11

Future of IT Engagement in Research

Exponential growth looks like nothing is happening,
and suddenly you get this explosion at the end.
—Ray Kurzweil

Futurists, policy makers, and technologists are fond of using revolutionary rhetoric to describe the role, impact, and potential of information technology (IT). On July 10, 2006, Microsoft technologist Gary Flake proclaimed "The changes the Internet is bringing about are every bit as profound as previous historical milestones in the evolution of society—like the Renaissance or the Industrial Revolution" (Flake, 2006). Analogies like these, while possibly overstated, recognize that many of us believe that "high-performance computing and the high-performance networks that interconnect and facilitate information-sharing between the high-performance computing centers are key elements of the nation's 'innovation infrastructure'" (Waters, 2006). With more than a billion people now using the Internet, there can be little doubt of the profound social and economic reach of IT.

Rhetoric notwithstanding, the scientific research applications of high-performance digital computation, networking, storage, and displays have without a doubt had a transformational impact. Scientific research is perhaps IT's greatest triumph. IT has demonstrated the capacity not only to enhance the efficiency and effectiveness of research, but ultimately to transform the very nature of scientific inquiry. Recognizing this, the President's Information Technology Advisory Committee (PITAC) stated in June 2005 that "computational science has become the third pillar of the scientific enterprise, a peer alongside theory and physical experiment" (PITAC, 2005). PITAC cochairs Marc R. Benioff and Edward D. Lazowska advised U.S. President George W. Bush in 2005 that "computational science provides a unique window through which researchers can investigate problems that are otherwise impractical or impossible to address" (Benioff and Lazowska, 2005). Other influential leaders like IBM's Irving Wladawski-Berger argue that "supercomputers enable scientists to either make discoveries that would be difficult (perhaps impossible) to accomplish experimentally or to point researchers in new directions" (Wladawksy-Berger, 2006).

The Engagement of IT in Research So Far: A Real Transformation

In the 60 years since ENIAC and Manchester's Baby were powered up, researchers—particularly scientists—have been very quick to embrace new digital tools. Today, high-performance computers model complex weather systems as well as environmental and astrophysical events. Ultrafast networks and supercomputing techniques have conspired to hasten the mapping of the human genome, birthing new academic disciplines such as proteomics, computational biology, nanoscale science, and many others. Computational techniques are used to simulate traffic flows, chemical interactions, or even nuclear detonations that might otherwise be impossible, too time-consuming, or unacceptably expensive to undertake. These tools and techniques are saving lives, monitoring our environment and health, and protecting property. High-performance computers linked by blindingly fast networks—such as with the Earth Simulator—accumulate and transport huge amounts of real-time data from instruments, sensors, satellites, and other sources to produce models of meteorological and other environmental events that enable temporally and geographically accurate evacuations or other protective interventions. The large-scale simulations also make it possible to model the impact of environmental "events" in ways that facilitate international scientific and policy dialogue on important topics such as global warming. Science and computation have worked hand in glove since World War II. In many ways, this partnership has fulfilled Franklin Roosevelt's vision for science: "New frontiers of the mind are before us, and if they are pioneered with the same vision, boldness, and drive with which we have waged this war we can create a fuller and more fruitful employment and a fuller and more fruitful life" (Roosevelt, 1944).

Rapid improvements in high-performance computation have facilitated major shifts in the very nature of scientific research. One shift has been described as the shift to Big Science. The conduct of scientific research has shifted from 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 data derived from iterations of studies, to
- teams working across laboratory, institutional, and national boundaries, to
- the opening up of entirely new research realms based on both new discoveries and on the existence of new tools and techniques for discovery, analysis, and representation of research data (Braman, 2006).

High-performance computation has also facilitated a fundamental shift in the scientific method (Muggleton, 2006).

Changes to Traditional Science with Automation

Traditional Science	Automated Science
Hypotheses	Machine-encoded logical hypotheses
Chemical knowledge	Machine-encoded chemical algebra
Experiments	Chemical Turing Machine programs
Experimental design	Design theory

Much challenge remains. Notwithstanding computation's extraordinary contribution to research, Vannevar Bush's caveats and admonitions have never been timelier: "There is a growing mountain of research. But there is increased evidence that we are being bogged down today as specialization extends. The investigator is staggered by the findings and conclusions of thousands of other workers—conclusions which he cannot find time to grasp, much less to remember, as they appear. Yet specialization becomes increasingly necessary for progress, and the effort to bridge between disciplines is correspondingly superficial" (Bush, 1945).

So What's in Store?

If scholarship at its heart remains the capacity to build on past knowledge or theory by constructing (manually or otherwise) hypotheses, designing experiments, and gathering, analyzing, and reporting results, then IT will likely focus on the creation of a research cyberinfrastructure that is organized to:

- extend our senses,
- increase our capacity to analyze data and test theories,
- capture, store, and protect our data and research findings, and
- publish our findings (and our tools and data).

More specifically, if IT-enabled scholarship is to continue to grow in quality and quantity, our public policies and investments in research cyberinfrastructure will need to be organized to:

- incent economically optimal choices (for example, to reward scale);
- promote the sustainability of the research cyberinfrastructure, especially the university-based research cyberinfrastructure;
- facilitate and mediate access to instruments and data; and
- foster collaboration.

Indiana University's Brad Wheeler calls this enhanced research capacity Scholarship 2.0, which he defines as the sum of scholarship from archaeology to zoology raised by the power of IT! (Wheeler, 2006). These goals, trends, pressures, or even mandates suggest an action agenda for the future research cyberinfrastructure that includes developing:

- Governance and incentives that promote alignment and reward scale;
- A road map for high-end and sensor-scale (or smaller) computing;
- A research policy environment and funding that promote economic leverage, sustainable infrastructure, and the preservation and reuse of historical investments (including data);
- Resources to promote access to "the totality of our accumulated global cultural heritage now scattered throughout libraries, archives, and museums" (ACLS, 2006).

