Information Technology and the Transformation of Scientific and Engineering Research[1]

Daniel E. Atkins
Professor of Information and EECS
The University of Michigan, Ann Arbor
atkins@umich.edu

Background

This paper is part of a series describing information technology (IT) impact on transforming various economic and social dimensions of our world. The focus of this piece is the impact of IT on transforming scientific and engineering research and allied education.

The history of modern information technology begins about 60 years ago with the invention of electronic computers motivated by the prospects of automating tedious scientific and engineering calculations. Computers began as massive devices, housed in centers, and with use highly intermediated by a new breed of experts. But the invention of the transistor, electro-optical devices, and the astonishing perfection of integrated circuit technology soon provided the physical basis for a new world of ubiquitous “digital enterprise” built upon software and knowledge. Science and engineering applications launched the information age, have continued to help define the leading edge, and now the science and engineering research communities are in the position to harvest the broader benefits of information technology evolution. Going beyond computation, information retrieval, or communication, these communities are starting to create integrated and comprehensive knowledge environments that can serve individuals, teams and organizations in ways that revolutionize what they can do, how they do it, and who participates. This trend also has profound broader implications for education, commerce, and social good.

For purposes of this paper we define the following waves of information technology: (1) pre-Internet, (2) Internet, (3) World-Wide Web (WWW), and now a fourth wave we will call *comprehensive virtual federation* (built upon *cyberinfrastructure*). Pre-Internet computers were standalone systems in super- through micro- sizes, and accessed locally or using remote alphanumeric terminals connected by telephone lines. The marriage of computing and communication through DARPA R&D projects together with the later deployment as the NSFNET, initially to connect advanced scientific computing resources supported by the NSF, spurred the adoption of the Internet for scientific, engineering and academic work in general. The primary use was email, remote login, and text file sharing. The invention of the World Wide Web (WWW) at CERN coupled with the first easy-to-

---

[1] Prepared for the First International Conference on the Economic and Social Implications of Information Technology, January 27-28, 2003. See http://cip.umd.edu/transform.htm

use browser, MOSAIC, at the University of Illinois provided a seminal distributed application for the Internet that has created an explosion in the extent and nature of use of the Internet and extended the media object in routine use to graphics, pictures, audio, video, computational modules, and multimedia composites of all of these plus text. The WWW has now laid the technical basis and sparked visionary possibilities for federating diverse distributed resources to produce new organizational forms to serve a wide variety of human endeavor. The Internet, the World Wide Web, extended and augmented in ways not yet fully defined, can be part of a "cyberinfrastructure" layer on which to build and use *comprehensive virtual federations* of *diverse resources* specialized for many types of knowledge intensive activities, including but not limited to scientific and engineering research and education.

The *resources* potentially available for federation include people, data, information, computational tools and services, and specialized instruments and facilities necessary for the functioning of a specific research team (or more generally, a specific community-of-practice.) The term *comprehensive* is included to imply that the extent and nature of the resources available through cyberinfrastructure could approach functional completeness for a specific community of practice. Astronomers, for example, can find all of the colleagues, data, literature, observational stations, and tools they need to work at the leading edge of their field. We will also add that in many scientific and engineering contexts we need to also add the term *advanced,* implying both high-performance technology and leading-edge research. So science and engineering research are moving into an IT era of *advanced comprehensive virtual federation.*

The general idea of virtual communities or organizations mediated through information technology is not new and has been, for example, the object of design and study by the field of computer-supported cooperative work (CSCW)[1] since the 1980s. More specific to virtual research communities for science and engineering, there are three primary lines of endeavor around IT-based systems for collaborative science and education under different names, but with similar goals: *collaboratory* or *co-laboratory*, *grid*, and *e-science*. These names are not necessarily mutually exclusive and the terms are sometime combined, as in "a grid-based collaboratory*,"* or "a collaboratory for e-science."

Collaboratory - The concept of a *co-laboratory* or *collaboratory*[2] – a laboratory without walls built upon distributed information technology – was defined at an invitational NSF workshop at Rockefeller University[2] in 1989 and later elaborated and sanctioned in an

---

[2] The term *collaboratory,* a play on the word *collaborate*, was coined by William Wulf while serving as the NSF Assistant Director for CISE at the time of the 1985 collaboratory workshop. The term has stuck and is now more often used than *co-laboratory*. It is, however, sometimes confused with the more limited notion of *collaboration technology* – applications such as Net Meeting, Centra, WebEx, or the Access Grid offering such features as synchronous conversation, video conferencing, shared viewing/pointing, and group editing. A collaboratory includes collaboration technology but is not just communication tools – it is a virtual organization of people, information, computational tools, and other facilities to support team-based knowledge discovery and dissemination – including but not limited to collaboration technology.

