Developing a Coherent Cyberinfrastructure from Local Campus to National Facilities: Challenges and Strategies

A Workshop Report and Recommendations

EDUCAUSE Campus Cyberinfrastructure Working Group
and
Coalition for Academic Scientific Computation

February 2009

EDUCAUSE
uncommon thinking for the common good

CASC
Coalition for Academic Scientific Computation

EDUCAUSE is a nonprofit association and the foremost community of IT leaders and professionals committed to advancing higher education. EDUCAUSE programs and services are focused on analysis, advocacy, community building, professional development, and knowledge creation because IT plays a transformative role in higher education. EDUCAUSE supports those who lead, manage, and use information technology through a comprehensive range of resources and activities. For more information, visit www.educause.edu.

The Coalition for Academic Scientific Computation (CASC) is an educational nonprofit organization with 57 member institutions representing many of the nation's most forward-thinking universities and computing centers. CASC is dedicated to advocating the use of the most advanced computing technology to accelerate scientific discovery for national competitiveness, global security, and economic success, as well as develop a diverse and well prepared 21st century workforce. For more information, visit www.casc.org.

Developing a Coherent Cyberinfrastructure from Local Campus to National Facilities: Challenges and Strategies

Table of Contents

Preface

A fundamental goal of cyberinfrastructure (CI) is the integration of computing hardware, software, and network technology, along with data, information management, and human resources to advance scholarship and research. Such integration creates opportunities for researchers, educators, and learners to share ideas, expertise, tools, and facilities in new and powerful ways that cannot be realized if each of these components is applied independently. Bridging the gap between the reality of CI today and its potential in the immediate future is critical to building a balanced CI ecosystem that can support future scholarship and research.

This report summarizes the observations and recommendations from a workshop in July 2008 sponsored by the EDUCAUSE Net@EDU Campus Cyberinfrastructure Working Group (CCI) and the Coalition for Academic Scientific Computation (CASC). The invitational workshop was hosted at the University Place Conference Center on the IUPUI campus in Indianapolis. Over 50 individuals representing a cross-section of faculty, senior campus information technology leaders, national lab directors, and other CI experts attended.

The workshop focused on the challenges that must be addressed to build a coherent CI from the local to the national level, and the potential opportunities that would result. Both the organizing committee and the workshop participants hope that some of the ideas, suggestions, and recommendations in this report will take hold and be implemented in the community. The goal is to create a better, more supportive, more usable CI environment in the future to advance both scholarship and research.

July 2008 Workshop Organizing Committee

Patrick Dreher, *Chair*	Renaissance Computing Institute
Vijay Agarwala	Penn State University
Stan Ahalt	Ohio Supercomputer Center
Guy Almes	Texas A&M University
Sue Fratkin	Fratkin Associates
Thomas Hauser	Utah State University
Jan Odegard	Rice University
Jim Pepin	Clemson University
Craig Stewart	Indiana University

Executive Summary

The rapid growth of information technology (IT) in research and education over the past decade has supported new initiatives, projects, and methods of interactions among faculty, staff, and students worldwide. Evidence of this expansion can be seen at all levels, from the individual to the campus and through to national computational resources. Today we see wide recognition of the concept of cyberinfrastructure.

For purposes of this report, we adopt the following definition:

> Cyberinfrastructure consists of computational systems, data and information management, advanced instruments, visualization environments, and people, all linked together by software and advanced networks to improve scholarly productivity and enable knowledge breakthroughs and discoveries not otherwise possible.

The expanded and sophisticated capabilities of CI have evolved in a disjointed manner, however. In many cases, faculty, staff, students, and researchers who have tried to access and integrate IT resources on a local level through the campus or national infrastructure have encountered serious roadblocks in interoperability, usability, and availability. Budget and organizational choices made at each level have exacerbated these problems. Each infrastructure layer has focused on growing functionality within that layer without considering how such capabilities interrelate to the other layers of the national CI ecosystem.

This practice by funding agencies and institutions of enhancing CI functionality within a single level has created an environment with dysfunctional access to available resources. For example, a number of federal agencies are making substantial investments in key components of nationally funded CI. At the same time, campuses are making local CI investments ranging from minimal capabilities up through multi-teraflop computational systems with support facilities. Lacking is the larger goal of developing a coherent, coordinated vision to leverage these capabilities among the individual, campus, and national facilities.

The Coalition for Academic Scientific Computation and the EDUCAUSE Net@EDU Campus Cyberinfrastructure Working Group recognized the importance of addressing these issues and jointly sponsored a two-day workshop in the summer of 2008 on the theme “Developing a Coherent Cyberinfrastructure from Individual Campuses to National Facilities: Challenges and Strategies.” An invited group of experts from across the research community, campus IT, and staff from national computational centers met in Indianapolis in July 2008 to:

- Identify key issues
- Identify possible options and strategies to build a coordinated CI
- Develop strategies and recommendations and information to advise the members as to how to leverage and implement a coherent CI on their campuses

- Recommend steps to integrate the nationally funded CI centers with the activities at the campus and individual layer

Workshop discussions focused on a wide spectrum of CI capabilities and technologies grouped under the broad categories of computational systems, information management, and the human/social aspects of CI. In each area workshop participants discussed the rationale for a coherent national CI strategy, potential short- and long-term recommendations to facilitate effective CI implementation on university campuses, and ideas for improving coordination between universities and funding agencies for better implementation of CI capabilities.

Four key areas within CI were identified where a focused effort to improve the current status would have major positive impacts. :

- Harnessing campus and national resources
- Information life cycle: accessibility, usability, and sustainability
- Identity management, authentication, and authorization
- Human resources and broader impact

Within each of these major categories, workshop participants offered actionable strategic and tactical recommendations that can be implemented today to help develop a coherent CI from the local to the national resource level. The strategic recommendations are:

- Campuses in partnership with national resource providers and governmental agencies should support, promote, and develop a coherent, comprehensive set of computing and data facilities.
- Agencies, campuses, and national or state network organizations must improve the aggregate national network infrastructure needed to address the data-transfer and remote resource access needs of a coherent CI.
- Agencies and campuses must work together to create technical and nontechnical architectures to enable researchers and other CI users to make the most effective use of campus and national resources.
- Funding agencies and institutions must fund both (1) operational implementations of data preservation to meet immediate needs and (2) research on data preservation and reuse to guide future activities.
- Federal agencies, disciplinary communities, institutions, and data management experts should develop, publish, and use standards for provenance, metadata, discoverability, and openness.
- Funding agencies, research institutions, and communities must collaborate to develop a combination of policy and financial frameworks to ensure maintenance of important data over time scales longer than the career of any individual investigator.

- Agencies, campuses, and national and state organizations should adopt a single, open, standards-based system for identity management, authentication, and authorization, thus improving the usability and interoperability of CI resources throughout the nation.
- Agencies and campuses should support a strategic investment in human capital and curricula in order to build a pipeline of qualified experts who can develop the full capacity of CI.
- Agencies and campuses should develop technologies and tools to use the emerging CI for education and scholarship.
- Agencies and campuses should invest in partnerships between industry and academia.

As individuals, campuses, and national facilities embrace opportunities presented by a robust and pervasive CI, it becomes an urgent priority to make them interoperate seamlessly. To achieve these goals will require a coordinated effort in the design and implementation of a bold CI strategic vision, robust CI architecture, frameworks for CI governance, and well-developed and coherent interoperability strategies. Without such concerted effort, the potential impact of investment in CI resources at all levels will not be realized.

1. Toward a Coherent Cyberinfrastructure

For the United States to remain competitive in the discovery of new knowledge and rapid pursuit of new research directions, it needs to maximize the use of advanced technology. Advanced technology provides for the creation of robust new tools that, when organized and coordinated seamlessly, allow the free flow of information, ideas, and results. Fully realizing this goal requires resources that extend from the individual faculty member through medium-scale campus layer resources to large national centers such as the NSF-funded TeraGrid and Department of Energy Leadership Computing Facilities. In the future, it may be necessary as well to leverage cloud computing efforts as they are deployed by industry and academic enterprises.

Despite progress toward providing coordinated access to national CI resources, there is room for improvement. Over the past 15 years, for example, we have seen significant advances in the development and deployment of a robust national network infrastructure. The evolution of the Internet enables relatively transparent point-to-point access to information technology resources between individuals and local campus layers up through national computational facilities. The academic community has capitalized on these communications advances, generating national and international collaborations and creating virtual organizations focused on educational scholarship and research activities.

