Infrastructure for eScience and eLearning in Higher Education

Ed Lazowska	Peter Lee	Chip Elliott	Larry Smarr
University of Washington	Carnegie Mellon University	BBN Technologies	UC San Diego

Version 11: December 22, 2008[1]

Recent rapid advances in information and communication technologies – both hardware and software – are creating a new revolution in discovery and learning, laying the foundation for a more competitive US economy in the second decade of the 21st century.

Over the past several decades, computational science – the large-scale simulation of phenomena – has joined theory and experiment as a fundamental tool in many branches of science and engineering. Today we are at the dawn of a second revolution in discovery – a revolution that will have far more pervasive impact. The focus of this new approach to science – called *eScience* – is *data*; specifically:

- the ability to manage orders of magnitude more data than ever before possible;
- the ability to provide this data directly and immediately to a global community;
- the ability to use algorithmic approaches to extract meaning from huge volumes of data.

Enormous numbers of tiny but powerful sensors are being deployed to gather data – deployed on the sea floor, in the forest canopy, in gene sequencers, in buildings and bridges, in living organisms (including ourselves!), in telescopes, in point-of-sale terminals, in social networks, in the World Wide Web. These sensors (and, indeed, simulations too) produce huge volumes of data that must be captured, transported, stored, organized, accessed, mined, visualized, and interpreted in order to extract knowledge. This "computational knowledge extraction" lies at the heart of 21st century discovery.

The fundamental tools of eScience include sensors and sensor networks, databases, data mining, machine learning, data visualization, and cluster computing at enormous scale as pioneered for other purposes by companies such as Google and Amazon.com[2]. These companies have created entirely new businesses by capturing enormous volumes of data, mining it for new knowledge, and making it freely available on the World Wide Web in useful ways, transforming how people find and make use of information on a daily basis. The same technologies are helping to usher in the era of eScience. eScience, even more than computational science, illustrates the extent to which advances in all fields of science and engineering are married to advances in computer science and the mathematical sciences.

The fact that eScience is a *networked* science has an important democratizing effect – data can be made available to everyone, not only to professional scientists but also at the same time to students and teachers. Beyond new scientific discoveries, we are at the dawn of a new revolution in learning due to information and communication technologies. The broadband and un-tethered

[1] For the most current version of this essay, as well as related essays, visit http://www.cra.org/ccc/initiatives
[2] See a related essay on "Big-Data Computing": http://www.cra.org/ccc/docs/init/Big_Data.pdf

world – information available ubiquitously via high-bandwidth wired and wireless networks and a wide variety of information devices – makes learning an anywhere, anytime activity. We must allow students to live in this future, harnessing the creativity and ingenuity of our young people. We have the opportunity today to reinvigorate the wildly successful government-industry-university partnership that gave birth to and nurtured the growth and evolution of the current Internet, stimulating future advances in broadband networking and its wide-ranging applications, by leveraging and adding value to the Administration's plans to make America a world leader in broadband access[3].

At the foundation of our recommendations is a series of investments to create *balanced* high performance cyberinfrastructure for hundreds of U.S. colleges and universities which will stimulate the development, deployment, and application of a new generation of data-intensive discovery. *Research universities are the central engine of the innovation economy.* But this role depends critically on having state-of-the-art cyberinfrastructure as a foundation for eScience research and education activities. Like the physical infrastructure of roads, bridges, power grids, and water systems that support modern society, "cyberinfrastructure" refers to the distributed computer, information and communication technologies combined with the personnel and integrating components that provide a long-term platform to empower the modern scientific endeavor.