In higher education terms, MIT President Emeritus Charles M. Vest said it best: "We are seeing the early emergence of a meta-university—a transcendent, accessible, empowering, dynamic, communally constructed framework of open materials and platforms on which much of higher education worldwide can be constructed or enhanced" (Vest, 2006).

Extending Our Senses

Like all research instruments, those associated with the conduct of research serve to extend humans' capacity to sense and probe the physical environment (experiment) and increasingly to situate information into intellectual constructions and thereby raise information's explanatory value (create theory and knowledge). These tools and techniques as well are serving to render unimaginably large data sets in forms that humans can comprehend (visualization and presentation) and communicate with others. In these fundamental aspects, we can reasonably assume that the past will be prologue. By that we mean that science and science fiction will continue to become increasingly indistinguishable from one another as computational tools and techniques drive increasingly toward improving on the human being. Computational devices and techniques will be used with growing frequency to see farther, probe deeper, and collect more widely.

If indeed one key purpose of computational instruments is to extend our capacity to sense the environments we wish to understand, the future of research and computation will likely be at once a continuation of past trends, and nearly unimaginably magical. The major elements of this research infrastructure include the following elements.

The Network

Assuming a major goal of the emerging cyberinfrastructure is to extend humans' capacity to sense the physical and biological world, then the nature, size, capacity, and security of the network remains a matter of central concern. If computers are the distributed intelligence of a global metalaboratory, then surely the network that links them is the cortex of the research enterprise. Science fiction writer Vernor Vinge speculates that in 15 years, we are likely to have processing power that is a thousand times greater than today, and an even larger increase in the number of network-connected devices. He maintains that "the Internet will have leaked out to become coincident with Earth" (Vinge, 2006).

To continue to globalize research and to eliminate barriers of distance and time, networks will need to be faster. In the field of medical research, for example, the newest ultra-high-voltage electron microscopes produce network bandwidth requirements that approach 100 gigabits per second. It is expected that the TeraGrid will require sustained data flows of 1 terabit per second between seven U.S.-based supercomputing centers (Waters, 2006). To optimize network throughput and data quality, most technologists anticipate the widespread acquisition and use of dedicated optical channels—lambdas—by researchers.

In the decades ahead, networks will also need to reliably and securely connect up to a thousand times the number of intelligent devices that they do today. Much of this connectivity will be wireless, demanding significant jumps in the performance and security of wireless networks. In addition, much of this connectivity will be among sensors and other embedded devices, posing new challenges and complexity for routing, securing, and prioritizing data flows.

Big Data Collectors (Digital Instruments)

The propensity and capacity to conceive of and build complex instruments to extend sight and hearing is a distinguishing feature of our species. This propensity will continue to find expression in the future. Importantly, all instrumentation in the future will be digital and all will be connected via networks. The Large Hadron Collider (LHC) at the CERN particle physics lab will go live in November 2007. It is a collaborative effort among more than 2,000 physicists in 34 countries. Digital particle detectors like the Compact Muon Solenoid (CMS) incorporate millions of silicon microstrips and tens of millions of pixel elements to track and record the collisions, paths, and decays of protons. Complex instruments that are expected to collect petabytes of data are under development in physics, atmospheric sciences, geoscience, space sciences, and life sciences. Developing and deploying instruments of this kind often requires decades of research and development. Big Science projects involve hundreds or thousands of scientists and engineers, and they depend on complex interactions and investments among governments, universities, and private firms. They are often global in nature.

Small Data Collectors (RFID, sensors, telemetry, motes, nanodevices, etc.)

While the future engagement of IT in research will depend on the massive investment of governments and others in a global (and extraglobal) complex of scientific instruments, the continued miniaturization of thinking machines is likely to exert profound changes on the course of both research and of IT management. Passive radio frequency identification (RFID) chips measure just 0.15 mm x 0.15 mm, cost less than US$0.05 each, and can be read from a distance of four inches. They are being used chiefly to track inventory. Active RFID tags have their own power sources and can be read from distances of 300 feet. These tags have memory and when married to sensors can transmit and receive steady streams of research data at very low costs. RFID and telemetry are now being used to track wildlife in biology population studies. Sensors are small devices or biological organs that detect, or sense, signals, physical conditions, and chemical compounds. Sensors have long been used in research, but their uses are expanding exponentially as they become simultaneously more robust and more economical. Sun Microsystems, for example, is developing commercially available sensors that feature a 32-bit processor with 256K RAM and 2 MB of flash memory. These devices are battery powered and equipped with telemetry and security software as well as light and temperature sensing capacity, and they are smaller than a credit card. When coupled with telemetry, network standards, security, and middleware, sensors like these will be increasingly used to detect, store, and transmit research information from seismically active areas, biological systems, transportation systems, human and animal systems, and so forth. These sensors can measure temperature, pressure, electrical and magnetic fields, chemical levels and behaviors, and more. In the future, sensors will likely be of an increasingly biological nature. Malikarjun Shankar, a sensor Web expert with the Oak Ridge National Laboratory sums it up: “This is where computer science hits the tangible, the palpable world. It is the next frontier” (Butler, 2006).

The low cost of sensors will of course facilitate their widespread adoption. This is likely to stimulate research production while simultaneously raising issues regarding complexity, standards, security, and privacy.