National Research Council report published in 1993 [3]. The report includes the following assertion:

> *The fusion of computers and electronic communications has the potential to dramatically enhance the output and productivity of US researchers. A major step toward realizing that potential can come from combining the interests of the scientific community at large with those of the computer science and engineering community to create integrated, tool-oriented computing and communication systems to support scientific collaboration. Such systems can be called collaboratories.*

This report called for a major initiative by NSF to explore the design and applications of the collaboratory. Although investment at the recommended level of investment did not occur, several collaboratory projects were funded by NSF, for example UARC-SPARC [4], and more were funded recently by the Department of Energy [5] and the National Institutes of Health [6]. An overview of collaboratory research, especially at the intersection of the technical and social dimensions can be found in [7], [8] and [9].

Recent the NSF has funded a Science of Collaboratories (SoC) project under the Information Technology Research (ITR) Program. The SoC Project is examining a series of collaboratory projects over the past decade that has been funded by NSF, DOE, NIH, and other agencies – some successful and some less so. The project *aims to define, abstract, and codify the broad underlying technical and social elements that lead to successful collaboratories*. The SoC website [10] includes an inventory of about 125 collaboratory projects. This site as well as the website for the Collaboratory for Research on Electronic Work (CREW) [11] includes numerous papers and other resources concerning the collaboratory R&D community.

Grid - Foster and Kesselman [12] have articulated the concept of the Grid[3]. The term *Grid* was adopted as an analogy to the dynamic linking of electric power generators over distribution grids to balance supply and demand over a mixture of geographic regions and power companies. A computer grid is intended to provide a user extraordinary computational power by aggregating high-performance computational resources over wide areas and diverse administrative entities. The NSF is now a funding a large Distributed Terascale (Tergrid) Project [13] to explore the creating a grid of supercomputers and virtual organizations between five high-performance computational centers. Many of the technical activities focus on network performance and creating standards and a middleware toolkit, Globus [14], to support efficient and secure interoperability between diverse machines and their hosting domains. The Globus project overview includes the following analogy:

---

[3] The terms *grid* and *access grid* are sometimes confused. The Access Grid (AG) is a suite of applications to support human interaction over a grid network. The Access Grid supports large-scale distributed meetings, collaborative work sessions, seminars, lectures, tutorials and training.

> *The development of the World Wide Web has revolutionized the way we think about information. We take for granted our ability to access information from all over the world via the Web. The goal of the Globus project is to bring about a similar revolution with respect to computation.*

Middleware is a central component of cyberinfrastructure to enable federation of resources across networks. It is the focus of a recent NSF Middleware Initiative (NMI) [15]. The NMI includes the following definition of middleware:

> *Middleware is software that connects two or more otherwise separate applications across the Internet or local area networks. More specifically, the term refers to an evolving layer of services that resides between the network and more traditional applications for managing security, access and information exchange to (1) let scientists, engineers, and educators transparently use and share distributed resources, such as computers, data, networks, and instruments, (2) develop effective collaboration and communications tools such as Grid technologies, desktop video, and other advanced services to expedite research and education, and (3) develop a working architecture and approach that can be extended to the larger set of Internet and network users.*

The NMI, among other things, is promoting synergy and commonality between major middleware grid-based activities such as Globus and the Internet2 Shibboleth Project [16] that is developing architectures, policy structures, practical technologies, and an open source implementation to support inter-institutional sharing of web resources subject to access controls in the higher education world. The emergent Web Services [17] and more ambitious Semantic Web [18] activities are also likely to be important parts of the middleware architecture of cyberinfrastructure.

Although much the focus of the Grid is transparent interoperability between high-performance computers, the use of the Grid encompasses a broader notion of resource sharing than only computational engines and in this context a *grid community* is equivalent to a collaboratory or co-laboratory. The term *datagrid* has also come on the scene [19]. Grids in this broader sense are about “resource sharing and coordinated problem solving in dynamic, multi-institutional virtual organizations.” The international community of academic and commercial participants coordinates Grid development and application through the Global Grid Forum[20] and the related GlobusWorld [21].

e-Science - In the U.K. research community [22] and the European Union Sixth Framework Program [23] the term *e-science* is used to describe science done through distributed global collaborations between people linked by the Internet with each other, with very large data collections, terascale computing resources, high-performance visualization, and instruments and facilities controlled and shared over the network. In describing e-science reference is made to both grid architecture and collaboratories. The UK is also calling for 1000x increase in the computational processing power and data transport bandwidth.