In making these one-time stimulus expenditures, we recognize that care must be taken to make investments that lead to substantial short *and* long-term gains. Since network infrastructure investments promise useful lifetimes of 8-10 years and have the potential to impact millions of researchers and students, our priorities place network infrastructure investments over computing infrastructure investments. Knowledgeable people are an appreciating asset. *Providing network-enabled opportunities for more students and faculty to work with large-scale data-intensive computing and other cyberinfrastructure will yield high returns over many years.*

We assume that a set of overarching broadband stimulus programs will provide all colleges and universities with greatly enhanced bandwidth in the national research networks (Internet2, NLR) and regional networks, but not in campus connectivity or other renovations in campus infrastructure. Hence, the recommendations below address needs and opportunities in building additional network *and* computing capability onto this foundation that enable new advances to take our nation into a new era of data-intensive scientific discovery:

1. **Provide $630 million for next-generation network infrastructure to 400 colleges and universities**

    a. $120 million would be used to capitalize 10 gb/s connections from existing GigaPoPs and RONs to 100-150 research-oriented colleges and universities (Carnegie Doctoral/STEM and Doctoral/Professional institutions). Networking is a geographic activity; this program would be structured similarly to the extremely well conceived NSF Connections program, with proposals expected from 10-20 GigaPoPs and RONs representing geographical clusters of institutions.

[3] See a related essay on "Innovation in Networking": http://www.cra.org/ccc/docs/init/Networking.pdf

b. $70 million would be used to capitalize 1 gb/s connections to an additional 250-300 colleges and universities, again through a program structured around existing GigaPoPs and RONs.

c. $300 million would be used to provide a 50% subsidy for high-speed campus wired and wireless network upgrades for the 400 institutions in (a) and (b). This would allow students and faculty across disciplines to "live in the future" with ubiquitous high-speed connectivity, access to cloud computing facilities, widespread adoption of high-definition teleconferencing to support multi-institutional teams, etc. Such upgrades should be "research enabled" in the dual senses of enabling research on innovative networking technologies and applications, and enabling users to be "early-adopters" of these innovations. Backbone, regional, and campus networks can be "research-enabled" by introducing equipment that supports greater programmability deep into the network, thus opening up the infrastructure for innovations in network security, reliability, robustness and performance. (The marginal cost of this research-enablement is small.) This research-enabled infrastructure will unleash a new wave of innovations. See (e) below for additional suggestions on the structuring of this investment.

d. $100 million would be used to create advanced wireless and sensor networks at 100-150 research-oriented colleges and universities (Carnegie Doctoral/STEM and Doctoral/Professional institutions). This would allow large-scale experimentation with next-generation wireless network technology (such as cognitive radios), hands-on experiences for students, and a means by which students on those campuses may participate 24x7 via wireless in new national research experiments.

e. $40 million would be used to greatly expand the national cyberinfrastructure backbone by procuring 10 nationwide 10 gb/s wavelengths, including seven years prepaid maintenance and operations. These can be relatively inexpensively procured because of prior investments by National LambdaRail and Internet2. These are the "interstate highways," with high speed parallel lanes enabling data traffic to be mixed (shared Internet) on some lanes and streamed at much higher speeds over other lanes ("HOV lanes"). It is critical that the campus infrastructure in (c) be architected so that it can provide direct connectivity to the backbone at 10 gb/s to specific laboratories and scientists; technically, providing "VLANs for science" – essentially, a private network infrastructure for science, computer science, and networking research rather than the more common Layer 3 networks which are used for general institutional traffic. See also (c) in the section below.

2. Provide $200 million to research universities for an "MRI Plus" program in advanced eScience cyberinfrastructure

a. $100 million would be used to build a networked collection of 10 geographically distributed mesoscale DISC (Data-Intensive Scalable Computing) systems, in the spirit of the OpenCirrus initiative[4], but with greater coordination, oversight, and network bandwidth between them. Some of these machines would be made available to

[4] http://research.yahoo.com/node/2328

researchers pursuing applications of big-data computing, while others would be made available to systems researchers exploring operations and low-level systems issues.

b. $50 million would be used to provide up to 12 awards for medium-scale supercomputers and advanced instruments.

c. $50 million would be used to provide 200 $250,000 supplements to existing NSF peer reviewed grants to fund the enhanced connectivity to laboratories referred to in (e) in the section above, as well as local data environment enhancements such as scalable storage, compute, and visualization capabilities.