Synthetic and Virtual Worlds

Synthetic and virtual worlds are computer-simulated environments that have been created for users to interact with each other through avatars. Several of these are MMORPGs (massively multiplayer online role-playing games) like *EverQuest* and *World of Warcraft*. These virtual environments, constructed for entertainment, host millions of players and revolve around competition and cooperation. Increasingly these games are attracting the attention of scholars who are being presented with the unprecedented opportunity to study collaborative and competitive social systems on a large scale in real time (Castronova, 2005).[1]

Visualization Techniques and High-Performance Displays

The human ability to comprehend is strongly tied to our ability to visualize. While it is well established that multiple intelligences exist, such as logical-mathematical, spatial, and linguistic, for many of us seeing really is believing (Gardner, 1983). Huge progress in the development of rendering engines and in rendering technique is making it possible for researchers to display gigabytes of data—spanning multiple variables, time scales, and so forth—in ways that they and other researchers can understand, discuss, and explain. In this way, data that describe subatomic or cosmological events that are otherwise completely inaccessible to the human senses become comprehensible. Facilities like the Synthesis Center of the San Diego Supercomputer Center (SDSC) provide unique, wall-sized, high-definition displays to help scientists from a variety of disciplines gather, visualize, discuss, and interpret their large-scale data. Future breakthroughs in rendering, display, and networking will make it possible for researchers to gather virtually across very fast networks to observe large data sets and to manipulate them in real time. High-definition displays will also make it possible to substitute simulations for economically or technically infeasible research alternatives and to exploit the full value of large-scale data sets that have already been collected but not yet visualized. At the same time, everyday displays are likely to change radically. Inventor Ray Kurzweil predicts that "by 2010, computers will disappear. They will be so tiny that they will be imbedded in our environment, our clothing. We will have eyeglasses … that display images directly in our retina: contact lenses for full immersion virtual reality" (Kurzweil, 2003). Developments like smart paper will likely also promote mobility in future research activities.

Increasing Our Capacity to Analyze Data

Computational Power

Moore's law predicts that computational capacity (as defined by the number of transistors per square inch on integrated circuits) will double every 18 months at a constant cost. Most computer scientists expect this prediction to hold for at least two more decades. In the meantime, successor technologies such as optical computers or even quantum computers are being researched and developed. Quantum computers make direct use of distinctively quantum mechanical phenomena, such as superposition and entanglement, to perform operations on data. The first working quantum computers based on nuclear magnetic resonance were developed in 1998. Even today's brain power is hard to comprehend: IBM's Blue Gene/L supercomputer, for example, processes (at peak) 360 trillion floating-point operations per second. In 2005, the Blue Gene line was expanded to include Blue Brain, an effort to model the human neocortex.

More computational power in the future means simply more research, faster research, and (likely) better research. Life scientists will complete the genomic sequences for mammalian and other species more quickly than planned, and they are turning their attention to measuring micro RNA and protein expression levels and will ultimately identify the markers of potential disease long before such diseases express themselves. And while it took a decade in the 1990s to determine the sequence of one human genome, computers in 2050 are expected to be able to sequence the genome of every individual on earth (Muggleton, 2006). In the future, nanodevices will be able to enter cells to perform repairs. For this, scientists will need to calculate atomic level quantities. This accomplishment is a major goal of nanoscience, but it is beyond our current computational capacity. More computation also makes it possible to "jump" research innovation curves, often with significant changes in cost structure and in performance. For example, high-performance computers are now being used to assemble data from "software telescopes" that produce unprecedented resolving power by using thousands of low-cost radio antennas spread across large geographical regions.[2]

The existence of faster computers in the future will be complemented by two other important trends in computing. First, computers will continue to get smaller. Many of these devices will be sensors or even nanoscale devices. The proliferation of small intelligent devices will allow researchers to monitor nearly all biological, physical, and even psychological and social systems, creating unparalleled prospects for research breakthroughs while simultaneously stimulating the exponential growth of research data and the attending problems surrounding its management, curation, processing, use, and preservation. The annual doubling of scientific data has profound consequences. As two scientists speculate, "with data correlated over many dimensions and millions of points, none of the old steps—do experiment, record results, analyze, and publish—is straightforward…. Many [predict] that few traditional processes will survive in their current form by 2020" (Szalay & Gray, 2006, p. 413). Second, it is likely that computing power in the future will be a network service. As progress with grids and virtualization of computing continue, it will be possible and desirable for researchers to call upon a variety of high-performance tools as the need arises. In this context, virtualization is meant as the process of providing a logical grouping of computing resources that is not constrained by the physical configuration, geographic location, or specific implementation of the underlying computing resources. In some cases, campus, regional, and national computational resource managers will work to harvest unused computing cycles to parse computational tasks, memory, and data among millions of underutilized personal computers as has been done successfully at institutions like Purdue University, UC Berkeley, and with projects such as SETI@Home.

Databases

Our ability to undertake the complex analysis of distributed and multiscale research data sets that are terabytes or even petabytes in size is also a function of the quality of our computational algorithms and our capacity to organize large data sets and find patterns in them. These demands can only be satisfied with ongoing improvements in databases. In fact, the problem today is serious. Szalay and Gray (2006, p. 413) report that as computers increasingly collect the data and perform experiments on data, it is difficult to "capture all the model numbers, software revisions, parameter settings and process settings in an enduring format." This means that experiments are becoming increasingly hard to document and nearly impossible to reproduce.

Increasing Our Capacity to Manage and Protect Our Data and Knowledge

Szalay and Gray (2006) refer to the simultaneous proliferation of data from large-scale instruments and small-scale sensors as "scaling up and scaling out." Researchers refer to a data deluge. The growth in the rate of data production is not limited to scientific endeavors. Indeed Indiana University reports that "In the humanities, new uses of information technology are similarly generating TeraBytes of multimedia data at an accelerating rate" (Indiana University, 2005, p. 5).