In a speech to the Royal Society in May 2002, Prime Minister Tony Blair included the following references to e-science:

> *What is particularly impressive is the way that scientists are now undaunted by important complex phenomena. Pulling together the massive power available from modern computers, the engineering capability to design and build enormously complex automated instruments to collect new data, with the weight of scientific understanding developed over the centuries, the frontiers of science have moved into a detailed understanding of complex phenomena ranging from the genome to our global climate. Predictive climate modeling covers the period to the end of this century and beyond, with our own Hadley Centre playing the leading role internationally.*
>
> *The emerging field of e-science should transform this kind of work. It's significant that the UK is the first country to develop a national e-science Grid, which intends to make access to computing power, scientific data repositories and experimental facilities as easy as the Web makes access to information.*
>
> *One of the pilot e-science projects is to develop a digital mammographic archive, together with an intelligent medical decision support system for breast cancer diagnosis and treatment. An individual hospital will not have supercomputing facilties, but through the Grid it could buy the time it needs. So the surgeon in the operating room will be able to pull up a high-resolution mammogram to identify exactly where the tumour can be found.*

General properties of virtual federations for science and engineering research - The general goal of all of these efforts is to use information technology to relax barriers of time and distance in bringing together the expertise, information, tools, instruments and facilities necessary for the discovery, dissemination and application of knowledge. Time can be synchronous, asynchronous, or “relevant” (just-in-time to not slow the workflow). Distance can mean geographic, organizational, or disciplinary. Interest in virtual federations is driven in part by a growing demand in research for more global, collaborative, and multidisciplinary approaches to discovery and problem solving. Mediated by information technology, the team work can in theory proceed not only in physical proximity but also easily flow through three other variations of same- and different- time and place. Observational instruments, unique because of cost and/or need to be in special locations can be shared. Arrays of smart sensors can provide unprecedented resolution or coverage in monitoring natural phenomena. Research community frameworks for multi-level, multi-system simulations can be collaboratively created and executed across grids of supercomputers. Archival data from many specialized fields can be made to interoperate in well-curated data repositories and used to drive comprehensive physical models in supercomputers or to extract new knowledge thorough data mining. Digital libraries can off anytime and anyplace access to the complete literature of a field and software agents to help maintain current awareness.

Some fields begin a virtual community motivated by collaboration around instruments and data gathering campaigns, some around collaboration in creating and using community data, and some around high-performance computational modeling, Experience has shown, however, that if these various starting points are successful that communities evolve the virtual environments to incorporate more capabilities. They move in the direction of functional completeness --- the (virtual) place to be for the project or field.

As mentioned earlier, the emerging over arching use of information technology in scientific and engineering research is no longer only about computation or information retrieval. It is rather about the potential for new comprehensive infrastructure, environments, and organizations to enable the overall enterprise of discovery, dissemination, and use (including education, public awareness, and informed policy making). The architecture, design principles, and analysis of impact of such IT-based environments must therefore include a merged technical and social/cultural/behavioral perspective. The past is littered in both industry and science with virtual organizations projects that have failed due to lack of knowledge or consideration of the social dimensions of team activities.

What's happening now

Collaboratory, grid community, and e-science environments are all names for virtual federations of resources based upon information technology to create knowledge environments (or organizations) for science and engineering research. There is similar work also underway using none of these terms. Recently the leadership of NSF observed a growing number of grass-roots projects, many funded under their Information Technology Research (ITR) program recommended by Presidential Information Technology Advisory Committee, that included major activities and expenditures to create and use collaboratories, grids, or e-science environments in specific, often multidisciplinary, projects. Many also need access to the next frontier of high-performance computing and are concerned by the danger of the US not maintaining leadership in this area.

Examples of these projects include:

1. **The National Virtual Observatory (NVO)** [24], a collaboration to create standards for astronomical data collections that will be used by the wide astronomical community. It will make data easier to use, easier to find, and easier to join with other data to create a *digital sky*. A second thrust is exploring the use of high-performance computing resources for discovery in astronomy.

2. **Network for Earthquake Engineering Simulation (NEES)** [25], to provide an unprecedented infrastructure for research and education, consisting of networked and geographically distributed resources for experimentation, computation, model-based simulation, data management, and communication. Rather than placing all of these resources at a single location, NSF has leveraged its investment and facilitated research and education integration by distributing the

shared-use equipment among nearly 20 universities throughout the US. To insure that the nation's researchers can effectively use this equipment, equipment sites will be operated as shared-use facilities, and NEES will be implemented as a network-enabled collaboratory. As such, members of the earthquake engineering community will be able to interact with one another, access unique, next generation instruments and equipment, share data and computational resources, and retrieve information from digital libraries without regard to geographical location.

3. **The Grid Physics Network (GriPhyN) [26]** project brings together a team of experimental physicists and information technology researchers to develop Grid technologies for scientific and engineering projects that must collect and analyze distributed, petabyte-scale datasets. GriPhyN research will enable the development of Petascale Virtual Data Grids (PVDGs) through its Virtual Data Toolkit.