The complex mix of advanced computing resources, people, and capabilities is sometimes referred to as cyberinfrastructure (CI). For the purposes of this report, we define CI as follows:

> Cyberinfrastructure consists of computational systems, data and information management, advanced instruments, visualization environments, and people, all linked together by software and advanced networks to improve scholarly productivity and enable knowledge breakthroughs and discoveries not otherwise possible.

Today, the nation's CI resources have the power to spark discovery and innovation. Tomorrow, they might enable transformation of the process by which we engage in discovery and innovation. Progress in deploying CI resources nationally remains fragmented, however, and lacks the overall coordination and planning that could seamlessly integrate the different users and layers. The absence of a well-developed, coherent interoperability strategy prevents the United States and its researchers from capitalizing on CI's full potential.

The NSF TeraGrid, for example, attempts to deliver a national-scale computing resource with tremendous computational power that is particularly suited for supporting large computational models. The TeraGrid focuses on projects requiring thousands of processing cores performing tightly coupled computations. The resulting implementation at nationally selected computing centers has fallen short of actively engaging and enabling a wide spectrum of potential users, however, and has thus not enabled scholarship in its broadest sense. A notable gap exists between the processes and policies used to access local, campus, and national facilities such as the TeraGrid and the processes and policies used to access non-TeraGrid computational resources and stored data. This barrier limits an individual researcher's ability to effectively leverage local, campus, and national resources in a natural and seamless way.

This problem is especially acute in the campus layer. The governance and culture of educational institutions present a complex mixture of resources that are not easily adaptable to the rapid changes of the 21st century. Furthermore, most campuses are organized on the principle of researcher autonomy and self-funding, with tenure and promotion processes focused on recognizing and rewarding individual accomplishments. Many current practices of funding agencies and hiring incentives for faculty contribute to the problem by structuring awards of funding in ways that encourage faculty to make individual CI investments that fragment and balkanize the campus CI.

Over the past decade the balkanization of campus CI has grown more acute due to commoditization of CI, permitting the rapid growth and deployment of more sophisticated IT technologies. Yet we are failing to effectively leverage these localized investments toward a coherent CI that can drive research and education both locally and nationally. New technologies allow researchers to work in large collaborations and address problems that a few decades ago would have been considered intractable, and these large groups frequently require facilities beyond the reach of any single university. As a result, no one location contains all the resources needed for a project. These complex collaborations increasingly have the characteristic of being multidisciplinary, multi-institutional, and multi-instrument, with intellectual, computing, and data capacity distributed among participating sites that span regional, national, and sometimes international boundaries.

Senior campus leadership must choose how to apportion limited resources across a spectrum of campus CI needs, from administrative applications such as e-mail, web, file, and print services to advocating for greater investment in CI for research and education, to enable and support deployment of disruptive new capabilities. Some campuses are moving forward and developing cross-disciplinary centers and campus-wide support structures to encourage these new paradigms for academic and research collaborations; other campuses have been resistant or slow to adapt to this model's implications. The operational requirements of large collaborations and virtual organizations have stretched the capacities of the campus IT staff to support CI, in particular the infrastructure critical to advancing the research and scholarship dimension. Similarly taxed are campus and regional organizations that build networks, computing facilities, and other critical support infrastructure, including identity management, data archiving, access to distributed instruments, and an increasing number of sensor nets producing enormous volumes of data.

Comprehensive CI requires a full spectrum of support and resources stretching from local labs, through the campus layer, and up to the national centers. The continued evolution of CI hinges on our better understanding and adapting to the complexity of this challenge. Leveraging CI resources and enabling new research and educational initiatives will depend on our ability to design and implement a coordinated strategic vision with coherent plans for implementation.

A roadmap for coherent CI will require an overall strategic plan that spans the local through the national layers. We also need to implement a workable CI best-practices advisory framework. Developing and articulating a framework for CI strategies at all layers will strengthen the value proposition supporting a business case for educational institutions to invest scarce resources into maximally leveraged CI deployments. Such an overall framework will spark federal, campus, and other stakeholders to invest in the development of critical elements of CI. These investments include fiber-optic networks, massive file systems, support for visualization, scalable authentication/authorization infrastructure, and curriculum materials to help students aggressively use the emerging CI.

These and many related topics have been part of a recurring discussion by both CASC and CCI for the past two years. During the summer of 2008, EDUCAUSE CCI and CASC jointly organized a workshop to dive deeper into the issues related to national coordination of CI. The workshop also considered the confluence of architectural and social questions being asked by NSF's Office of Cyberinfrastructure (OCI)Internet2, and other governmental entities with the goals of developing a deeper understanding for the problems we face and making recommendations for a path forward. The invitational workshop was hosted at the University Place Conference Center on the IUPUI campus in Indianapolis. Over 50 individuals representing a cross-section of faculty, senior campus IT leaders, national lab directors, and other CI experts attended the two-day workshop, which explored three dimensions of CI:

- *Computational systems,* including high-performance and high-throughput computing, networks/communications, visualization, advanced instrumentation, and other similar systems.
- *Information management,* including data creation, storage, handling, retrieval, distribution, interpretation, and security; policies on research data; long-term preservation; provenance; and metadata as well as identity management, security, authorization, and authentication.
- *Human/social aspects of CI,* including campus communities and outreach to nontraditional computing groups, education and training, CI-enabled learning, CI partnerships for faculty and virtual organizations, as well as industry, federal, and campus partnerships.

This report encapsulates the issues, comments, observations, strategies and recommendations resulting from the workshop. All breakout sessions identified software as a critical component to success in leveraging increased investments in CI. While the workshop did not have a specific software breakout, it is important to recognize that software such as middleware and applications are the glue that bring seemingly disparate IT technologies together and enable researchers to leverage increased investments in CI. Compounding the challenge is that software has always trailed hardware in making effective use of new features. For example, the current shift by vendors toward increasing clock speed as a way of increasing the overall CPU processing power rather than increasing the number of cores per processor creates a need to rethink many parallel algorithms. Educational programs will need to evolve to reflect this dramatic change as well.

2. Recommendations from the Workshop

The workshop "Developing a Coherent Cyberinfrastructure from Individual Campuses to National Facilities: Challenges and Strategies" identified a number of specific actions that, taken together, will increase the coherence and effectiveness of national CI that supports the increasingly data-intensive and computation-intensive research and other scholarly work of our universities. Diverse CI efforts have been made through national, campus, and other structures across the country, but greater coherence and coordination would support university work much more effectively. To that end, the workshop participants make the following recommendations.

2.1 Harnessing Campus and National Resources

Strategic Recommendation 2.1.1: Campuses in partnership with national resource providers and governmental agencies should support, promote, and develop a coherent, comprehensive set of computing and data facilities.

The following tactical recommendations support the strategic recommendation:

- Tactical Recommendation 2.1.1a: Integrate national resources with the campus layer in a way that ensures transparency, scalability, and ease of use.
- Tactical Recommendation 2.1.1b: Develop funding models that enable and demand integration of resources (data, computing, instrumentation) from lab to campus to national center.
- Tactical Recommendation 2.1.1c: Develop and deploy processes and policies that ensure flexibility for principal investigators (PIs) to choose local or national resources.
- Tactical Recommendation 2.1.1d: Governmental funding agencies at both the federal and state levels should implement contract and grant terms that encourage sharing and effective use of resources at all layers while eliminating disincentives for researchers to use campus or other shared resources.
- Tactical Recommendation 2.1.1e: Campuses should encourage resource sharing where local governance and policy allow it, thus helping improve scholarship.
- Tactical Recommendation 2.1.1f: Campus IT organizations should take an active role in exploring new technologies by serving as a conduit via the CIO (or equivalent) to promote and develop new capabilities and access to resources that are external to the campus.

Strategic Recommendation 2.1.2: Agencies, campuses, and national or state network organizations must improve the aggregate national network infrastructure needed to address the data-transfer and remote resource access needs of a coherent CI.

The national network consists of backbone, regional, state, and campus elements. This report emphasizes the performance and robustness of the end-to-end network that connects national and campus CI resources to each other and to their users.

Better communication is needed among constituencies on the academic and administrative sides of universities. Campus network planning must take into account both the general CI needs of the campus community and the special cybersecurity, performance, and robustness needs of CI.