Data Management

In the future, in addition to needing enhanced databases, the research community will need considerably better practices and standards for moving data between instruments, sensors, and storage systems on a global basis, as well as enhanced capabilities for managing the long term "annotation, curation, provenance management, and archiving of data" (Indiana University, 2005, p. 4). In the future continued attention will be paid to issues of research vocabulary management including semantics, natural language processing, vocabulary authorities, ontologies, and others. These activities are becoming pressing as research problems increasingly span academic disciplines, each possessed of a unique language, culture, standards (or not), and scholarly communication convention. The need for such practices and standards increases in importance as huge efforts are under way to link and thereby leverage historical datasets. The Virtual Observatory Alliance, for example, is integrating most of the world's astronomical archives, but it is constrained by the need to adopt standard language to make searching and exchanging this data possible. There is compelling evidence that managing research data will be further complicated by the complexity of the data itself. Proteomic data (as well as astrophysical and others), for example, are three-dimensional, requiring the deployment of so-called data cubes to preserve and store data in three dimensions.

Knowledge Management

While the economic productivity of storage devices has followed a slope consistent with that of semiconductors, our efforts to manage and preserve "knowledge" have not. Whether this is about (1) the systematic processes by which knowledge is leveraged for an organization to succeed (Rumizen, 2002), (2) augmenting or enhancing the completeness of a participant's ability to interpret information within a context (Voeller, 1999), or (3) the art of creating value from intangible assets (Sveiby, 2001), much can and must be done to manage the mountain of accumulating research data, to place it under intellectual control, to ascertain its provenance, and to manage it both within prevailing theoretical constructs and for the long term. The metaphor for such robust management of research data is, of course, the library. Charles Severance of the University of Michigan offers some organizing principles and attributes for an idealized research data management system:

- Generally one (data management) system for the [research] area;
- Long-term strategic choice for the institution;
- System focused on accessing, indexing, curation, and storage;
- Millions of high-quality objects properly indexed;
- Data and metadata quality;
- Must enforce standards and workflow to insure data quality; and
- Most use is very purposeful: search, publish, and add value. (Severance, 2006)

Publishing Our Findings, Tools, and Data

Of course, managing research data is not enough. Vannevar Bush reminds us that "Mandel's concept of the laws of genetics was lost to the world for a generation because his publication did not reach the few who were capable of grasping and extending it; and this sort of catastrophe is undoubtedly being repeated all about us, as truly significant attainments become lost in the mass of the inconsequential" (Bush, 1945). The term *publishing*, in fact, is a misnomer in the research context and refers only to that small aspect of scholarly communications that plays itself out through formal processes of peer review and official dissemination. Indeed, access to research information and scholarly communication—broadly defined—really represents more fully this area of future speculation.

Middleware

Promoting effective access to research data and other evidentiary information integral to the establishment of the validity, credibility, or repeatability of research is central. Access must be global since the community of scholars operating within any academic discipline is a global community. Regulating access to research data, to research instrumentation, and to scientific and other research networks and resources is therefore of paramount concern. In the future it is likely that the standards of authentication for researchers that are imposed by their home institutions will rise to levels imposed by the sponsors of research and the operators of research grids and collaboratories. The authentication of institution-based researchers against externally imposed standards will make it possible for research federations to prosper, obviating the need for the authenticity of researchers to be repeatedly confirmed at each research facility they invoke or with each networked research instrument they use. While there remain technical issues associated with both hard authentication and with federated identity management, the issues surrounding these future developments are likely policy and cultural issues. Hard authentication technologies like biometrics, for example, are already in wide use among state and federal U.S. first responders associated with homeland security. These technologies meet the highest standards for efficacy. Nevertheless, biometrics are not 100 percent reliable, and it is possible to spoof them or counterfeit results from retinal scans, fingerprints, and the like. Such imperfections raise anxieties and doubts about their widespread uses. Too, the academy has long cleaved to higher—perhaps unnecessary—standards and values as regards the protection of student and employee privacy. No doubt these beliefs and cultural standards will constrain and ultimately circumscribe higher education's approaches to managing identity. This said, the stakes are high and rising. In particular, as devices become simultaneously smaller, cheaper, and more powerful (for example, sensors), we will find them in widespread, and potentially unregulated, use. Imagine the risks posed by spoofing data in sensor networks that serve to advise researchers monitoring seismic activities, violent weather systems, tsunamis, and the like.

The Research Library

Vannevar Bush also reminds us that "a record—if it is to be useful to science—must be continuously extended, it must be stored, and above all, it must be consulted" (Bush, 1945). Bush goes on to describe the need for a memex, which is a device "in which an individual stores all of his books, records, and communications, and which is mechanized so that it may be consulted with exceeding speed and flexibility. It is an enlarged, intimate supplement to his memory." The idea of a unified compendium of the human written record is a longstanding one. It is a vision that may be within our reach in the future. In August 2006, the American Council of Learned Societies issued the report of their Commission on Cyberinfrastructure for Humanities and Social Sciences. This report identifies a single grand challenge (with numerous tactical implications) for the humanities and social sciences: "to create an integrated digital representation of the cultural record, connecting its disparate parts and and making the resulting whole more available to one and all, over the network" (ACLS, 2006, p. 15). Of course the creation of such as resource is audacious in its magnitude and would demand investments in search engineering, natural language processing, semantics, storage, databases, and on and on. Nevertheless, the point is that information technologies are now poised to fulfill the ultimate dream of many in the academy. To even begin to decompose such an immense vision with an eye toward implementation is indeed a tribute to the state of IT and its contributions to research.