4. **National Ecological Observatory Network (NEON) [27]** will be a network of networks, a system of environmental research facilities and state of the art instrumentation for studying the environment. Each node in NEON will be a regional observatory, comprised of a core site and associated sites that are linked via cyberinfrastructure. These observatories will be geographically distributed based on the US Forest Service defined eco-regions of the US. NEON will enable integrative research on the nature and pace of biological change at local, regional and continental scales. Its advanced technologies and continental scale connectivity will be used to measure all factors that affect the structure and function of ecosystems, for example the power of genomics will be applied to predicting how the spread of invasive species will affect native biodiversity.

5. **The ATLAS Experiment for the Large Hadron Collider (LHC**) [28] is creating a collaboratory to support a world wide community of physicists engaged in building and using the LHC centered at the CERN Laboratory in Switzerland. Its goal is to explore the fundamental nature of matter and the basic force that shape the universe. ATLAS is the largest collaborative effort ever attempted in the physical sciences. There are 2000 physicists participating from more than 150 universities and laboratories in 34 countries. The ATLAS Experiment Collaboratory is also extraordinary in terms of data handling requirements. Estimates are that by about 2012 it will need an exabyte (10^{18} bytes) archive for data from four LHC experiments. These data need to be readily available, not just in a central location, but to scientists all over the world. In addition computation rates of 0.30 petaFLOPS (0.3×10^{15} floating point arithmetic operations per second) will be required to process the data.

6. **Environmental Research and Education (ERE)** is a multidisciplinary field of increasing importance and is a high priority in the NSF research planning process. An NSF Environmental Research and Education Advisory Committee has just released a ten-year outlook for the NSF's ERE programs in a report entitled *Complex Environmental Systems: Synthesis for Earth, Life, and Society in the 21st Century.* [29] The report contains the following comments on building

> capacity: *Long-term dynamic partnerships that cross national and regional jurisdictions and international boundaries are needed to address multi-scale challenges. Developing the requisite cyberinfrastructure – advanced data assimilation and curation, networking, modeling, and simulation tools for large-scale, systems level, integrated applications – is key to making progress in the decade ahead.*

These and other research communities assert that new high-capacity IT-based virtual environments are absolutely critical to their future research aspirations.

NSF has coined the term "cyberinfrastructure" for an information technology layer that could be a platform for such discipline- or project- specific knowledge environments. The NSF Assistant Director for CISE, Ruzena Bajcsy formed a "Blue Ribbon Advisory Panel on Cyberinfrastructure[4]" to provide NSF advice on the nature of this movement and the opportunities and challenges it presented to NSF, a vision of how NSF might respond, and advice on how the current major investments by NSF in cyberinfrastructure, largely in the form of supercomputing alliances, should fit into this vision. The charge includes the notion that creating and using advanced cyberinfrastructure requires synergy between the computer science and engineering research community and the greater domain science communities who want to benefit from it. Cyberinfrastrtucture and its use is both an object of research as well as an enabler of research. The Panel conducted surveys, hearings, reviewed prior relevant recommendations, requested comments on a draft report, and has recently submitted a final report to NSF that will soon be made public [30].

Extrapolating in large part from prior NSF investments in cyberinfrastructure including high-performance computing, networking, middleware, and digital libraries, trends in the IT industry, and the vision and innovation coming from many research communities, the Panel asserts that *the capacity of information technology has crossed thresholds that now make possible a comprehensive "cyberinfrastructure" on which to build new types of scientific and engineering knowledge environments and organizations and to pursue research in new ways and with increased efficacy*. The Panel shares the belief of many who testified that advanced cyberinfrastructure could be the basis for revolutionizing the conduct of scientific and engineering research and affiliated education. It could also have broad impact in many other domains of knowledge intensive activity.

The Panel also found that *such environments and organizations, enabled by cyberinfrastructure, are increasingly required to address national and global priorities such as global climate change, protecting our natural environment, applying genomics-proteomics to human health, maintaining national security, mastering the world of nanotechnology, and predicting and protecting against natural and human disasters; as*

---

[4] The author of this paper served as Chair of this Panel. Dr. Peter Freeman, moderator of the session for which this paper was prepared, has superceded Dr. Bajcsy as the NSF AD for CISE.

well as to address some of our most fundamental intellectual questions such as the formation of the universe and the fundamental character of matter.

Although this Panel was commissioned by the NSF and concentrated on the NSF research community it also found that similar visions and needs are emerging in research communities supported by other federal agencies most apparently by the Department of Energy (for example [5]) and the National Institutes of Health (for example [31]). And as mentioned earlier, heavily funded e-science initiatives based on cyberinfrastructure are underway in Europe and the Japanese recently set new records in one aspect of cyberinfrastructure applied to research in the Japanese Earth Simulator Project [32]. They have developed and are running a comprehensive Earth systems simulator program at unprecedented levels of resolution and scale and at record setting computation rates.