The use of conventional perimeter firewalls, which might be appropriate for parts of the campus constituency, must not burden high-speed flows between on-campus users and resources and those off campus.

The following specific, more tactical recommendations support the strategic recommendation:

- Tactical Recommendation 2.1.2a: Campus networks must be designed to support cybersecurity while also supporting the performance and robustness needed by CI.
- Tactical Recommendation 2.1.2b: Network leaders must choose architectures and patterns of interconnection of (backbone, campus, and mid-level) network elements to support the broader coherent national CI.
- Tactical Recommendation 2.1.2c: Agency and campus leaders must invest, both locally in through mid-level and national organizations, to accomplish this.

Strategic Recommendation 2.1.3: Agencies and campuses must work together to create technical and nontechnical architectures to enable researchers and other CI users to make the most effective use of campus and national resources.

Realizing this strategic recommendation requires innovations and improvements in areas such as workflow tools, virtual organization frameworks, federated authentication tools capable of recognizing campus credentials, flexible authorization tools, data-access tools that use these authentication and authorization tools, scheduling and allocation tools, and inclusion of remote visualization and remote instrument access in these schemes. Policies must be developed that work in concert with these technologies to support collaboration and shared access to resources. The key in all these innovations is to focus on the combined national and campus resources as a coherent CI in support of research and other scholarship. Similarly, researchers and other CI users will need support to effectively use combined campus and national resources to best meet their needs.

The following tactical recommendations support the strategic recommendation:

- Tactical Recommendation 2.1.3a: Agencies must include campus CI leaders in planning the evolution of national CI resources such as the TeraGrid and the Open Science Grid.
- Tactical Recommendation 2.1.3b: Campuses must prepare to integrate new and existing campus resources into the resulting architectures.

2.2 Information Life Cycle: Accessibility, Usability, and Sustainability

Strategic Recommendation 2.2.1: Funding agencies and institutions must fund both (1) operational implementations of data preservation to meet immediate needs and (2) research on data preservation and reuse to guide future activities.

- Tactical Recommendation 2.2.1a: Since there are more unsolved problems than solved ones in the areas of data taxonomies, metadata, and provenance management, agencies and institutions should fund research to develop and operationally use better techniques and tools for long-term data preservation, discovery, and reuse. We have neither the tools nor the reward system required to ensure that important data already stored—and of potential value in perpetuity—can be maintained in usable form.

Strategic Recommendation 2.2.2: Federal agencies, disciplinary communities, institutions, and data management experts should develop, publish, and use standards for provenance, metadata, discoverability, and openness.

- Tactical Recommendation 2.2.2a: Research institutions and communities should develop, vigorously disseminate, and adopt standards for data provenance, metadata, discoverability, reusability, and openness for all phases of the data life cycle. Institutions of higher education and research communities should strive to achieve consensus on standards in these areas. Where data are published openly, standards should be developed for giving credit to data providers.
- Tactical Recommendation 2.2.2b: Research institutions must define internal data life-cycle processes, including identifying parties responsible for management, oversight, and delivery of services in support of data preservation (which, at different institutions, might be librarians, archivists, or the IT organization). Such people or organizational subunits then function as the stewards of the data for the university.
- Tactical Recommendation 2.2.2c: Research institutions should develop and adopt on an institutional basis standards regarding ownership of data within the institutions and from those standards derive policies on responsibilities for data preservation over time. Given that data might be valuable in perpetuity, research institutions and communities should investigate the important philosophical question of whether responsibility for long-term data preservation resides with individual institutions, libraries, virtual organizations, or federal funding agencies. There is a need for standardized mechanisms that will allow the storage, discoverability, and usability of data over long periods of time while maintaining information about the provenance and authenticity of data sets.

Strategic Recommendation 2.2.3: Funding agencies, research institutions, and communities must collaborate to develop a combination of policy and financial

frameworks to ensure maintenance of important data over time scales longer than the career of any individual investigator.

- Tactical Recommendation 2.2.3a: Current policy development by funding agencies for distribution of data must expand to explicitly address maintenance of data over periods of time longer than the career of a single investigator so that data collected with federal or state funding will persist as societal assets as long as they have value.
- Tactical Recommendation 2.2.3b. In addition to developing policy frameworks, federal and state funding agencies and research institutions must develop financial and management strategies that assure availability of funds for maintenance of data that have been identified as important long-term societal assets.

2.3 Identity Management, Authentication, and Authorization

Strategic Recommendation 2.3.1: Agencies, campuses, and national and state organizations should adopt a single, open, standards-based system for identity management, authentication, and authorization, thus improving the usability and interoperability of CI resources throughout the nation.

- Tactical Recommendation 2.3.1a: The global federated system for identity management, authentication, and authorization that is supported by the InCommon Federation should be adopted with an initial focus on major research universities and colleges. After an initial deployment in research-oriented functions involving research universities, such an identity management strategy for CI should be implemented generally within funding agencies and other educational institutions.

2.4 Human Resources and Broader Impact

Strategic Recommendation 2.4.1: Agencies and campuses should support a strategic investment in human capital and curricula in order to build a pipeline of qualified experts who can develop the full capacity of cyberinfrastructure.

CI is fundamentally changing the ways in which research is conducted and in which teaching and learning take place. Furthermore, the shortage of well-trained computational scientists impacts the rate of adoption of large-scale computations in industry as well as in academia. Thus, there exists an overall need for more “computational science” education—starting with undergraduate and graduate minors and extending to graduate programs—so that computational science becomes a more distinct discipline in its own regard.

- Tactical Recommendation 2.4.1a: Institutions should commit to supporting the development and delivery of modules, workshops, and courses to address the growing need for CI literacy.
- Tactical Recommendation 2.4.1b: Curricular materials for computational scientists should include systems, architecture, programming, algorithms,

and numerical methods and should prepare them to think across disciplinary boundaries.

- Tactical Recommendation 2.4.1c: National organizations and/or open-source mechanisms should be used to share curricular materials.

Strategic Recommendation 2.4.2: Agencies and campuses should develop technologies and tools to use the emerging CI for education and scholarship.

- Tactical Recommendation 2.4.2a: A diverse set of communities should commit to the implementation of advanced CI technologies before there is an obvious return on investment. Examples include deploying federated identity management systems, the Access Grid, data repositories, wikis, and other middleware technologies.
- Tactical Recommendation 2.4.2b: Investigate whether technological and organizational factors that support effective virtualization can be standardized or provided as commoditized infrastructure. Commoditized, on-demand computational and storage systems may offer practical and economical solutions.
- Tactical Recommendation 2.4.2c: Offer awards for supporting community services at all levels, including the development of new scientific applications, operation of technology infrastructures, and ongoing maintenance of these services. For example, create funding models that encourage the development, standardization, and reuse of CI infrastructure. These funding models should also encourage the involvement of technology experts, social scientists, and human-computer interaction specialists because organizational and technological issues are inseparable. An effective CI must integrate both.

Strategic Recommendation 2.4.3: Agencies and campuses should invest in partnerships between industry and academia.

- Tactical Recommendation 2.4.3a: These partnerships should work with businesses to adopt the use of computational science and supercomputing and assist the transfer of new computational science and supercomputing technologies from sponsored research projects to small and medium-sized businesses.
- Tactical Recommendation 2.4.3b: These partnerships should identify industry needs for new modeling software, adapt software to run effectively on modern supercomputer platforms, and provide a repository for sharing this software.
- Tactical Recommendation 2.4.3c: Academia and industry should adopt a sensible model for sharing intellectual property. The NSF Industry/University Cooperative Research Center program could provide a viable model.
- Tactical Recommendation 2.4.3d: Academia and industry need to develop effective strategies to encourage students from traditionally underrepresented groups to pursue academic careers in computational science and to address workforce needs in industry.

3. Building Blocks of Cyberinfrastructure: Issues and Opportunities

The work of creating a coherent CI from the local campuses to the national facilities must focus on major challenges such as:

- Interconnecting campus and national resources in a well-coordinated way
- Managing the information life cycle
- Identity management
- Supporting changes needed in organizational structures and dynamics in response to CI and the emerging importance of virtual organizations
- Education, outreach, and training

3.1 Harnessing Campus and National Resources

One of the linchpins of a robust CI is a concerted effort to coordinate and leverage the activities and interests of diverse stakeholders. To that end, policies and practices are needed that explicitly promote partnerships between individuals, institutions, regional centers, national centers, and industry. Progress toward that goal hinges on a confluence of trends in technology and critical related factors.