E-Collaboration

Much of this chapter suggests a continuation of the linear and exponential historical trend lines with respect to the dominant piece parts of the research cyberinfrastructure. The question that needs to dog our scientific and policy leaders is this: "Is the sum of continued advances in computation, storage, networking, visualization, e-collaboration, and so forth, sufficient?" The likely answer to this question is no. Indeed, the phrase that crystallizes our hope for research opportunities on the horizon is "collective intelligence." It's about how we organize thousands of people and institutions and millions of intelligent devices to solve problems quickly, accurately, and creatively. Grid.org, for example, provides a platform for large-scale nonprofit research projects that includes services for sharing computational power, high-bandwidth networking, and data storage. The Eli Lilly spin-off, InnoCentive, is another example. InnoCentive is a virtual and global R&D dating service. It creates a community—really a marketplace of science—linking together scientists across the world through shared financial incentives to solve problems and challenges posed by sponsoring companies. Wikipedia, of course, is created by thousands of people who are not getting paid but who produced the world's biggest and arguably one of the best encyclopedias in less than five years.

The future is likely to witness a redoubling of effort to understand and improve the nature of e-collaboration in the research domain. The notion that significant research breakthroughs occur at the interstices between disciplines is thoroughly socialized within the research community. So too is the idea that research collaborations in the ideal must include researchers from disparate institutions; researchers who are united by a common passion for the topic and who possess complementary research capabilities. And yet, the evidence seems to suggest that multi-site research projects achieve lower outcomes than others (Cummings & Kiesler, 2005). These findings suggest the need for tools and environments that foster effective communications and other elements of collaboration. Severance (2006) describes the attributes of systems that might constitute such a collaborative environment:

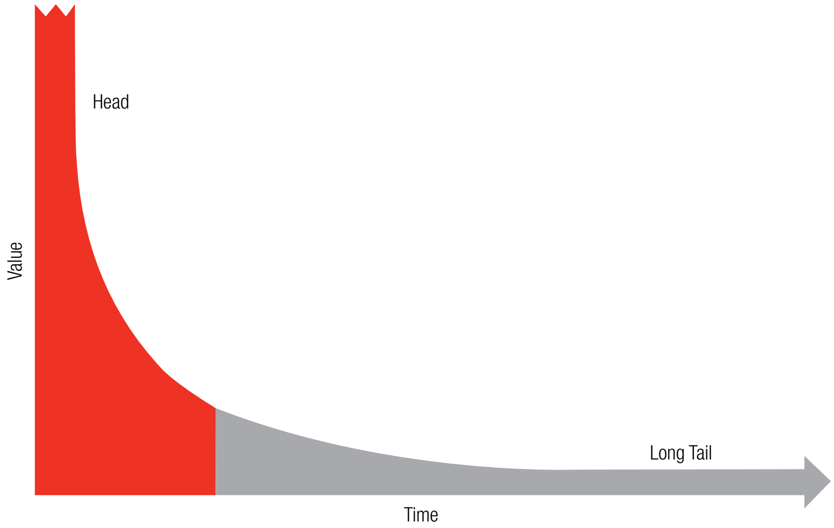
- Many different systems may be active at the same time.
- Systems evolve, improve, and are replaced every few years.
- Systems focus on the dynamic needs of users and applications.
- Systems support thousands of simultaneous users.
- Performance tuning of such systems must be undertaken continually.
- Systems must be so easy to use as to be "unnoticeable."
- Systems of this kind will be used informally hundreds of times per day.

Intellectual Property

This topic deserves its own research study. Contemporary researchers live in a cyber world where the watchwords are cut, paste, rip, remix, and share. Many longstanding Western standards of authenticity, credibility, and identity are under strain within the new communication milieu. At the same time, the network has extended the value of information by virtually eliminating the costs of storing inventories of books and journals. This phenomenon has been described as the long tail, referring to the notion that the value (economic or scientific) of information "is in the millions of niche markets at the shallow end of the bit stream" (Anderson, 2004) (see Figure 11-1).

These changing practices, techniques, values, and economics push Anderson to draw powerful recommendations for owners of content in the digital era. He characterizes these recommendations as rules:

Figure 11-1. Time Value of Information



1. Make everything available.
2. Cut the price in half, now lower it.
3. Help me find it.

In essence, it is reasonable to conclude that constantly improving digital technologies are having numerous disruptive effects on industries that can be described as "content industries." While scholars do not typically think of markets for their work or of scholarship as a content industry, indeed it is, and the pressures above to cut, paste, rip, mix, open up, slash prices, and render accessible are running headlong into changing copyright laws that serve to extend the rights of copyright holders chiefly out of publishers' awareness of the emerging long-tail economics of publishing. In essence, published works have greater value over a longer term because the cost of supplying them in digital form is lower than the cost of keeping them in digital storage.

The future is likely to witness sharpening battles over intellectual property as publishers' legal and technical ability to control and repurpose content collides with scholars' needs for open discourse and low-cost access to research data and publications.

Swimming Toward the Future Against an Economic Tide

No discussion of the future research cyberinfrastructure is meaningful without taking at least a passing glance at the economic environment facing the research community and enterprise. In this it is perhaps only slightly too dramatic to suggest a climate that could be described as a big bang. A collision is likely in store between the growing demand for research investment on one hand and the decreasing capacity or willingness of mature, first-world economies to finance it. Demand for research funding will be stoked simultaneously by the rising cost of research projects and the aging of the developed world and China. Rising research costs are chiefly a function of complexity: as the problems we aim to solve get harder, the cost of solving them rises. The aging of much of the planet puts demand on the scientific establishment to deliver new medical and pharmaceutical breakthroughs, but it also puts stress on much of the world economy as public funds are pressed into serving other welfare needs of the elderly. In the United States and elsewhere, huge budget deficits conspire with these structural demographic pressures to dampen the supply of available research funds.

In a context of sharpening competition for research funds, or worse, in an era of possible absolute declines in funding, it is essential to optimize the resources that are available. In many cases, such optimization suggests enhancing our capacity to increase the scale of research.