Figure 1 below, taken from the Panels's report is a schematic of what cyberinfrastructure includes and how it fits within a stack of activities. Starting at the bottom, cyberinfrastructure is built on electro-optical integrated circuit technologies for computation, storage, and communication (data transfer). Cyberinfrastructure begins with networking, operating systems, and middleware providing the generic capabilities for management, transport, and federation of systems and services (tools) described in the five columns. A community-specific, customized knowledge environment can ideally be created efficiently and effectively using facilities, tools and toolkits provide at the cyberinfrastructure layer to federate the requisite resources.

This model assume significant effort to capture and benefit from commonalities across science and engineering disciplines and appropriate levels of coordination and sharing of facilities and expertise to minimize duplication of effort, inefficiency, and excess cost. To achieve advanced capability it also assumes real collaboration between domain scientists and engineers and computer scientists and engineers, and constructive participation by social scientists to help understand the social and cultural issues.

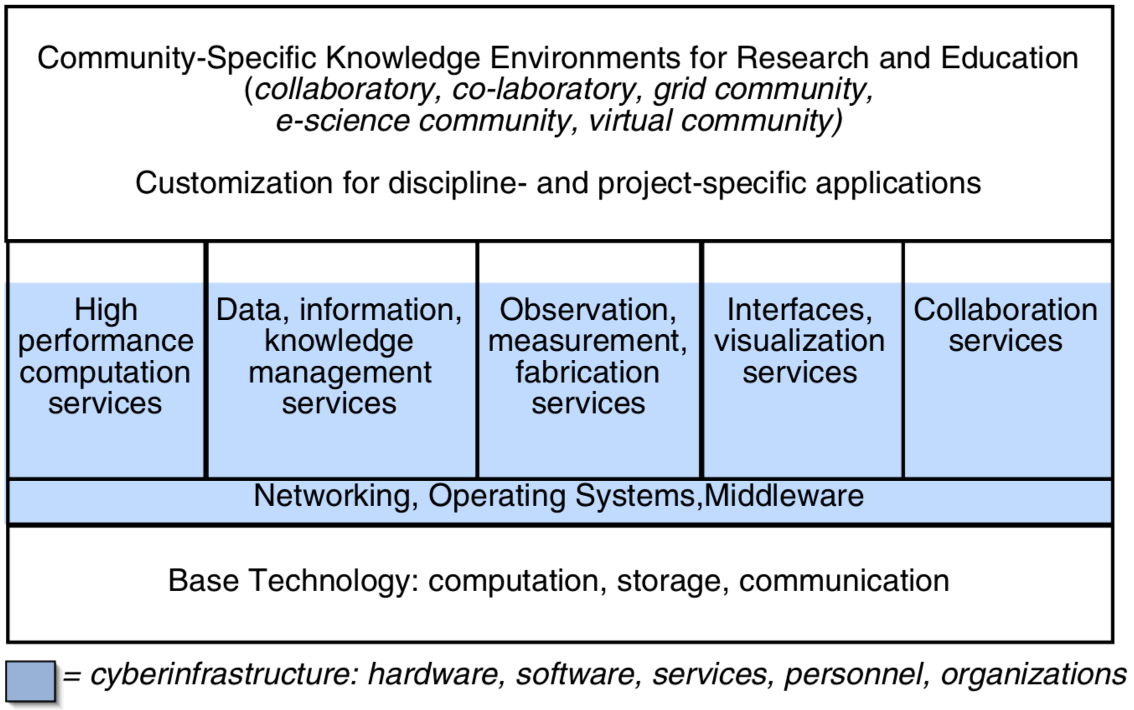

Figure 1 - Integrated cyberinfrastructure services to enable new knowledge environments for research and education.

The Panel goes on to recommend that the NSF seek significant new funding and assume the leadership of a Foundation-wide Advanced Cyberinfrastructure Program (ACP) with close coordination with other U.S. and international R&D agencies. The report emphasizes that investment in advanced cyberinfrastructure involves significant and long term investment is hardware and software, but above all it is an investment in people: researchers to create and use it, professionals to build and operate it, and a broad spectrum of people to educate themselves and others about new methods and opportunities for doing research. It also emphasizes that a significant part of the challenge is to find ways to identify and implement infrastructure that can be easily used by a wide range of the science domains and to engage a broad set of contributors including industry and universities. Details on the nature of cyberinfrastructure, potential impact, budget level, and organizational approach are included in the Panel's report [30].

A central goal of ACP is to define and build cyberinfrastructure that facilitates the development of new applications, allows applications to interoperate across institutions and disciplines, insures that data and software acquired at great expense are preserved and easily available, and empowers enhanced collaboration over distance, time and disciplines. The individual disciplines must take the lead in defining specialized software and hardware environments for their fields based on common cyberinfrastructure, but in a way that encourages them to give back results for the general good of the research enterprise. Achieving this vision will challenge our fundamental understanding of computer and information science and engineering as well as parts of social science, and it will motivate and drive basic research in these areas.