3.1.1 Relevant technology trends

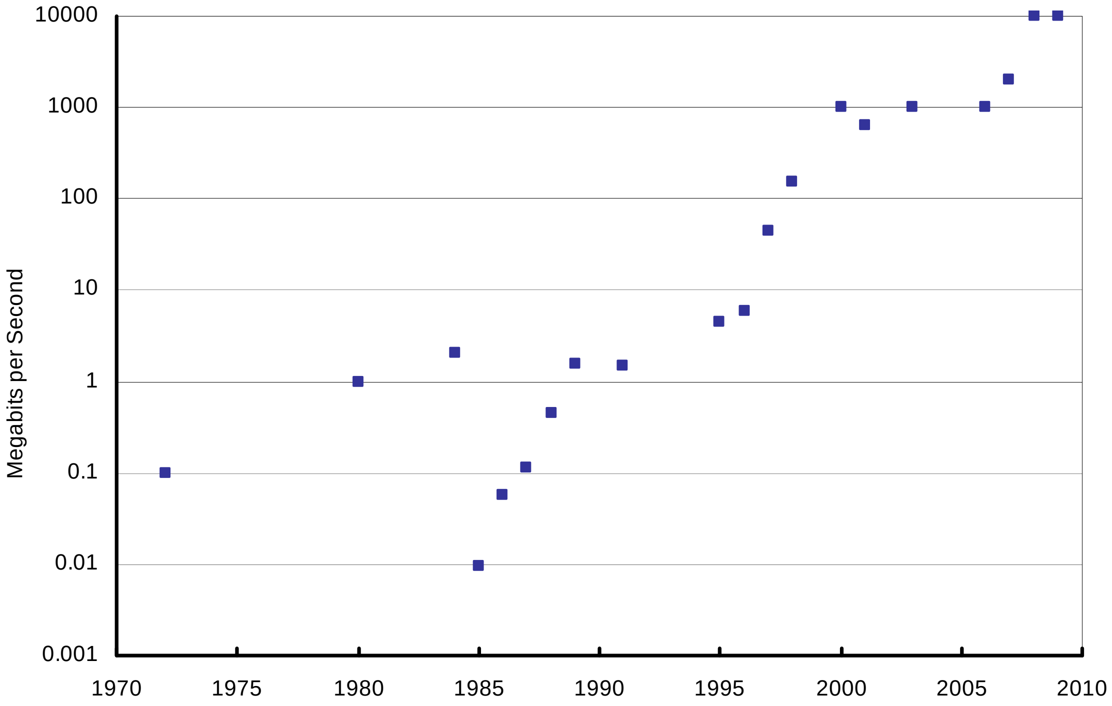
Storage trends: The capacity per dollar of rotating magnetic storage continues to double about every 12 months. Combined with other trends, this enables rapid increases in the size of data sets that must transit university-based, regional, national, and international computer networks. It also motivates investment in sensors and instruments that sample more frequently and with more bits per sample by making tractable the long-term storage of massive amounts of collected data. Coincidentally, this trend makes effective data management a greater challenge. An everyday example is the explosion of cell-phone cameras at ever-increasing pixel density. The same trend applies to science data created by instruments of all kinds, as well as the rapid digitization of social science and humanities data and making library collections available.

Network trends: The current decade has witnessed an overwhelming trend toward the aggressive use of owned fiber-optic cables, leveraged by wavelength-division multiplexing (WDM) (including Dense WDM over the wide area and Coarse WDM over metro and campus areas). Though progress is uneven, this trend will eventually encompass the national backbone, plus the regional/state and campus-layer network infrastructure. Over the past several decades, there has been significant growth in overall capacity. It is important to note, however, that while the bit-per-second data rate of individual circuits grew rapidly during the late 20th century from 50 Kbps to 10 Gbps, 21st-century growth seems to be in the numbers of parallel 10-Gbps circuits. This has led to explosive growth in aggregate bandwidth, counting our current wide-area fiber-optic networks (both university backbones and regional optical networks in much of the country) as capable of, say, 100 lambdas of 10 Gbps each. Nonetheless, the networks that result from this wide-area fiber deliver only several 10-Gbps lambdas. Several reasons contribute to this disparity:

- The cost of 10-Gbps optronics available in switches and routers is falling very slowly during the present decade.
- The cost of metro-area carrier-based 10-Gbps circuits is also falling very slowly, for reasons that include market forces and diminished competition.
- The network architectures in use by the research community are generally not capable of taking advantage of the large numbers of parallel links that result from the emerging fiber/WDM physical infrastructure. Other countries are leading in the architectural innovations needed to make use of parallel links.
- Transport protocols and their congestion-control algorithms are not keeping up even with the limited improvements seen in raw circuit speeds.

As a consequence, our ability to move the increasingly large scientific data sets that result from other trends is growing at a slower pace than our ability to generate and store data. To put these technology (and market) trends into perspective, recall some historical data regarding ARPAnet bandwidth versus computational power. Comparing a time when the ARPAnet was provisioned at 50 Kbps rates with large-scale systems at 1 Mflops, today we have ordinary campus clusters with 10 Tflops, connected at perhaps 10 Gbps. Thus, while the ratio of computing speed to wide-area circuit bandwidth in 1970 was 20 flops/bit, today it is 1,000 flops/bit. This computing-to-network performance ratio has thus increased by a factor of about 50. An even more compelling increase would apply to the ratio of large data-set size to network bandwidth. While the present situation indicates that data-set size is doubling every 12 months and processing capacity is tracking Moore’s law and doubling every 18 months, the university-to-university network performance is growing much more slowly, doubling every 48 months at best. Without concerted investments, these trends will likely continue. See Figure 1.

Figure 1. Network Capacity Based on a Limited Sample from the University Participants Involved in the Workshop



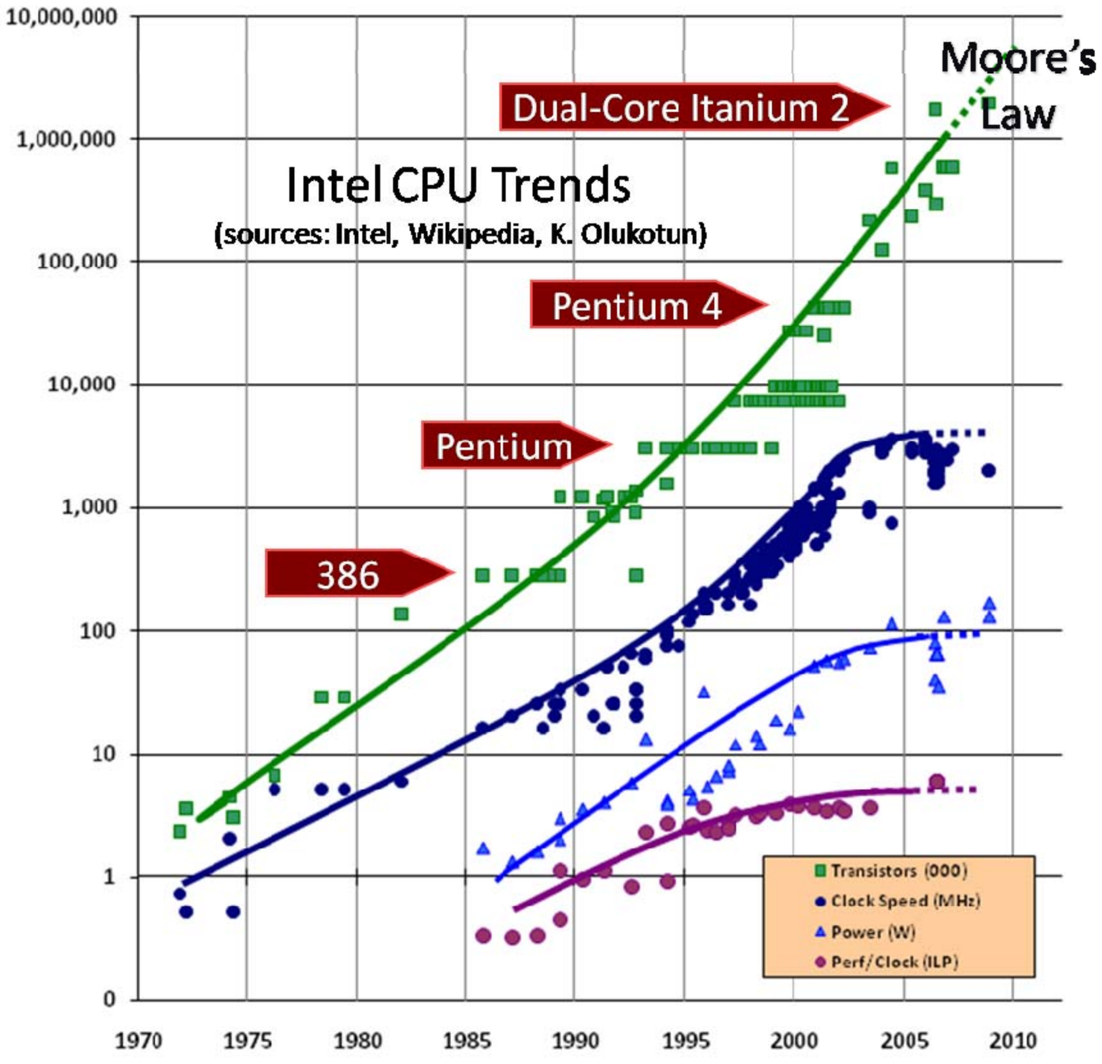
Source: Chart created by Jan Odegard, Rice University; data collected by Guy Almes, Texas A&M University