Enhancing our capacity to increase the scale of research or even more audaciously, implementing the vision of collective intelligence, is less about making sustained progress on the speed and performance of elements of the cyberinfrastructure (though it is assuredly about this) and more about the unification and harmonization of those elements. In this regard, the challenge of this vision is a familiar one to any technologist or technology policy maker over the age of 50. Creating the future cyberinfrastructure is perhaps best understood as a trip back to the future. In the 1970s, computing became something more than simply an amazing and interesting anomaly. By the early 1980s, computers—while mostly not interconnected—touched many of us. From the 1970s through the 1990s, a series of actions were taken to unify and harmonize thousands of disparate devices operating on disparate networks. The Internet emerged from a patch quilt of campus and regional networks unified through the National Science Foundation (NSF) backbone network to a global, distributed, public utility that performs incredibly well, in some ways. And while the

Internet has many clear deficiencies, it has attracted a global user base of more than a billion people and performs with considerable reliability and predictability. Creating a research cyberinfrastructure bears many of the hallmarks of the past harmonization of the Internet. There exists in the world a vast and growing array of digital telescopes, microscopes, simulators, earthquake shaking tables, sensors, computers, grids, and other research resources that will need to be stitched together.

Of course this massive stitching together is incredibly complex. Notwithstanding this complexity, once the key to the future is understood to be harmonization and unification, one can begin to deconstruct the stack of challenges associated with such a problem and to apply solutions across different elements in that stack.

Cyberinfrastructure's Key Challenges

In light of the economic tide that is likely to push against the emergence of an integrated research cyberinfrastructure, policy makers, education leaders, information technologists, researchers, and others will need to address a number of key challenges.

Foster Economically Optimal Decisions Through Changed Governance and Incentives

Elements of the cyberinfrastructure today are owned and operated by a wide variety of government entities (of different nations, of course), universities, private firms, and individuals. Cyberinfrastructure, like the Internet, can hope only to be a very loosely coupled federation. That said, one question and challenge for the future might be, So what are the governance counterparts—in cyberinfrastructure—to the IETF, WC3, ISO/IEC, ICAAN, and other international bodies concerned with promoting standards? If the research cyberinfrastructure is to evolve into a loosely coupled federation, who will draft and oversee the articles of confederation? This problem seems to demand a unified voice from the government sector, and facilitating alignment among government scientific policy makers will not be easy. Nonetheless, this task should reside on the list of future tasks. The problem of governance extends through all layers of the cyberinfrastructure. Grant recipients at all levels of U.S.-based research respond to conflicting mandates and incentives from different research sponsors. Government labs develop proprietary networks and databases, and they operate instruments under different agency rules. Scientific consortia construct technical designs, and implement standards, operating procedures, terms and conditions of use, metadata, and data exchange protocols that preclude easy interconnection or interoperation. Campus-based principal investigators spar with central campus IT service providers over control of physical resources, energy management, technical standards, IT security and privacy, and so forth. The reinforcement—through technology—of bunkers within academic disciplines is rampant as everyone engages in local suboptimization. As a result, at a time when we believe most in the value of cross-institution research, we struggle with the effects of massive coordination challenges. Duke University Professor Jonathon Cummings concludes, "Projects with PIs from more institutions were significantly less well coordinated and reported fewer positive outcomes than single institution projects" (Cummings, 2004, p. 2). Bunker construction, or more positively spun, local suboptimization is likely an artifact of the incentives inherent in the current research endeavor and environment. In higher education, concern over incentives is growing. On group gathered recently under the auspices of the NSF, Internet2, and Pennsylvania State University concluded that "major negative reinforcements exist in the current environment … grant approaches at several major agencies seem to favor 'autonomous, small clusters in closets' over more sustainable and secure resources" (Klingenstein, Morooney, Olsansky, 2006).

A Road Map for High-End Computing

It has been often observed that if you don't know where you're going, you'll probably end up somewhere else. The U.S. federal government has, despite duplication and fragmentation, been particularly good at identifying grand challenges that could benefit from high-end computation such as nanoscale science, climate research, and genomics. The U.S. federal government has also been effective at articulating the technical components needed to enable scientists to make progress in areas of immense challenge (High-End Computing Revitalization Task Force, 2004). The essential message here is that there is general alignment on a vision of how high-end computation or research cyberinfrastructure can promote 21st-century research. There is also considerable agreement on the technical standards, architecture, components, and institutional elements of a research cyberinfrastructure. Therefore in many ways, the complexities of creating a technical roadmap to guide the ongoing development of the research cyberinfrastructure are widely understood and are solvable. What prevents concerted progress are politics and money. Differing and historical philosophies, stakeholder expectations, and power bases continue to challenge us in executing a shared vision of the research cyberinfrastructure. Too, the private sector and the university sector have not been brought fully into this roadmapping activity, leading to missed opportunities for leveraging federal investments and to probable holes and weaknesses in the cyberinfrastructure. The missing university voices are often the voices not of the scientific community but of those charged with providing a campus research cyberinfrastructure.

Funding for a Durable Cyberinfrastructure

Science and engineering research performed at U.S. colleges and universities in 2003 (the most recent year of reporting) exceeded US$40 billion (National Science Foundation, 2003). Thirty percent of this research was performed at 20 universities. In 2003, respondents to the NSF survey of research and development expenditures at universities and colleges also reported spending more than $1.4 billion on research outside scientific and engineering academic disciplines (see Table 11-1). While this research spending is by no means wholly or even dominantly comprised of spending on IT, the spending on IT is not insubstantial. And, as more and more academic disciplines' research capabilities come to be dominated by computation, the IT spending in pursuit of research will continue to grow as a percent of overall research spending. This college- and university-based spending on research is supplemented by enormous federal spending on the research cyberinfrastructure in the form of support for national laboratories, supercomputer centers, and specialized interdisciplinary efforts like the National Ecological Observatory Network (NEON), the Sloan Digital Sky Project, the Laser Interferometer Gravitational Wave Observatory (LIGO), the Network for Earthquake Engineering Simulation (NEES), and others.

