data—interactions not previously possible—were so sophisticated that they provided new theoretical insights into biological and physical phenomena.

In 1987, the White House Office of Science and Technology Policy (OSTP) issued a report that looked to networked computing to support research.<sup>3</sup> In 1993, after a number of additional studies and congressional hearings, the U.S. Congress established a \$4.7 billion High Performance Computing and Communications Program (HPCC). Several foci of the HPCC remain important today. First, there was an emphasis on the importance of high-performance computing and communications to national security as well as to the economic competitiveness of the United States. Second, projects to be



**Federal and state governments both enable (through funding) and constrain (through laws and regulations) research institutions and activities.**

funded were framed within the context of scientific and engineering “grand challenges.” Finally, research projects using high-performance computing were to be collaborative ventures along three dimensions, involving (1) partnerships among entities from academia, the private sector, and government; (2) interdisciplinary work within the sciences; and (3) computational scientists who would be treated as full collaborators with researchers rather than as service providers.

The NSF operationalized the concept of grand challenges by establishing funding programs for work that joined disciplinary researchers, computer scientists, and emerging information technologies to solve fundamental science and engineering problems. In 1993 the NSF Blue Ribbon Panel on High Performance Computing summarized the intentions and goals of these projects, which differed from others funded by the NSF in that they were intended to accelerate progress in virtually every branch of science and

engineering concurrently (rather than being directed at single disciplines) in order to stimulate the U.S. economy as a whole.<sup>4</sup> Admitting that much of how science and engineering are practiced would be transformed if the panel’s recommendations were implemented, the NSF identified several issues that needed attention: the removal of technological and implementation barriers to the continued rapid evolution of high-performance computing; the achievement of scalable access to a pyramid of computing resources; the democratization of the base of participation in high-performance computing; and the development of future intellectual and management leadership for high-performance computing. Lewis Branscomb, who edited the report for the committee, also analyzed its implications for further policy. Policy-makers, he argued, should recognize emerging technologies essential to meeting the grand challenges as “critical technologies” the development of which would need other

types of legal and regulatory support, such as exemptions from antitrust law and changes in the tax structure.<sup>5</sup>

At the time, the continued U.S. domination of the world in science and technology was assumed. Only a decade later, however, the situation had changed significantly. Successes in analysis of the genome had dramatized the utility of computational science. It was no longer clear that the United States dominated the world in science and technology, and the political environment regarding definitions of the knowledge and technologies critical for survival had shifted. The NSF reconsidered its position on supercomputing centers and, in 2003, issued a report focused on the physical and biological sciences.<sup>6</sup> This report outlined a new strategic orientation around “cyber-infrastructure,” referring to the evolution of distributed, or grid, computing, which uses the telecommunications network to combine the computing capacity of multiple sites. The concept of networked com-

puting expanded to include tasks such as the harvesting of computing cycles, massive storage, and the management of data in both raw and analyzed forms over the long term. Such matters as middleware, top-level administrative leadership, and the development of new analytical tools and research methods were all identified as key to the knowledge-production infrastructure. The NSF called for increasing the speed and capacity of computers another 100–1,000 times in order to serve current research needs.

A second NSF report focused on the particular cyberinfrastructure needs of the social sciences, arguing that making progress in this area would be important not only for research on society but for knowledge production in general because of what can be learned from social scientists about human-computer interactions and about the nature of collaboration at a distance.<sup>7</sup> Recommendations included emphasizing funding and institutional support for training social scientists in

the uses of advanced computational techniques, as well as the skills needed to succeed within the context of distributed research teams. The report also identified policy issues such as security, confidentiality, and privacy.

The American Council on Learned Societies joined, in 2005, with a number of other groups to offer recommendations regarding the computational needs of those in the humanities.<sup>8</sup> This report noted that data produced by those in the humanities contributes to the cultural record, currently fragmented across institutional boundaries, often made obscure through redaction and other archival-processing activities or even made unavailable because it is proprietary. As in the social sciences, there is still a great deal of work to be done developing computational environments, software, and instrumentation in the humanities, though the first humanities project to use digital infrastructure—Project Gutenberg, which offers free digital versions

of full texts of important books—was launched as long ago as 1971. Those in the humanities increasingly collaborate with researchers in the physical and life sciences, as appreciation for the value of expertise in such areas as document and image analysis grows. Humanities databases can be as massive as or even more massive than those of the physical and biological sciences. For example, the Sloan Digital Sky Survey used x-ray, infrared, and visible-light images to survey more than 100,000,000 celestial objects and produced a dataset of more than 40 terabytes of information; by contrast, Survivors of the Shoah Visual History Foundation's analysis of video interviews of Holocaust survivors has produced 180 terabytes of data.<sup>9</sup>