3. Provide $225M to upgrade existing HPC capabilities

a. $80 million would upgrade the HPC and networking capabilities at the four major HPC centers (NCSA (IL), TACC (TX), UTK (TN), PSC (PA)), and increase bandwidth to the six nearest institutions for each center (including providing the capability for dedicated 10 gb/s optical paths).

b. $120 million would provide infrastructure upgrades for Track-1 and Track-2 awardees in 27 EPSCoR jurisdictions, as well as minority-serving institutions.

c. $25 million would be used to upgrade five Teragrid resource providers (LONI (LA), Indiana University (IN), Purdue (IN), ORNL (TN), University of Chicago (IL), SDSC (CA)), and provide connectivity to three additional sites.

4. Provide $140 million per year ($700 million 5-year total) to allow NSF to accelerate its research initiatives in networking technologies and applications and in data-intensive computing and eScience

a. $70 million per year would support Broadband Innovation Incubator grants for broadband and broadband-enabled research, education, and innovation[5].

b. $70 million per year would support a similar initiative in data-intensive scalable computing and eScience.

These initiatives can be launched immediately. The higher education community, its partners (e.g., NLR and Internet2), and the National Science Foundation (specifically CISE and OCI) are prepared.

Stimulus Impact – Immediate and beyond

Government/university/industry partnerships in technology infrastructure have proven, time and again, that putting leading edge networking, computing, and storage into the hands of researchers and students can transform our society and the entire world's economy.

[5] See the essay on "Innovation in Networking": http://www.cra.org/ccc/docs/init/Networking.pdf

The flowering of creativity and innovation sparked by the networking and high performance computing programs of the 1970's and 1980's, for example, led to surges of interest in research, new inventions, and startup companies that have completely changed our economy and further fueled our nation's leadership in science and technology. Programs such as the one proposed here have produced thousands of computing and engineering graduates with the technical skills, motivation, and vision to create new industries and transform old ones. Just as important is the fact that all graduates become citizens who are comfortable with – and, in fact, demand – technology infrastructure such as broadband networks in their work and home lives. Today, these educated consumers help move our economy forward and make their businesses more efficient and innovative.

Although such longer-term transformations will take 2-5 years to gain significant economic force, *immediate* stimulus to the nation's economy will be significant in the 3-18 month timeframe, and will help stabilize the country's rapid decline in high-tech employment. Challenger Gray & Christmas has reported that 140,422 jobs were lost in the telecom, electronics and computer industries in the first three quarters of 2008[6]. The initiatives outlined here can help win back thousands of those jobs and, equally importantly, help preserve tens of thousands of jobs that otherwise might be lost if struggling companies were to collapse in the coming months. Specifically, in the networking area:

- Capitalization of 10 gb/s connections to 100-150 research-oriented colleges and universities (1a, $120 million) and the expansion of the national cyberinfrastructure backbone through the procurement of 10 nationwide 10 gb/s wavelengths (1e, $40 million) could be accomplished within the first 12 months, with the bulk of the dollars going to telecommunications and cable suppliers, construction/installation companies, and network equipment manufacturers, many of which have started deeper layoffs in recent months in the face of rapidly shrinking revenues.

- Typical companies that would benefit most directly would include Ciena, Cisco, Force10, Hewlett Packard, Infinera, Juniper, AT&T, Level3, Qwest, SBC, and Verizon. The impact could be significant for leading-edge companies. Ciena and Infinera, for example, have annual revenues of $900 million and $245 million, respectively. These networking infrastructure initiatives could provide a significant offset to revenue decline, thereby avoiding further layoffs and even making possible some expansions.

Laboratory, HPC, and Teragrid upgrades also would show immediate impact:

- Supplements to existing peer-reviewed grants for enhanced laboratory connectivity (2c, $50 million) can immediately exploit national backbone and regional network upgrades, building out new enhancements from the edge of the campus directly into researchers' laboratories. Similarly, HPC site upgrades (3a, $80 million) and Teragrid site upgrades (3c, $25 million) can be implemented within the first 3 months, with immediate impact on productivity.

[6] *Barron's*. "Tech Job Losses Headed For Worst Year Since '03; Challenger Sees CSCO, NOK, QCOM Upping The Total." Eric Savitz, November 14, 2008, blog.