While a significant portion of the funding for the research cyberinfrastructure arises through direct investments from the NSF, NIH, and other research sponsors, considerable funding derives from highly decentralized, complex, and widely varying practices on university campuses involving principal investigators, chairs, deans, business officers, CIOs, and provosts. Investigators, of necessity, want to put as many resources as

Table 11-1. R&D Expenditures in All Fields at Universities and Colleges, FY 2003 (in millions of 2006 dollars)

Field	Dollars
Science and engineering (S&E)	40,077
All non-S&E	1,371
Business and management	165
Communications, journalism, and library science	55
Education	597
Humanities	135
Law	41
Social work	56
Visual and performing arts	39
Other non-S&E fields	241

Source: NSF Division of Social Science Resources, Statistics, Expenditures at Universities and Colleges, FY 2003

possible directly into their experiments. They will starve the infrastructure if such behavior does not diminish their research. When they do invest in infrastructure, it is too often in the "autonomous, small clusters in closets" that are managed to ensure direct accessibility and control by the investigators. Technology management practices in this highly decentralized kind of environment range from gifted amateurism (for example, support and maintenance by very bright graduate students) to highly professional. Rarely is the locus of laboratory practice the creation and maintenance of sustainable and sharable networked resource. The focal point of the grant is the investigator and the organizational and programmatic locus of the investment in technology is the laboratory and the experiment. When the experiment is completed, the hardware, software, and other elements of a potentially valuable (and extensible) institutional cyberinfrastructure are deployed in local and often idiosyncratic ways. Importantly, the same can also be said of the treatment of research data sets acquired or produced through this localized system of research. In defense of historical localization of research cyberinfrastructure, researchers and CIOs will quickly agree that until recently, few central IT organizations were well positioned to serve local research needs.

To take cyberinfrastructure to the level anticipated in the Atkins Report (National Science Foundation, 2003), a new compact will need to be formed among principal investigators, their departments and schools, their universities, and the sponsors of their research. An infrastructure is only as strong as its weakest link, and the current funding model will not yield a research cyberinfrastructure that is durable, shared, resilient, or secure at the institutional level. Decker and Neas (2003) focus on the opportunity to rebuild the partnership between researchers and central campus IT providers. They outline a series of questions that have only risen in relevance since they were written:

- What opportunities exist to reduce personnel expenses through appropriate partnerships between central IT and the research groups?
- Are there useful approaches to handling common licensing, distribution, and maintenance tasks and costs?
- How can IT facilities be cooperatively designed and supported to maximize utility and reduce cost to the institution for the long term?

- What are the best and most cost-effective models for research administration support systems?
- Is the overall research and IT environment designed and protected to manage security vulnerabilities?
- What role can central IT organizations and campus CIOs play in helping institutions mitigate growing reporting and information sharing demands? What infrastructure implications may derive from new policies in these areas?

Federal research sponsors, research leaders, and higher education leaders will need to raise similar issues in regard to different links in this partnership chain.

A Research Policy Environment That Promotes Openness, Accessibility, Economic Leverage, and Sustainability

The American Council of Learned Societies articulated well the "necessary characteristics" of the envisioned research cyberinfrastructure (ACLS, 2006). The future research cyberinfrastructure, according to the ACLS criteria should:

- be accessible as a public good,
- be sustainable,
- promote interoperability,
- facilitate collaboration, and
- support experimentation.

While the achievement of such goals will make significant technical demands on cyberinfrastructure in areas like the management of security and identity, it will require especially the development of public policies that foster these characteristics. In particular, the future research cyberinfrastructure must aim to be inclusive in the broadest possible sense. To do this, public policies and investments must be targeted to stimulate the emergence of a research enterprise that leverages not only a global installed base of networked instruments and smart devices but also the collective, and in many cases surprising, engagement of many in research. To accomplish this goal, universities and policy makers in the future may need to examine fundamental questions about the nature of authentic research and the nature of researchers. For just as tools like blogs and wikis are redefining our notions about professionalism and expertise in journalism and beyond, so the emerging cyberinfrastructure will challenge many historical divides. Divides include:

Academic Disciplines

As mentioned, only about 3.3 percent of college and university expenditures on research are made outside scientific and engineering disciplines. For the most part this is due to the monetization of research by federal and other sponsors. In fact, the proportion of research by humanists and social scientists is likely much higher, but it is not accounted for as this expense is subsumed under the broad category of faculty salaries. Nevertheless, a large gulf in research spending does exist between the sciences and the humanities and social sciences. The recent report of the ACLS provides a roadmap for reducing this gap and thus for stimulating similar breakthroughs in human understanding of social systems, our history, philosophy, art, and culture.

Undergraduates and Graduates

In citing U.S. President George W. Bush's 2006 State of the Union address announcement of an initiative to "encourage innovation throughout our economy," Arthur Ellis (2006) in a recent editorial called for the creation of a "culture for innovation." In particular, Ellis calls for an escalation in higher education's mentoring of undergraduate students in research. Through the NSF's Undergraduate Research Collaboratives (URC) program, colleges and universities are creating research-oriented curricula for first- and second-year undergraduates. Undergraduates in these programs are *using* the research cyberinfrastructure. Importantly, the URCs are lowering other barriers to a culture for innovation by extending benefits to students at two-year institutions. Ellis correctly observes that "those students represent a huge, diverse, largely untapped talent pool. By involving them and their instructors in research, we can both build our institutional capacity for innovation and encourage talented students …" (Ellis, 2006, p. B20).