Building on work by the Library of Congress and other federal agencies, the National Science Board, which governs the NSF, produced a 2005 report<sup>10</sup> on long-lived data needs and problems: the relative ephemerality of digital storage

media; the constancy and rate of technological innovation; the desire to revisit data, collected for one purpose, through different analytical lenses that will throw light on additional research questions; and the growing complexity and time-sensitivity of data curation. The data collection universe for any single institution now includes data collections, related software, and hardware and communication links along with data authors, managers, users, data scientists, research centers, and other institutions. Three different types of data-storage entities were distinguished in this report. Focused *research collections* support single research projects in which a limited group of researchers participate. Intermediate-level *resource collections* span various research projects and teams from a specific facility or center. *Reference collections* have global user populations and impact, incorporating data of multiple types across disciplinary, institutional, geopolitical, research project, and data type boundaries.

In June 2005, the President's Information Technology Advisory Committee (PITAC) issued a report reemphasizing the importance of computational science for U.S. competitiveness and for the nation's capacity to address problems such as nuclear fusion, the folding of proteins, and the global spread of disease.<sup>11</sup> In addition to strengthening computational and long-term data-storage capabilities and supporting the interinstitutional relationships needed to carry out large-scale distributed research, PITAC recommended the creation of national software sustainability centers to ensure that archived data would remain accessible irrespective of innovations in hardware and software. The committee warned that without new initiatives, U.S. leadership in computing and computationally intense science would continue to deteriorate. Interestingly, the long-standing White House committee was dissolved by President George W. Bush shortly after it issued this report.

Also in 2005, the Mathematical Association of America released a report that exemplified how disciplinary associations are thinking about the changes taking place.<sup>12</sup> For most of the twentieth century, mathematics was linked most

closely with physics and engineering, but today biology is seen as the stimulus for innovation. This report noted that evidence is now as often mathematical as observational, and it remarked on the shift in the relative prestige of various disciplines as they become more computationally intense.

National policies remain critically important for the nature of academic research. Unfortunately, an appreciation of the knowledge economy does not always translate into economic support for colleges and universities as the sites of knowledge production and distribution. Still, the number of institutional players continues to grow.

### The Contemporary Research Environment

As the *New York Times* science writer George Johnson has noted, today "all science is computer science."<sup>13</sup> Over a longer history, from the invention of the printing press and movable type in Western Europe more than five hundred years ago through the sensor web of the twenty-first century, technologies have made possible developments in the production of knowledge along several dimensions. We've gone from

- individuals working alone, to
- individuals interacting in face-to-face conversations with others, to
- individuals thinking about the work of others across space and time as accessed by print, to
- individuals developing new ideas based on the systematization of written records, to
- individuals observing the natural world, to
- individuals systematizing observations about the natural world, to
- individuals working in laboratories, to
- teams working in laboratories, to
- teams working with massive amounts of data derived from iterations of studies, to
- teams working collaboratively across laboratory, institutional, and national boundaries, to
- entirely new realms of research opening up as a result of high-performance computing.

The cumulative effects of these developments have formed the contemporary research environment, which is characterized by a number of features pertinent to the higher education IT field.

*Computation is the "third branch" of science, along with theory and experimentation.* For modern science as it developed over the last several hundred years, knowledge was produced through iterative interactions between theory and experimentation. Today, computation is seen as a third fundamental element of knowledge production. Kenneth Wilson's distinction among theory, experimentation, and computation has been particularly influential.<sup>14</sup> Experimentalists design and use scientific instruments to make measurements, undertake controlled and reproducible experiments, and analyze both the results of those experiments and the errors that appear within them. Theorists focus on relationships among experimental quantities, the principles that underlie these relationships, and the mathematical concepts and techniques needed to apply the principles to specific cases. Computational scientists focus on algorithms for solving scientific problems and their operationalization via software, the design of computational experiments and the analysis of errors within them, and the identification of underlying mathematical frameworks, laws, and models. Some describe visualization as the fourth branch of science because it has turned out to be so valuable in stimulating new conceptual and theoretical insights.<sup>15</sup>