Figure 2 is also taken from the Panel's report and is an attempt to convey a general idea of the focus and impact of an advanced cyberinfrastructure program (ACP). One dimension is technology *capacity* by which we mean a measure of the aggregate performance of the computational, storage, and data transfer technology. Capacity is measured in units such as floating-point operations per second (FLOPS), bytes, and bytes per second. It is a measure of the processing power, data volumes, and date rates that can be handled in the IT-based research environment. The other dimension is *functional comprehensiveness* (or extent of coverage or completeness) of the assets needed by a research team or organization to do their work. This is a qualitative measure indicating, for example, what percentage of colleagues, relevant scientific literature, archival data, instruments and other critical facilities are easily available through the cyberinfrastructure based environment, i.e. through the virtual organization – the collaboratory – the grid community.

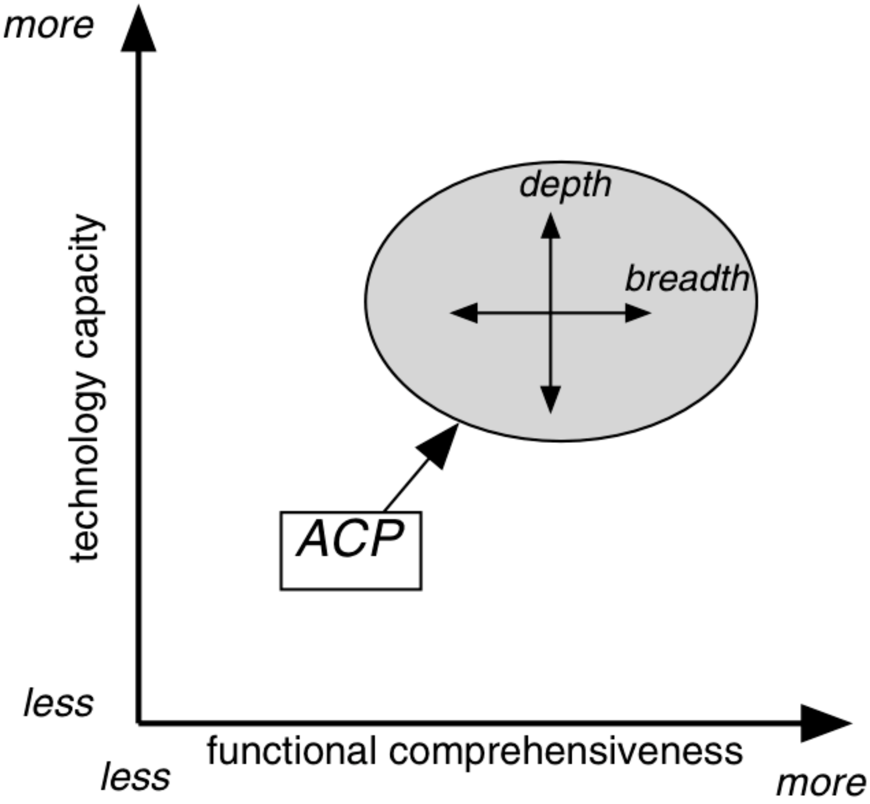

Figure 2 - Increasing technology capacity and functional comprehensiveness of cyberinfrastructure enabling depth and breadth approaches to discovery.

NSF and others have made significant successful investments in research to move higher along the technology capacity dimension and to make such capacity available through, super-computing initiatives. They have also invested in creating understand and prototypes of important functional components of a research knowledge environment:

digital libraries, visualization tools, collaboration technology as well as the middleware glue to link them together. The opportunity now presented to the research community is to establish new working environments that are advanced, innovative, and innovating in terms of both capacity and comprehensiveness. The shaded region represents such a region and the general goal of an ACP is to move both *within* and *among* broad fields of science into these regions and to move the regions up and to the right. To borrow a phase from Dan Reed, Director of the PACI organization at the University of Illinois, the challenge and opportunity for an advanced cyberinfrastructure program is to "build up" in capacity and "build out" in function and use,

As the combined region capacity and functional comprehensiveness increases, and is adopted more broadly, the payoff will likely derive from enhancing both "depth" and "breadth" approaches to discovery.

In a depth opportunity, for example, atmospheric scientists could use higher-performance computation (together perhaps with denser and smarter distributed networks of sensors and with higher quality archival data) to improve the resolution and accuracy of a weather prediction model. Astronomers could use a more capable telescope to look more deeply into their favorite region of the universe.