Another reality is that applications requiring low-latency human-computer interactions will work best when the computer is physically near the human. The latency induced by the combination of long distances and the necessary number of network transition points can significantly interfere with human-computer interactions at long distances. These factors have important consequences:

- The practical ability of science users to move data to/from campus resources will increasingly exceed their ability to move data to/from national resources. Thus, in the short term at least, campus resources will have increasing importance for data-intensive CI applications.
- For the sake of conserving the value of national resources, the community must work together to address the various causes of anemic growth in actual wide-area network performance.

Computing trends: Moore's Law, which posits that the number of transistors per chip doubles roughly every 18 months, should continue to hold true for at least the next decade. In a manner reminiscent of the comments above concerning WDM, the 21st-century impact of Moore's Law is primarily in making available a growing number of 3-GHz (or so) processing cores. While this might sound promising at first, each core is not getting faster (the limitations of physics). If we want to continue scaling application performance, we need to rely on advances in software that enable large-scale parallel computations. Figure 2 shows CPU trends in terms of transistor counts and effective performance gains. Note the increasing gap between the green curve and the dark-blue, light-blue, and purple curves. This is alarming because it means easy performance gains are past. To leverage the effect of future technology capacity improvements, we need to develop parallel codes—a very difficult task not suitable for the average programmer.

Figure 2. Comparing Transistor Counts Against Dates of Introduction and Effective Performance Gain Achieved by Technology Shrinkage



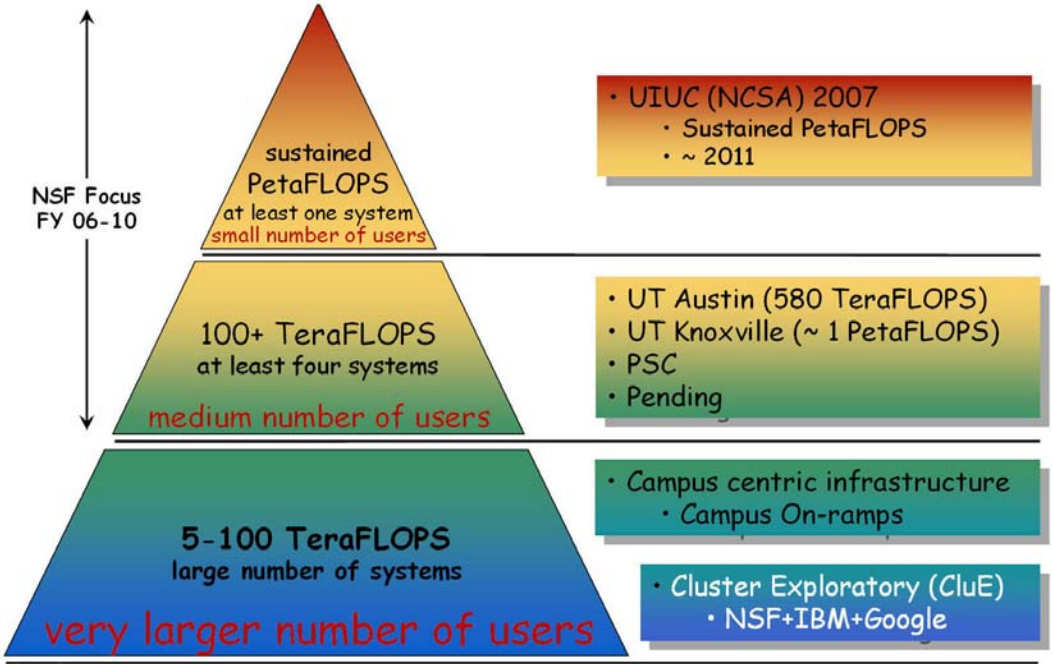
The result is a computing environment where improvements in performance will require effective use of a rapidly growing number of processing cores. For those applications characterized by large numbers of independent computations (also frequently referred to as "embarrassingly parallel" or "throughput computing"), this presents relatively few problems. For those whose applications require very tightly coupled parallel algorithms, the current MPI and OpenMP programming models require great skill on the part of the algorithm developer and programmer to make effective use of more than, say, 2,000 processing cores. Without significant sustained investment in software, algorithms, and tools, extremely large clusters such as the future NSF Track 1 and even (to some degree) today's NSF Track 2 systems will provide and support excellent scalability for only a very few software applications. Correspondingly, it has been more common to see several 1,000–2,000-core clusters available on many individual campuses. This implies that it is not only practical but also optimal to solve a large number of computational problems at the campus level.

- For computationally intensive applications, the multi-core phenomenon makes it less expensive to build and manage medium-sized clusters (in 2008 terms, roughly 2,000 cores).
- These systems can use shared memory techniques within single nodes as well as increasingly affordable high-speed interconnect fabrics between nodes.
- Since the state-of-the-art tightly coupled parallel applications cannot efficiently use more than 2,000 cores, this often means large national resources, rather than meeting their full potential, are used as aggregations of medium-sized clusters with very expensive massive interconnection technology.
- Medium-sized campus layer clusters are efficient at running high-throughput computations. In this role, they offer excellent economies of scale and are a strong complement to national resources.

As these trends evolve, the need for closer coordination and coherence between campus and national resources will increase.

Figure 3 shows the structure of the existing computing landscape as viewed through the traditional lens of the Branscomb Pyramid. At the top of pyramid are national centers with CI resources characterized by tremendous power and a focus on supporting large computational models; they reach relatively few users today. At the bottom of the CI pyramid is the campus infrastructure, ranging from a single researcher's lab to regional collaborations. Similar illustrations are possible looking at other national and open resources.

Figure 3. Structure of the Existing Computing Landscape as Viewed through the Lens of the Branscomb Pyramid in the Context of the NSF-OCI Infrastructure Investments



The TeraGrid has greatly advanced coordination of access to national resources and has set a vision for a nationally integrated computing landscape. However, the emergence of coherence between national computing resources has not extended to create coherence, or even particular ease of migration, between the TeraGrid and computational resources at the campus layer or at national centers outside the TeraGrid. Further, we must recognize that success in broader integration across systems and levels is likely more a social challenge than a hardware and software challenge. As a result, we fail to benefit from the potential value of broader coherence and integration. For science users this frustrates what should be natural ways to use campus and national resources in a seamless way—for example, by sharing scientific data sets across resources or setting up workflows whose steps use both campus and national resources.

Nontrivial barriers inhibit entry for local users trying to access TeraGrid resources. Ideally, no barrier would separate the local CI and the TeraGrid, and users of local CI resources could transition between local, regional, and national resources with minimal hurdles. Allocations on the TeraGrid could be set aside for local campus centers to transition large users, with both partners getting "credit" for these collaborations. In addition, a number of initiatives focus on delivering computing that has the potential for further disrupting the CI landscape, such as:

- *Cloud Computing*. One relevant technology trend to keep an eye on is how cloud computing, particularly applications that use commercial infrastructure, integrates with CI. Cloud computing is an evolving term. It can be thought of as a virtualized elastic (computing) resource from which end users can purchase access as needed. Note also that the academic community has been deploying variants of cloud computing since the early ARPAnet era. As we deploy campus, regional, and national clouds today, identity management is increasingly a key factor. Emergence of several commercial providers of cloud computing services will pose new challenges and questions on several fronts such as, if it becomes economical to buy computational services from vendors, how would or should it impact future development of campus, regional, and national CI; how tightly would this CI have to be integrated with that of private-sector companies; would federal funding agencies have new policies and guidelines on what computing services individual or teams of investigators can acquire from private sector companies, and if those companies would have to be U.S.-based for data or national security reasons.