*Every discipline is becoming computationally intense.* The first discipline to become computationally intense was physics, followed by chemistry and then biology. Today, all disciplines have taken up computation. For example, the term "megatronics" is used to describe the integration of computers and mechanical engineering. In the field of English, a database of every piece of scholarship by and about *Hamlet*, linked to every line of the play, is now critical to the over five hundred items still written on the subject each year; and software written to analyze the genome is now being used to determine which versions of Chaucer's texts are the earliest. Musicians are being fully

wired inside and out so that researchers can study what happens with their bodies as they play and sing. And ethnographers are using computers to comparatively analyze results from studies conducted at diverse sites.<sup>16</sup>

*Large research projects are more likely to be transdisciplinary than to be disciplinarily bound.* The expectation that computationally intense research projects would be interdisciplinary in nature was inherent in the design of the Internet and has been enunciated as a basic principle for all research involving information technology since the appearance of the NSF's grand challenges. The notion of computing and communications as research infrastructure, introduced in the Bardon Report of 1983,<sup>17</sup> has been helpful in this regard, orienting attention around problems to be solved rather than around disciplinary boundaries. The heterogeneity of research teams is a corollary.

*Research is more likely to be carried out in the context of a problem-oriented application than within the context of a specific academic community.* The enunciation of grand challenges started the process of shifting research funding attention toward specific problems and away from discipline-bound and theory-driven basic research. In recent years, U.S. national security concerns have raised the salience of research results intended to have short-term utility.<sup>18</sup> As a corollary, researchers are increasingly required to respond to social and political demands to be accountable rather than isolating themselves within an "ivory tower."

*Large research projects are more likely to involve working with the results of data from multiple studies at multiple sites across multiple time periods than working with single studies alone.* High-performance computing has made it possible to analyze much larger aggregates of data, expanding vision across space and time in ways previously not possible. Researchers working with such datasets need support in developing software capable of analyzing patterns across types of data, as well as institutional arrangements that ensure access to that data irrespective of where it resides or in what form.

*The distinction between "basic" and "applied" research has fallen away.* Though Vannevar

Bush's 1945 distinction between basic and applied research is still common in public discourse, sociologists of knowledge reject the distinction between basic and applied science as one-dimensional and replace it with a more complex array of linkages among various kinds of scientific activities. Similarly, the distinction between science and technology is falling away, since information technology is simultaneously the subject and the tool of inquiry. Technological innovation can develop out of existing technologies without additional fundamental research—and the use of technologies can itself generate new basic knowledge. For higher education institutions, these changes provide additional reasons to encourage the institutional development or adaptation of software to serve campus research needs, since doing so can result in patents and other successes for the college or university.

### Research and IT Challenges for the Future

Colleges and universities seeking to increase the extent to which those in information technology engage with research face challenges in three areas: computation, networking, and data. Of course, these three are intertwined. Computation often involves networking, networking can involve data, and many data problems must be resolved before computation can take place.

#### Computation

In the area of computation, the problem of providing enough capacity not only endures but may be growing as the number of disciplines dependent on high-performance computing increases, the amount of data being created and stored each year multiplies, mature supercomputing facilities fail to maintain their refresh cycles, and the need for access to high-performance computing spreads to the classroom.

Grid computing fills some of the gap, but often campus networks form bottlenecks for data transfer; in addition, not everyone is able to access grid computing capabilities, although the 2005 NSF commitment to further support grid computing nationwide should make a difference in the future. Meanwhile, new techniques being used to maximize the utility of existing resources include the following: harvesting unused cycles available on a single campus; harvesting unused cycles available within the multiple campuses of a research network; harvesting unused cycles contributed to specific research projects by members of the community; and using older computing systems in classrooms where teaching also now involves high-performance computing, rather than discarding obsolete equipment. There are institutional innovations as well. At Princeton University, researchers pooled resources to purchase a supercomputer to be used by all, and at the Virginia Institute of Technology, one thousand ordinary Mac G5s were linked together using volunteer student labor to produce a supercomputer for use by researchers.

Institutional support systems often lag behind the available technological systems. The management of LambdaRail connections, for example, requires a conceptual shift because it is such a different way of organizing activity. There is some concern that unfamiliarity with the management of qualitatively different types of technological systems, such as LambdaRail, may contribute to their underuse. This issue suggests the need to reconsider the types of training available to CIOs and other upper-level administrators, as well as to faculty who would make use of these opportunities.