In a breadth opportunity, a multidisciplinary team of earth scientists could use the availability of more computational power, more complete multi-dimensional data, enhanced observation capability, and more effective remote collaboration services to bring together an entire earth system simulation framework capable of supporting usefully predictive environmental simulations. Astronomers, given access to a federated "digital sky," could explore the breath of the known universe over the entire available electromagnetic spectrum to seek, for example, rare or new objects or phenomena. We can only begin to glimpse the impact of blended depth and breath approaches, especially as they weave together complementary expertise from multiple disciplines.

Although the breadth and extent of grass roots activity underway to innovate research through the application of cyberinfrastructure is impressive and provides evidence of a pull from the research communities, it also suggests the potential for lost opportunity costs. Although the Panel's report stresses the vast opportunity for creating new research environments based upon cyberinfrastructure, it also notes significant risks and costs if the research community does not act quickly and at a sufficient level of investment. The dangers, all increasing with the passage of time, include adoption of incompatible data formats in different fields; permanent loss of observational data due to lack of well-curated, long-term archives; increased technological ("not invented here") balkanizations not interoperability among disciplines; wasteful redundant system-building activities among science fields or between science fields and industry; lack of synergy among information technology research, the IT industry, and domain science users resulting in under- or overestimating technological futures; lost opportunity from not driving basic computer science research with advanced applications; loss of leadership to other countries and a falloff of research and economic vigor; lack of understanding of social/cultural barriers to new ways of doing research; inadequate supporting or supported educational activities; and an inadequate, piecemeal cyberinfrastructure program.

It is too early to tell where these recommendations will lead but there is no doubt that various research communities at different rates and from different starting points are moving into a new age of adoption of information technology in both capacity and extent of application.

Broader Impact and the Future

The authors of these "transformation" papers were asked to also comment on the significance of transformation in their designated area on other aspects of the digital enterprise. With respect to our designated area we were also asked to comment on (1) drivers and barriers to further change, and (2) on what further we would like to know and what research is needed to find out. We will conclude this paper with some partial responses to these topics.

Impact on other digital enterprise - Our treatment of the topic of IT-impact on science and engineering has focused most directly on the conduct of scientific and engineering research with only indirect treatment of education and practice. We have noted the growing needs of many research fields to work in more interdisciplinary global teams using larger, richer, and more diverse sources of expertise, information, observation, and computational tools. These needs coincide with a state of information technology that has now crossed thresholds of capacity and function such that we have the opportunity to meet these needs through "cyberinfrastructure" that enables virtual federation of resources to create new forms of knowledge work environments. Obviously success at this in the science and engineering research enterprise would have enormous impact on other knowledge intensive activities including commerce and education. We will say a bit more about education.

Studies of collaboratory projects to date has shown that they have been used by researchers to help train their colleagues at other locations to use other instruments, supercomputers or other techniques for discovery; they have enabled graduate students to participate earlier in their careers in authentic experiments involving scarce or remote instruments; graduate students have benefited from mentoring through the collaboratory by leaders of their field at other institutions; and faculty at undergraduate institutions have used collaboratories to participate in first-tier research and to embellish their undergraduate teaching.

The NSF sponsored Space Physics and Aeronomy Research Collaboratory (SPARC) [33]centered at the University of Michigan combined forces with the NASA sponsored Windows to the Universe Project [34] under the direction of Dr. Roberta Johnson (now at NCAR) to explore the concept of a dual-use collaboratory serving both a frontier research community *and* middle school earth science teachers and students. Windows itself is a type of collaboratory that provides not only instructional material but also facilities for science teachers to build a virtual community to help each other with using the material to enrich their teachers. These teachers and students also could interact with scientists, data, and instruments in SPARC thorough a special portal designed for the appropriate level of understanding and participation by the middle school teams. This dual-use experiment was not the subject of a careful longitudinal study, but anecdotal evidence suggest that collaboratories could be designed to enrich science education by making it more authentic, more fun, and more relevant. Another interesting example of a dual use collaboratory is the University of Illinois Bugscope Project [35].

Another great hope for cyberinfrastructure-based knowledge environments is that they will democratize participation for those that have been disadvantaged because of history, location, or physical reasons. The original NRC Collaboratory report and the new NSF Panel report treats this topic extensively.

The final comment on broader impact concerns the future of higher education, particularly the research university. The National Academies of Sciences in recent years has initiated a series of workshops and studies around the topic of information technology and the future of higher education [refs]. Cutting across all these studies is the notion that information technology potentially offers a vast array of new place and time independent opportunities for augmenting how the fundamental mission of teaching/learning, scholarship/research, and service/societal engagement is carried out. Restricted notions of “distance learning” in higher education are now broadening to that of using IT to facilitate the work of communities involved in knowledge creation, dissemination, and use. The same cyberinfrastructure and opportunities and challenges it offers research is essential the same for higher education at large…. the collaboratory maybe the tooling for university of the future. More elaboration on this topic can be found in [DAV book].

What are the drivers and barriers to further change?