With significantly more demand from academic, federal, and industry customers, a more robust computing services industry segment might emerge in the next few years. It will depend to a large extent on how research computation-related CI is funded and deployed in the future. The computing services industry, largely based in the United States because of availability of skilled workers, can offer not only on-demand computational capacity but also value-added simulation and analysis services. This is an area in which U.S.-based companies, in stronger partnership with academia, can be globally competitive. The academic community must think through its own set of challenges and opportunities in building its CI and embracing cloud computing.

- *Dealing with Complexity*. For decades, two factors have remained constant in supercomputing: the suppositions that supercomputers are too complex to program, and that the task of programming should be simplified. Ongoing work has sought to mitigate these challenges. For example, the Ohio Supercomputing Center’s “Blue Collar Computing” initiative aims to increase innovation by creating tools that dramatically ease the task of creating parallel programs in the 32–1,000-processor range. This approach dramatically reduces the barriers from “pretty good” to “quite good” use of parallel computing in science and engineering challenges.

  Another approach is to create interfaces that reduce the complexity of interacting with advanced CI. One particular approach, exemplified by science gateways in the TeraGrid, relies on web-based applications providing end-to-end support for scientific workflows. The LEAD gateway, for example, allows weather researchers to select Doppler radar data feeds in real time, preprocess the data, perform ensemble predictions of patterns of severe weather (such as tornadoes), create a visualization of those simulations, and see the results of the visualization on the screen of a laptop or cell phone. In

> so doing, the weather researcher is accessing sensor nets and using large-scale supercomputing, data management systems, and visualization software, all from a graphical user interface sufficiently simple that an undergraduate student in weather forecasting can use it.

Having an adequate social architecture and models for how we develop productive partnerships between the local campus CI and national resources will be critical in our effort to develop a nationally coherent CI that will encourage participants at all layers to coordinate CI investments by offering those at every level the opportunity to multiply their investments. While computational resources have been mentioned in these examples, similar observations apply to other CI elements, including fiber-optic networks, massive file systems, and visualization facilities. Without simplifying and streamlining authentication and authorization, however, as well as developing a process for allocation of infrastructure, national coordination will fail. Similarly, it is critical that we recognize the pervasiveness of CI and support the development of curricular materials that can be used across the educational ecosystem (K–16+) and allow all educators to participate in educating the current and next generations of students about the potential of understanding and using CI in all fields of scholarly activity.

Campuses today have to integrate, coordinate, and manage an increasingly complex set of CI resources, from supporting the laboratory needs of individual researchers to supporting regional collaborations. Increasingly complex problem spaces require large communities of researchers to investigate areas of interest. Funding agencies increasingly are creating large virtual communities at the campus, national, and international levels. This evolution has put a significant burden on the campus support organizations that support the communities and academic researchers with CI.

The complexity of building the supporting CI as well as supporting governance must be recognized and understood. The support structure on a campus can be implemented in many ways, such as research computing under a vice president of research, or an academic computing group in central IT. Each campus needs to look inward at their campus governance and culture while looking outward to the collaborations increasingly demanded by complex research and academic needs. Funding agencies also need to recognize the complexity and difficulty of supporting these large collaborations and support building local campus CI and partnerships as well.

The current historic moment, at which architectural questions are being considered within both NSF'sOCI and Internet2, presents us with a compelling, though difficult, opportunity to address these issues on behalf of America's research and educational community.

3.1.2 Flexible Use of Campus and National Computing Resources

The work of scientists and engineers whose research is computationally intensive would be strengthened if they could make flexible use of both campus and national resources. Examples of this include the following:

- Computational scientists often develop tightly coupled codes on medium-sized clusters. As their need for access to more computing resources grows (whether the code scales or not), they should be able to run those codes on larger national resources.
- Computational scientists often develop codes that, from one day to the next, might best be run on a campus or a national resource. Enhanced integration between campus and national resources will permit users to more easily move workload between local and remote resources as the computational workflow dictates.
- Computational scientists often build simple or complex workflows, in which some job steps are large tightly coupled MPI applications and some job steps are independent parallel applications. The ability to schedule different job steps on different and possibly distributed resources would enhance throughput and decrease time to discovery in workflow-oriented processes.
- Computational scientists might need to access, and perhaps modify, data sets that reside on campus or national storage resources.

While in the examples above we used the term *computational scientists*, the benefit is not limited to science but is equally applicable to research in engineering, social sciences, and the humanities.

3.1.3 Escalating Power, Air Conditioning, and Support Demands

Campus computing resources are increasingly critical to the research enterprise and are, at many campuses, viewed as an important enabler for research and discovery. Such resources come at a cost with which many universities are struggling. Senior leaders across the academic ecosystem face escalating demand requiring sustained funding for deploying CI, and yet the total cost of ownership of these investments is often not well understood. Further complicating the relationship is the fact that the positive impact on the institution of such infrastructure is difficult to define and detail on the same timeline as campus budget cycles. Positive impacts are often anecdotal or claimed in ways that are not testable (for example, there is no experiment that allows proof of the contention “we received funding for this project thanks to our campus cyberinfrastructure”). The community increasingly recognizes that the cost of procurement is in most cases not the biggest expense of CI; rather, a large—and perhaps the larger—cost lies in support staff, power, and cooling. Aggressive CI investments also generate needs for renovations of existing data centers and laboratories and, at an increasing number of institutions, require new facilities. The days when only wet-lab scientists need expensive laboratories with fume hoods are behind us. Moving forward, universities must factor in needs for computing, storage, visualization, and advanced communication equipment to support faculty across science and engineering, and other disciplines will soon follow.

It is critical that these issues be addressed immediately. Additionally, many of the issues will benefit from coordination beyond local campuses. Specifically:

- Strategies are needed to address growing power and air-conditioning demands. Current federal funding programs often provide computing equipment funds directly to individual PIs, often resulting in small systems being deployed close to the researcher (the computer in a closet[1]). If, instead, funding programs were more cognizant of this challenge and worked to develop incentives for investigators to participate in more sustainable campus computing investments, universities would help researchers better understand the total cost of ownership and over time be able to make smarter CI investments—conserving power and air conditioning. Several institutions are embracing the concept of condominium computing, and it might be relatively easy for funding programs to recognize this and make sure investigators maximally leverage funding by partnering locally.
- A key component of the escalating cost of supporting CI is staff support. While federal funding should remain focused on funding students engaged in research, the federal government and funding agencies must recognize that the cost of CI is increasingly associated with staff support. There are two components to this, both working against universities engaged in supporting CI:
    - Few CI funding programs that are campus-centric (i.e., not national resource providers) permit much if any budget for support staff. As a result, universities incur this growing cost with no direct revenue stream, only a second-order relationship.
    - In the cases where universities invest local funding in building and deploying CI in support of research, the 26-percent cap imposed by OMB-A21on administrative costs on federally funded research results in a limited (if any) impact on cost recovery that can be used to support the growing need for support staff. Adjusting or even eliminating this cap would permit universities to recover the true cost of supporting federally funded research.

As a consequence of these issues and in particular the lack of funding (direct or indirect), universities are struggling with how to develop local plans for sustainable CI deployment in support of research. Encouraging or even requiring more aggressive local sharing as well as eliminating the somewhat arbitrary cap on administrative cost would help universities address their power and air-conditioning problems and be more wiling to make strategic investments in CI and support staff.

3.2 Information Life Cycle: Accessibility, Usability, and Sustainability

Multiple challenges beyond those imposed by technology must be addressed to achieve successful information management throughout its life cycle, from appropriate policies to sufficient funding to business models for storage and archiving.

3.2.1 Managing Data for Information Content

Data are being generated in the United States at a rate that outstrips the ability to efficiently move it long distances, yet access to that data is increasingly critical to researchers across the entire research enterprise. Today data are stored in repositories that provide capabilities for access by individuals and researchers worldwide, with some locations serving as data archives "in perpetuity." However, the amount of data nearly doubles every year, while our ability to access the data is expanding at a much slower rate.