The problems raised by the effort to build large and long-lived datasets designed for multiple uses are discussed below. Once the data-conceptualization issues are



**Computation often involves networking, networking can involve data, and many data problems must be resolved before computation can take place.**

resolved, however, technical issues remain. For example, in order for computer programs to make use of the concepts in an ontology, there must be a common inference syntax. Standard interchange formats are required to translate the same knowledge into a variety of symbol-level representations. And it is proving to be quite time-consuming to conceptualize the structure of tasks that systems must perform in such a way that problem-solving methods themselves can become reusable. Because there are differences in the types of systems and computer architectures that best serve the resolution of various research problems, the desire to be able to subject data to multiple uses also presents new hardware design constraints.

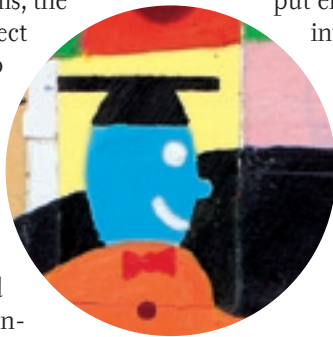
#### Networking

Networking issues include involvement with a variety of national and international grids, attention to the problem of shared interfaces at the user end, management of remote instrumentation and sensors at the input end, and governance problems.

The boundaries of any single institution's research computing infrastructure now include regional, national, and international resources. In addition to computational grids, research involves access grids (for collaboration among distributed researchers), data grids (for accessing and integrating heterogeneous datasets), and sensor grids (for collecting real-time data about matters such as traffic flows and electronic transactions). The fact that the architectures for each of these types of grids are often not contiguous with each other raises additional management problems. Indeed, researchers at a single institution may be involved in multiple access, data, and sensor grids.

Grid computing also creates issues that must be resolved at the points of data input and output. On the input end, there are as yet no consensual norms for how to

manage remote instrumentation and data collected through the sensor web; solutions are being developed on a project-by-project basis. The number of issues in this area continues to rise with technological innovation. Clinical medical practice, for example, is a forerunner of a field in which researchers would like to be able to put information into databases, analyze that information, and query databases for additional information while in the field and engaged in active problem-solving. Wireless networking and the ubiquitous sensor web present problems unique to their environment. On the out-



**The amount of data being created and stored each year continues to grow at exponential rates, with consequences for computing capacity.**

put end, there is a need for user interfaces that are consistent irrespective of computing platform, a matter particularly important in collaborations, since successful distributed research projects require WISIWYS (What I See Is What You See) interfaces.

Just as it will take decades to resolve governance issues raised by the global information infrastructure, so the disjunctures among organizational, political, geographic, network, and research practice boundaries present a number of new governance problems. Organizational structures being discussed and experimented with include management by disciplinary associations, leadership by individual campus IT organizations, expansion of the roles of the

Internet2 organization, oversight by the federal government, and creation of an entirely new organization to manage grid-related matters.

#### Data

Data issues now dominate many discussions about how to manage IT engagement with research: the amount of data that must be handled has grown; the research habit of using only one type of data in a research project has been replaced by the use of diverse datasets subjected to common analytical frameworks; there is

an interest in ensuring a long life for datasets to maximize possibilities for knowledge reuse; new types of data repositories suggest the need for administrator and researcher training in the management of that data; data can be sensitive in a variety of ways addressed by governmental as well as institutional policy; and there is still a paucity of software for the analysis of many different types of datasets. The impact of resolving these issues is expected to be significant, for as the analysis of gene sequences has demonstrated, the ability to mine very large data collections of global importance can open up entirely new paths of research.

The amount of data being created and stored each year continues to grow at exponential rates, with consequences for computing capacity. It has been estimated that in 2003, 5 exabytes ( $10^{18}$  bytes) of new information was created, with 92 percent born digital and 40 percent created in the United States. This is the same volume of information that is believed would result if all words ever spoken by human beings were digitized. Single experiments can now yield as much as 1 petabyte ( $10^{15}$  bytes). It is estimated that the amount of data in all of the holdings of all U.S. academic libraries is 2 petabytes and that, if digitized, all printed material in existence would have a total volume of 200 petabytes. The Large Hadron Collider just coming online at CERN is expected to yield multiple petabytes per experiment, to be used by 2,000 researchers around the world.