Amateurs and Professional Scholars

Perhaps the most remarkable instance of the role of amateurs in research is the famous story of Swiss patent clerk Albert Einstein, who had the audacity to submit three articles for consideration to the eminent journal *Annalen der Physik.* While the academy, at its worst, guards the gates of scholarship vigorously, at our best we welcome research and innovation regardless of the source. Of course, the *Annalen's* editors were quick to realize the genius in Einstein's papers and published them together and at once. The divide between professional and amateur scholars continues to exist and poses a real barrier to the creation of a culture for innovation. Participants at the 2006 Wikimania conference celebrated that encyclopedia's one millionth article. However, scholars and amateur encyclopedists sparred over professional standards including "often-indifferent prose" and debated the existence of a "hierarchy of knowledge" (Read, 2006). The potential of collective intelligence in service of research and innovation cannot be understated. Today, amateur astronomers partner with scholars to increase our coverage of the night skies beyond anything the academy could contemplate on its own. And as sensors are becoming available, amateurs will form communities of practice to monitor soil, water, and atmospheric conditions, track pathogens in the environment, measure health statistics, and so forth. The challenge for cyberinfrastructure is to recognize the potential of amateur research and to provide such communities with access to scientific and cultural resources. As with CNN's I-Report, which harnesses the time and talent of home videographers and reporters, the challenge for the academy is to leverage the collective intelligence of these communities to support meaningful academic purposes and standards and to incorporate gifted amateurs effectively into professional scholarly discourse.

Research-Intensive Universities and Others

The vision of the future cyberinfrastructure as a richly interconnected network of people, institutions, instruments, and so forth leverages the network's capacity to render location meaningless. At speeds of 10 gigabits per second and above, the location and academic affiliation of the research are rendered meaningless. This suggests that academic researchers in the future may have enhanced choices about how and where to spend their careers. Today, most scientists who use high-performance computation and networking work, of necessity, at research universities. As computing cycles become network resources and as dedicated optical channels become more widespread and affordable, some researchers will chose to work in BA institutions, research institutes, or elsewhere in the academic ecosystem. Today, despite the fact that most of the faculty in America's associate's, bachelor's and master's universities have earned terminal research degrees from the top 50 research universities, few of these highly trained researchers conduct research. The fulfillment of the NSF and ACLS visions of cyberinfrastructure creates the potential to return these professionally trained researchers to the fold. Such a return would go far toward creating a culture for innovation, but it would depend on making access to resources available broadly and on providing research funding for individuals at institutions not traditionally recognized for research.

Conclusion

Scientific research has transformed the world and information technologies have transformed scientific research. Research in the humanities and social sciences has been less affected by the revolutions in computation and networking, but it is poised to undertake the integration of humankind's cultural record using the research cyberinfrastructure. Investments in the traditional elements of cyberinfrastructure such as networks, servers, systems administration, support, and storage must be continued. New investments in security and middleware must be made. The frontiers of cyberinfrastructure will be extended through improvements in these areas, but perhaps even more, through improvements in the areas of data management and e-collaboration.

The daunting task of higher education and its research sponsors is to foster a culture for innovation. Such a culture depends—at its core—on the cyberinfrastructure of people, tools, and institutions described by NSF's Blue Ribbon Advisory Panel in 2003. The panel was correct in defining cyberinfrastructure more broadly than the complex of scientific instruments and instructional tools, for the chief barriers to the realization of this vision are economic, political, and cultural. Economically, new funds will always be needed to propel research forward. As has long been the case, such funds need to be viewed as investments in the common good. More important, if the vision of a future research cyberinfrastructure is to be realized, the fabric of public policies and economic incentives must be changed. Those who fund research must consciously and aggressively calibrate all incentives to promote accessibility, openness, economic leverage, and sustainability of the cyberinfrastructure and of research tools and data. In particular, a new partnership between federal research sponsors, the network of federal laboratories and research centers, and research universities should be contemplated. Such a new partnership—facilitated by new governance and incentives—must seek to stitch together all research instrumentation, to regulate access to it via contemporary identity management, and to protect it via accepted standards of security practice.

The fulcrum that balances where certain research resources reside may need to be moved if economic leverage is to be possible and if cyberinfrastructure investments and resources are to be sustainable. In many cases, computational and storage resources need to be housed centrally, if not managed centrally, to exploit advantages in energy management, network security, scale economics, and so forth. Along with the rationalization and integration of cyberinfrastructure resources comes the need for enhanced standards for data interchange and system interoperation, as well as semantic dictionaries that control vocabularies at least at the level of the academic discipline. These standards and tools will make it possible for the myriad parts of the research cyberinfrastructure to talk with one another.

While the goal of the research cyberinfrastructure is the promotion of innovation, the most powerful means to that goal is so-called collective intelligence. Collective intelligence not only assumes the creation of that richly interconnected fabric of research instruments, sensors, and so forth, but a real escalation in our ability to create cultures for innovation among the one billion planetary residents who now access the Internet. The potential for the research cyberinfrastructure to move humankind forward by embracing researchers from all corners of academe as well as from the ranks of amateurs, undergraduates, and others is without precedent. The future of scholarship may in fact be the sum of scholarship from A to Z raised by the power of IT times the sum of all possible researchers raised by the power of IT!

Endnotes

1. See the Synthetic Worlds Initiative at Indiana University at http://arden.indiana.edu/. See also, Castronova, E. "On the research value of large games: Natural experiments in Norrath and Camelot," 2005, at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=875571.
2. The Low Frequency Array (LOFAR) harnesses more than 10,000 simple radio antennas spread across the Netherlands and Germany and uses supercomputing to interpret data using high-speed calculations.