Providing access beyond the traditional geographic boundaries of the campus layer and across time scales of generations further complicates the formulation of stewardship, authorization, and access policies. Uniformity and standards as well as CI-focused practices and policies are needed across the information life cycle, starting at the point where raw data are created and continuing through long-term stewardship of both the raw data and the knowledge derived from the data. Much of this service, stewardship, and archiving has historically been the role of libraries and museums. As society transitions from a focus on holding books and physical artifacts to managing digital data and knowledge artifacts, libraries will likely evolve along this path as well. Indeed, such evolution is essential to support the effective development of research, discovery, and preservation of knowledge in scholarly communities.

3.2.2 The Data Life Cycle

When individuals create a data collection, they generally have the freedom to chose or create an arbitrary data context. In many cases this context is implicit and never explicitly documented. The second phase of the data life cycle occurs when an individual makes data available to another person or group. At this point, data formats, semantics, and the allowed manipulations of the data must be defined. In the case of data created by a group, these definitions may need to occur prior to creating the data so that the members of the group agree on what the data collection really means. Each time the data are made available to a broader community, the context may become more structured in terms of allowed formats and semantics. Each propagation of the data then constitutes another step in the data life cycle. The traditional phases of the data life cycle are:

- Creation
- Migration into shared collections
- Publication into digital libraries
- Preservation into persistent archives as reference collections

The two most urgent concerns regarding the information life cycle are (1) there are multiple data management standards and practices for managing data, metadata, and storage; and (2) there are no widely agreed-upon principles as to which entities (individual researcher, campus group, virtual organizations, or national entities) are responsible for maintaining the data life cycle and access to the data over long periods of time.

3.2.3 Policy Challenges in Information Management

Data files and collections of information are usually held by some combination of individuals, collaborations, or institutions. At present, there is no generally recognized and accepted practice or procedure for vetting issues that arise around data collection, storage, curation, and preservation among individual researchers, the campus infrastructure, or national facilities. These three distinct groups usually view the custodianship of data as fitting into two broad categories: it is either institutional data, or it is data from within a research project, collaboration, or community of practice.

At the institutional level, there may be an organization, most likely the library that has responsibility for long-term data stewardship. However, the capabilities, functions, organizational roles, funding, and other parameters for these organizations vary enormously.

At the community-of-practice level, data are initially organized as shared collections within projects. As the project matures and the quantity of data expands, it is usually desirable to assign the data to an identified institutional caretaker that can address and implement processes for cataloguing and publishing it in digital libraries or institutional repositories, some of which enforce restrictive access to only the members of that community. Finally, if the funding is available, the data are preserved over time as reference collections for general use in education and scholarly research.

Increasingly we are creating data that might be of value in perpetuity (for example, detailed weather data, or records of population genome sequencing projects). Today we lack the policy, financial, and technical tools needed to maintain data usability for a decade, much less a century or more.

Compounding these challenges, several key stakeholders and stakeholder groups—including individual researchers, individual institutions, communities of practice, and federal agencies—have somewhat different perspectives regarding who ultimately owns the data and who has the responsibility for managing its preservation, accessibility, and usability over time.

- *Individual researchers*. Some federal agencies now require that availability of data be maintained over time and have penalties for individual researchers who fail to meet these criteria. While such steps create tangible benefits to individual researchers and the U.S. scientific community in the short run, by themselves they will not result in preservation of data over the course of many decades. This approach also does not offer any aid with data collected outside of federally funded research.
- *Institutions of higher education, libraries, and museums*. Universities have actively pursued research, teaching, and other scholarship for hundreds of years. Similarly, libraries have routinely preserved information and artifacts for several hundred years. Museums began storing information and artifacts more recently, perhaps going back a few hundred years, but in many cases have created clearer strategies for long-term persistence of the objects stored. As institutions in society, universities and museums are likely to face increased pressure to provide open-access digital data and digital artifacts;

clear policies and processes for long-term persistence of digital information are needed and critical for publicly or privately funded institutions.

- *Communities of practice.* Both the institutional and community-of-practice perspectives use multiple types of data management applications that address different phases of the data life cycle. The present solution, while great for a single community, sometimes makes it difficult for multiple communities of practice to address larger scholarly research questions and topics that require multiple data collections to operate coherently. Federation of information resources requires exchanging structured information between resources, but since each information resource maintains internal information needed for interaction with the a specific collection by the community, an interface is required that retrieves the parameters needed for subsequent operations.

  An example of such an interface is the integrated Rule-Oriented Data System (iRODS) mounted collection. This consists of a set of standard queries needed to acquire the information that will enable manipulation of a file in the remote information resource. Since each information resource uses a different protocol, separate structured information resource drivers are written that map the queries to the required forms. This approach supports building a shared collection across files in independent information resources, enforcing management policies on the shared collection, and validating assessment criteria.

- *Federal agencies.* Many types of data are maintained by the federal government, from data generated in classified projects to data held by the National Library of Medicine and the Library of Congress. A larger federal role in long-term maintenance of data seems plausible but difficult to implement because as soon as data are aggregated across the country, it might be hard to properly identify data of long-term value and manage the storage of all appropriate data.

3.2.4 Technology Challenges in Information Management

A minimum standard for data preservation and metadata definitions would be of great value. While there might not be universal agreement on taxonomy and metadata structure for any particular area, there would likely be agreement on the basic elements required in a metadata catalog, and these definitions should themselves be made available and preserved. It would benefit universities a great deal if there were a model for data preservation that creates a basic structure but leaves room for universities to modify the details according to their own practices, procedures, and policies.

As data sets continue to grow, simplistic replication is not feasible. We need to develop both the necessary policies and technical solutions to support controlled, secure access to (campus, national, and global) data sets from (campus, national and global) computational resources. The emergence of clouds in the contexts of computing and storage will further drive the needs for such standards. These policies and practices can range anywhere from fully restricted to unrestricted access available to anyone on the Internet. (This topic is covered in the section below dealing with identity management; authentication, and authorization.)

3.2.5 Business Models for Information Management and Data Preservation

Data management involves both active management of the data and its preservation in an archive. From a technological standpoint it might not matter where data are located as long as you can access the data when needed. Still, at a practical level the question of who pays the costs to store and archive this information will affect the decision as to where data are stored (if they are stored), how access is controlled, and who gets credit for providing access to the data.

Ideally, the financial dimensions of data preservation should align with a university economic model. We cannot always assume that funding will come from sources outside of the university. Within a university, institutional funds can and will probably be used to preserve certain types of data collections deemed to be of particular value to the institution. A key challenge will be indexing the material to support discovery and browsing; managing data so that it is searchable will be more expensive than simply storing it.

Achieving sustainable data storage and archiving is challenging. As with computing infrastructure, the community lacks a sustainable funding model for supporting data storage. This is particularly important when the costs for managing data are rising faster than unit costs for data storage are falling. One option to explore is negotiating with the data owners to charge a fixed fee for a given period of time for the management and/or archiving of a fixed amount of data. The analogy would be libraries' procuring not simply books but also access to data archives in the future. Another option is to internally pay the costs to store and/or archive the information and then charge users individually to access the information for a fixed period of time. In both cases, it is unlikely that full cost recovery of storage and archiving can be achieved, so an additional source of funds will be needed to subsidize these operations. Other benefits may accrue, however. For example, institutions that have assembled unique data collections and the attendant expertise to use those collections are finding that such resources help attract faculty, students, and other research projects—and with them financial support.

3.3 Identity Management, Authentication, and Authorization

Identity management refers to the linkage of individual people with their electronic credentials and identities. *Authentication* is the component of this process that deals with verifying that a person asserting to have a certain electronic identity is indeed the proper person. *Authorization* is the process of determining what rights and capabilities are granted a particular electronic identity. Within tightly defined administrative groups, these challenges are met by existing technologies. Working across administrative domains—necessary in creating a nationally integrated CI—remains a challenge.

Any researcher today who uses *n* different CI resources likely has *k* different electronic identities. In most cases $k = n$, and only in a few cases is *k* less than *n*. Because each institution tends to implement a local solution that satisfies the needs for local computer security, the crux of the problem is developing a solution that applies broadly across institutions and interfaces with domain-based identity management. Flexible and nimble use of multiple advanced CI resources will be possible if and only if each researcher has one or at most a very small number of secure identities (*k* much less than *n*) recognized across a broad swath of U.S. (or better yet, global) CI resources. Widely usable authentication will decrease barriers to the use of CI at all levels. Eliminating the barriers created by the authentication management problems is a major challenge that will yield major benefits.