The amount of data keeps expanding, for a number of reasons. Many kinds of information originally produced in analog form are being digitized (e.g., manuscripts, transcripts, and visual information). The sensor web is vastly expanding the amount of data available, often in real time. Distributed computing and advanced networking support for the transport of data from one site to another increases the amount of data in circulation and the amount of data accessible from any single location. Often data is generated, transported, analyzed, and used entirely within the world of machine-machine communication, without human intervention. The NSF is trying to stimulate the development of

new uses of data, such as dynamic data-driven simulations that could be used for purposes such as forecasting tornadoes. Codification of tacit information also expands the universe of data. And knowledge reuse creates additional data from the results of research projects. The growing use of simulations and modeling means that data about possible futures also may need to have long lives. Since it is clear that our ability to meaningfully and effectively make use of all of the data available will not keep up with its growth, institutions, disciplines, and the government will need to make policy, infrastructure, and maintenance decisions regarding how much, and which, information to keep.

We're most likely to achieve our long-term data-storage goals if the collections of global importance are managed by just a few institutions serving as "community proxies." These institutions would have responsibility for collection access, collection structure, technical standards and processes for data curation, ontology development, annotation, and peer review. One example of what such a community proxy might look like is offered by the Data Central service recently announced by the University of California–San Diego. Available to researchers at any institution, this service includes a commitment to keep data collections accessible for one hundred years. Other types of data collections may best be handled by disciplinary groups, teams devoted to specific large-scale problems, or the institutions of key researchers. And since research funders are beginning to require that grant proposals give explicit attention to long-term data-storage issues, institutions with grant-funded researchers need to begin addressing this problem immediately.

The functions of researchers as both data authors and data users are tradi-

tional but are receiving renewed attention in terms of training, certification, monitoring, establishment of funded job lines, and assurance of career paths in today's environment. New categories of

professional responsibility are necessary in order to work with long-lived data storage—categories such as specialists in metadata and individuals who can provide an interface between researchers and collections. The University of Michigan is the first institution to step forward to provide leadership in training individuals for these new career paths.

The desire to maximize knowledge reuse has become a much more important factor in the design of libraries, databases, storage collections, and archives. Knowledge reuse occurs in several different ways: conducting a secondary analysis of a researcher's raw data, analyzed either by another in the same field or by people in other fields; revisiting data through new analytical lenses; analyzing data of multiple types through a single analytical lens; synthesizing the results

of many different types of studies for simultaneous analysis of multiple types of data about the same problem; adapting analyses of data for application in new contexts; and visualizing results of data analyses.

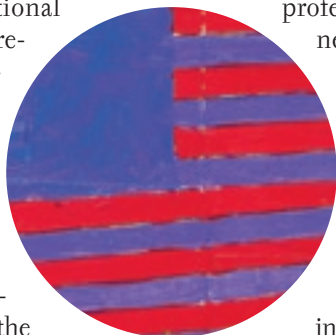
Meeting the needs of knowledge reuse, as well as the shared use of data across disciplines as required by interdisciplinary research, heightens the need for information infrastructure around data. Metadata—data about data—is of growing importance as the number, the size, and the range of uses of data collections grow. Cross-walks are a very traditional way for information scientists to link data from one discipline with data from another discipline. Efforts to reach common ways of thinking about data pertaining to a research problem that is shared

across disciplines builds new ontologies. Some people, such as those involved in the effort to develop a "semantic web," believe it will be possible to discuss data in ways that will be comprehensible to anyone from any discipline, geographic locale, culture, organizational context, or society.

A number of laws, regulations, and policy concerns affect the treatment of data. There is, for example, a tension between the need to centralize data that is being used by multiple parties for multiple purposes and the interest in dispersing data in order to respond to vulnerabilities highlighted by the 9/11 experience and to maximize access. Privacy and security issues are becoming more important, and more complex. International research collaborations generate policy problems when the laws of various nation-states regarding the treatment of data do not harmonize with each other. The growing ability to revisit data collected in the past raises questions about how to handle information collected from cultures or in periods in which there were very different ideas about what constitutes informed consent.

## Conclusion

Over the last few decades, a number of themes that shape the context for IT engagement with research in higher education have emerged at the intersection of national policy, disciplinary developments, the evolution of research methods, and institutional habits: the emergence of computation as a fundamental and distinct stage of the research process, the impact of globalization, the growth in scale of research projects, the trend toward interdisciplinary research collaborations, the recognition of the value of knowledge reuse, and the democratization of research. The tensions generated by these themes are familiar to those in information technology: centralization versus decentralization; the inertia of historical habit versus the need for innovations in practice; the speed of technological innovation versus the speed of institutional innovation; the requirements of the academic institution versus those of external funding agencies; desires of faculty versus mandates for administrators; the needs of the many versus