Hopefully a driver, or at least an agent for empowerment for future change will be a major Advanced Cyberinfrastructure Program emerging under the leadership of the NSF. The Cyberinfrastructure Panel’s report includes extensive treatment of drivers and barriers in this context. The barriers are probably more economic and organizational than technical. There is little doubt that grassroots activity from research communities will continue trying to innovate what and how they do research using cyberinfrastructure. The open question is where we will martial the will and the means quickly enough to seize the full potential of what many believe is a nascent revolution in science and engineering research and education.

What would we like to know and what research is needed to find out?

****More to be added here*****

References

1. Grudin, J., *CSCW: History and Focus*. IEEE Computer, 1994. **27**(5): p. 19-27.
2. Wulf, W.A., *The National Collaboratory - A White Paper, Appendix A in Towards a National Collaboratory, the unpublished report of an invitational workshop held at Rockefeller University, march 17-18, 1989 (Joshua Lederberg and Keith Uncapher, co-chairs.)*. 1989.
3. National Research Council (U.S.). Committee on a National Collaboratory : Establishing the User-Developer Partnership., *National collaboratories : applying information technology for scientific research*. 1993, Washington, D.C.: National Academy Press. xi, 105.
4. *Space Physics and Aeronomy Research Collaboratory Website,* *http://doecollaboratory.pnl.gov/history.html*.
5. DOE, *U.S. Department of Energy National Collaboratories,* *http://doecollaboratory.pnl.gov/history.html*. 2002.
6. *NIH Collaboratory Project Website, http://collaboratory.psc.edu/*.
7. Finholt, T.A., Olson, G. M., *From laboratories to collaboratories: A new organizational form for scientific collaboration*. Psychological Science, 1997. **8**(1): p. 28-35.
8. Finholt, T.A., *Collaboratories*, in *Annual Review of Information Science and Technology*, B. Cronin, Editor. 2003.
9. Olson, G.M., Finholt, T.A., Teasley, S.D., *Behavioral aspects of collaboratories*, in *Electronic collaboration in science*, S.H.K.M.F. Huerta, Editor. 2000, Lawrence Erlbaum Associates: Mahwah, NJ. p. 1-14.
10. *Science of Collaboratories Project Website,* *http://www.scienceofcollaboratories.org/*.
11. *Collaboratory for Research on Electronic Work (CREW) Website.* *http://www.crew.umich.edu*.
12. Foster, I. and C. Kesselman, *The grid : blueprint for a new computing infrastructure*. 1999, San Francisco: Morgan Kaufmann Publishers. xxiv, 677.
13. *Teragrid Project website, http://www.teragrid.org/*.
14. *The Globus Project website, http://www.globus.org/about/*.
15. *NSF Middleware Initiative website, http://www.nsf-middleware.org/*.
16. *Shibboleth Project website, http://shibboleth.internet2.edu/*.
17. W3C, *Web Services Activity website, http://www.w3.org/2002/ws/*.
18. W3C, *Sematic Web Activity website, http://www.w3.org/2001/sw/*.
19. *The DataGrid Project website, http://eu-datagrid.web.cern.ch/eu-datagrid/*.
20. *Global Grid Forum Website, http://www.gridforum.org/*.
21. *Globus World website, http://www.globusworld.org/*.
22. *The UK Research Councils e-Science Core Program website,* *http://www.research-councils.ac.uk/escience/*.
23. *EU Sixth Framework Program (2002-2006) website,* *http://europa.eu.int/comm/research/fp6/index_en.html*.
24. *National Virtual Observatory website, http://www.us-vo.org/*.
25. *Network for Earthquake Engineering Simulation website, http://www.nees.org/*.
26. *Grid Physics Network website, http://www.griphyn.org/index.php*.

27. *National Ecological Observatory Network (NEON) website, http://www.nsf.gov/bio/neon/start.htm*.
28. *The ATLAS Experiment for the Large Hadron Collider website, http://atlasexperiment.org/*.
29. NSF, *Complex Environmental Systems: Synthesis for Earth, Life, and Society in the 21st Century, available at http://www.nsf.gov/geo/ere/ereweb/advisory.cfm*. 2003, NSF.
30. NSF, *Final Report of the NSF Blue Ribbon Advisory Panel on Cyberinfrastructure: Revolutionizing Science and Engineering Through Cyberinfrastructure*. January 2003 (in press).
31. NIH, *Biomedical Informatics Research Network (BIRN) website, http://www.nbirn.net/*.
32. *Earth Simulation Center website, http://www.es.jamstec.go.jp/esc/eng/*. 2002.
33. *Space Physics and Aeronomy Research Collaboratory Website, http://intel.si.umich.edu/sparc/*.
34. UCAR, *Windows to the Universe website, http://www.windows.ucar.edu/*.
35. *Bugscope Project website, http://bugscope.beckman.uiuc.edu/*.