- In addition to the networking companies mentioned above, computer and storage vendors such as IBM, Hewlett-Packard, Dell, Sun Microsystems, nVidia, Intel, AMD, Cray, SGI, Network Appliance, and EMC would receive stimulus benefit from this portion of the program. For struggling companies such as Sun and SGI, this could have a substantial stabilizing effect on their workforce size (currently totaling about 14,300 employees).

To expedite the impact, NSF should complete its award processes within 3 months. Program directors should evaluate submitted proposals upon receipt. The processing of solicitations and awards should be streamlined, with campuses "shovel ready" to execute their upgrades upon receipt of the funds.

Almost immediately, thousands of new jobs, many of them blue-collar, would be created, as a major infrastructure installation process would begin within the first 3 months and ramp up over the first 12. In addition to on-campus installations and GigaPoP/RON upgrades, for campuses in rural areas, telecommunications companies would need to hire additional workers to install new trenched and aerial fiber, etc., as well as ramp up their technical support operations. *Within the first 12 months, upwards of 10,000 jobs would be created.*

Looking to the second 12 months, an even greater impact would be felt:

- For the 250-300 additional colleges that would receive 1 gb/s connection upgrades (1b, $70 million), the resultant economic stimulus would be similar to, but much larger than, that associated with the 10 gb/s upgrades listed previously. (We assume a slower start because this set of colleges and universities will be less prepared to carry out the upgrades and will need several months of planning to ramp up for this.)

- The 50% subsidy for campus wired and wireless upgrades (1c, $300 million) can start immediately with university fiscal years. While such upgrades normally require a period of careful planning, many universities would be ready to start within the first 6 months, as they have upgrade plans in place. We project that roughly 25% of these upgrades would commence within the first 6 months, and the rest progressively over the next 18 months.

- The mesoscale DISC systems (2a, $100 million), acquisition of medium-scale supercomputers (2b, $50 million), and EPSCoR upgrades (3b, $120 million) would commence within the first 3 months and ramp up (roughly linearly) over the next 12-18 months, dependent in part on how streamlined the NSF solicitation and award process turns out to be.

- The impact on networking vendors, including all of the ones listed above plus Atheros, LinkSys, Proxim, D-Link, and other WiFi and WiMax vendors, would be significant. Perhaps half of the investments would directly benefit them. A significant portion would also go to installers and university technology infrastructure staffs, averting, in many cases, layoffs. At colleges and universities in smaller communities, vendors such as Frontier and TDS also would see economic benefit.

Finally, in the longer term there would be direct benefits to companies who are desperate not only for new innovations but also trained graduates. Investments in new wireless and sensor networks (1d, $100 million) would provide a much-needed boost to tomorrow's wireless and sensor-technology startups. While such forward-looking campus buildouts may require 6 months of planning, the entire investment can be completed within 24 months.

The world is poised for a new revolution in eScience, eLearning and data-intensive computing. Ensuring that our nation is at the forefront of these critical fields requires more than infrastructure: it requires research investments that will produce the ideas and people that fuel innovation. Hence, research initiatives in broadband innovations (4a, $70 million per year) and DISC/eScience (4b, $70 million per year) are a critical element of this package, and over the first two years, these initiatives represent about 17% of the total cost (including the campus contribution to 1c). The benefits to companies such as Google, Microsoft, and Yahoo!, in new ideas and new manpower, are significant. In addition to setting the stage for the next great wave of innovation, this investment in research initiatives will quickly contribute to job creation, with the hiring of approximately 1500-2000 additional graduate students, faculty and staff. A number of additional high-priority high-payoff computing research initiatives are described at http://www.cra.org/ccc/initiatives.

The initiatives described here are vital for our nation's continued leadership in science and technology. As we enter the new era of eScience, infrastructure that sparks the creativity and training opportunities for a new generation of educated citizens becomes more crucial than ever. In addition to this long-term benefit, the immediate economic benefits of infrastructure deployment will result in new jobs and greater financial security for companies and industry sectors that need it.