**The desire to maximize knowledge reuse has become a much more important factor in the design of libraries, databases, storage collections, and archives.**

the needs of the few; and knowledge in service of the public versus development of intellectual property by an institution versus personal career ambitions. These transformations of the research enterprise offer new opportunities for researchers, administrators of higher education institutions, and IT specialists. *e*

#### Notes

1. National Science Board, *Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century* (Washington, D.C.: National Science Foundation, 2005).
2. Jerome A. Feldman and William R. Sutherland, "Rejuvenating Experimental Computer Science: A Report to the National Science Foundation and Others," *Communications of the ACM*, vol. 22, no. 9 (September 1979): 497–502 (the Feldman Report); William H. Press, ed., "Prospectus for Computational Physics," Report by the Subcommittee on Computational Facilities for Theoretical Research to the Advisory Committee for Physics, Division of Physics, National Science Foundation, March 15, 1981 (the Press Report); Peter Lax, ed., "Report of the Panel on Large-Scale Computing in Science and Engineering," Coordinating Committee NSF/DOD, December 26, 1982 (the Lax Report).
3. U.S. Office of Science and Technology Policy (OSTP), *A Research and Development Strategy for High Performance Computing* (Washington, D.C.: Government Printing Office, 1987).
4. NSF Blue Ribbon Panel on High Performance Computing, *From Desktop to Teraflop: Exploiting the U.S. Lead in High Performance Computing* (Washington, D.C.: National Science Foundation, 1993).
5. Lewis Branscomb, ed., *Empowering Technology: Implementing a U.S. Strategy* (Cambridge: MIT Press, 1993).
6. NSF Blue Ribbon Advisory Panel on Cyberinfrastructure, *Revolutionizing Science and Engineering through Cyberinfrastructure* (Washington, D.C.: National Science Foundation, 2003), <<http://www.nsf.gov/od/oci/reports/toc.jsp>>.
7. Francine Berman and Henry Brady, *NSF SBE-CISE Workshop on Cyberinfrastructure and the Social Sciences* (Washington, D.C.: National Science Foundation, 2005), <<http://vis.sdsc.edu/sbe/>>.
8. American Council of Learned Societies (ACLS), *The Draft Report of the Commission on Cyberinfrastructure for Humanities and Social Sciences* (New York: American Council of Learned Societies, 2005), <<http://www.acls.org/cyberinfrastructure/acls-ci-public.pdf>>.
9. Ibid.
10. National Science Board, *Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century* (Washington, D.C.: National Science Foundation, 2005), <[http://www.nsf.gov/nsb/documents/2005/LLDDC\\_report.pdf](http://www.nsf.gov/nsb/documents/2005/LLDDC_report.pdf)>.
11. President's Information Technology Advisory Committee (PITAC), *Computational Science: Ensuring America's Competitiveness* (Arlington, Va.: National Coordination Office for Information Technology & Research, 2005).
12. Lynn Arthur Steen, ed., *Math & Bio 2010: Linking Undergraduate Disciplines* (Washington, D.C.: Mathematical Association of America, 2005).
13. George Johnson, "All Science Is Computer Science," *New York Times*, March 25, 2001.
14. Kenneth G. Wilson, "Grand Challenges to Computational Science," *Future Generation Computer Systems*, vol. 5, no. 2/3 (September 1989): 171–89.
15. Author interview with Larry Smarr, Director of the California Institute for Telecommunications and Information Technology and Professor of Computer Science at the University of California–San Diego, August 4, 2005.
16. Burton Bollag, "Finding Land Mines before Lebanon's Children Do," *Chronicle of Higher Education*, June 17, 2005; Jeffrey R. Young, "Database Will Hold the Mirror Up to 'Hamlet,' with All Commentary on the Play," *Chronicle of Higher Education*, May 27, 2005; Katherine S. Mangan, "Medicine for Musicians," *Chronicle of Higher Education*, October 15, 2004.
17. Marcel Bardon and Kent Curtis, "A National Computing Environment for Academic Research," NSF Working Group on Computers for Research, July 1983.
18. Executive Office of the President of the United States, Office of Science and Technology Policy, "Science and Technology: A Foundation for Homeland Security," April 2005, <<http://www.ostp.gov/html/OSTPHomeland.pdf>>.
19. Peter Lyman and Hal R. Varian, "How Much Information 2003?" <<http://www2.sims.berkeley.edu/research/projects/how-much-info-2003/>>.