What is needed, therefore, is an identity management and authentication set of best practices and a system that is widely usable for U.S. researchers as well as international collaborators, service partners, and service clients. Adopting one standards-based solution to the authentication issue and adopting it across multiple administrative domains and levels of the CI ecosystem can maximize ease of access for individuals and facilitate ease of activities for virtual organizations. For example, the granting and revoking of authorization could be changed rapidly, supporting the work of virtual organizations and enabling implementation of best practices locally. This is a difficult problem because it is hard technically and because it requires social change and development of trust relationships and authorities (as with the certificate authorities) being used in grid computing today).

To succeed, such an authentication system should be based on open standards. It should be structured to allow the vast majority of U.S. postsecondary educational institutions to participate in a straightforward way. The implementation of an overarching identity management solution should not be confined simply to CI. The NIH has already announced adoption of one solution for identity management—the framework supported by the InCommon Federation. The federating system supported by InCommon is the only solution that is (1) based on open standards and (2) could be implemented within a matter of months. A relevant analogy could be the emergence in the early 1980s of many relatively equivalent network protocols that were technically sound but did not seamlessly interoperate until the decision was made to use TCP/IP. Here, the adoption of a single solution and diligent pursuit of its implementation is the key to changing the ability of researchers to flexibly use a national CI.

Given sufficient will, and leadership by the NSF in adopting one solution across the board, it should be possible to have a unified, standards-based approach to identity management in place in 18 to 24 months for many NSF functions related to research universities. With such an initial deployment in effect, broader implementation beyond research functions should be possible, and smaller campuses could build on lessons learned at research universities and start the process of transitioning local systems to integrate seamlessly with systems outside their respective campuses. Due to its relationship to U.S. higher education, NSF's leadership could have a profound impact on the U.S. national CI ecosystem.

3.4 Organizational Dynamics

CI concerns extend beyond deployment of computing and storage technology and identity management. There are critical considerations in the area of organizational dynamics, including organizational structures that need to be developed. Furthermore, the rise of CI is itself helping create fundamental shifts in research, education, and training paradigms.

3.4.1 Shifts in Organizational Structures

Several trends motivate support for scalable CI. These include:

- The concept of "Large Science" involves scientists at an increasingly large number of universities. The Large Hadron Collider (LHC) and other physics collaborations, for example, have driven the development of the Open Science Grid. The TeraGrid emphasizes support for "extreme" computational and data needs. The Network for Earthquake Engineering Simulation (NEES) includes NEESit, a CI focused on sharing data and shared participation in remote experiments.
- Data-intensive scholarship (content management, Creative Commons, open education repositories) is emerging in a similar way, though with more emphasis on data management and less on high-performance computing. One technology worth following is the application of the Google map-reduce primitive to nontraditional computational (cluster) applications. 3.4.2 Growth in Shared/Virtual Organization Services and Needs

3.4.2 Growth in Shared/Virtual Organization (VO) Services and Needs

A *virtual organization* (VO) is a group whose members and resources may be dispersed geographically and institutionally yet who function as a coherent unit using CI (see http://www.ci.uchicago.edu/events/VirtOrg2008/). Historically, VOs have formed around access to and use of high-performance computing or grid resources. However, the use of the term (and actual usage in scientific practice) is becoming much more flexible. An excellent example of a rapidly assembled VO was the group of experts collected to deal with the outbreak of severe acute respiratory syndrome (SARS) in 2002 and 2003, who in some cases communicated only through telecollaboration. In this case the VO approach allowed assembly of experts to be done quickly and, because initially the mode of transmission of the disease was not understood, safely. Safety is not often a consideration, but addressing the most pressing scientific needs facing the United States today might often require flexible creation and modification of VOs.

Keys to supporting VOs include the ability to (1) flexibly interconnect different elements of CI across different levels of the U.S. academic ecosystem, and (2)support collaboration and discovery as appropriate with computational, data management, and visualization tools. For the local campus CI provider, it is important to understand what demands VOs place on CI and how they may drive CI evolution. Conversely, enabling the greatest possible effectiveness of VOs may be one of the key justifications for local investment in CI and the national integration of CI.

3.5 Human Resources and Broader Impact

In any consideration of CI, human resources are vitally important. One critical CI consideration is the broad need to bring the higher education community up to competency in computational scholarship, both within disciplines and in a wider, cross-university sense. In powerful ways, CI is changing fundamental paradigms in the way research is conducted and how teaching and learning take place.

- *Research paradigm.* The advent of CI is essentially reshaping scholarly practice and inquiry around a new kind of science and creative activities enabled by CI, including developing and maintaining vibrant scholarly campus communities not restricted to traditional CI users.
- *Teaching and learning paradigm.* CI is also enabling teaching and learning with new tools, techniques, and practices. Professors are taking on new roles as mentors of students, for whom CI in turn makes possible deeper, hands-on engagement in learning. Teaching technologies developed over the past decade allow content creation and updating on an accelerated time scale.

No single approach has emerged as to how technology can transform education, but clearly considerable experimentation is under way. The open education resources (OER) movement, for example, represents a great opportunity to disseminate knowledge around the world. *Open education* is defined as digitized materials offered freely and openly for educators, students, and self-learners to use and reuse for teaching, learning, and research. Its focus, simply, is on making knowledge available for education. This is possible today as never before because so much information is in electronic form and thus easily shared, adapted, and improved. If quality content is provided, open education tends to flourish and spread.

3.5.1 Education and Training

While our technology-rich economy and society place ever-higher premiums on advanced skills, our educational and training techniques have changed little over the past few decades. Further, it is not always clear that the new techniques have positively transformed the way we learn, nor have we made great strides in enabling global lifelong learning. With the pervasiveness of IT in modern life in general, and in science and engineering in particular, we need to rethink the place of IT and the role of CI in the curriculum across the academy. Ubiquitous access to CI is fundamentally transforming scholarship, and it is critically important that we recognize this and educate and train the workforce of the future to master the skills required to embrace and leverage CI in their respective disciplines and to exploit it when available. In turn, CI offers academia rich opportunities to improve the quality and reach of teaching and learning.

A recent report from the President's Council of Advisors on Science and Technology (PCAST) makes several important recommendations regarding workforce development aimed at increasing the supply of professionals with bachelor's, master's, and doctoral degrees in networking and information technology (NIT).[2] While the PCAST recommendations focus on actions that should increase the supply of skilled professionals in the United States in the short term, it is critically important that the academic community not only embrace these recommendations but also expand programs such as STEM (Science, Technology, Engineering, and Mathematics) in addressing long-term needs. In fact, while STEM shows signs of success, we need to continue to strengthen and expand the emphasis on STEM disciplines in elementary and secondary education so as to increase the absolute numbers and relative percentages of high school graduates who plan to enter college in an NIT-related discipline. Furthermore, we should expand the scope of STEM by recognizing that IT is a universal enabler and including computing as a core component for a C-STEM program.

On one hand, each institution needs to consider how they teach “digital fluency,” which is the basic understanding of IT and CI that every educated citizen needs to have. On the other hand, with the growing importance of computational methods in science and engineering, each institution needs to reconsider the role that IT and CI ought to play in the science and engineering curricula, not just what role they have in the computer science curriculum. Similarly, as the quantity of digital data collections expands opportunities to access information that in the past required travel to a library, these collections will in the future help reshape the way scholarship is practiced and the kinds of questions that can be asked and answered.

Connexions and the OpenCourseWare Consortium were created to support mechanisms for creating, hosting, and disseminating open education resources. While their approach (and those of similar projects) may differ, the overarching objectives are to support, encourage, and openly share education and training materials rapidly and freely. These projects, developed by university researchers, seek to leverage the convergence of the maturing Internet infrastructure, evolving document standards such as XML, and open-source software. As an example, Connexions empowers educators, students, and self-learners by revolutionizing the interplay between education and information creation and dissemination. While inspired by local needs, the open educational repository is unique in that it has grown into a global grass-roots movement for developing and sharing educational and scholarly material. The resulting framework enables us to realize the vision of “a teacher for every learner.” Going beyond the development of authoring and dissemination tools, Connexions researchers have developed a distributed post-publication peer-review system for the World Wide Web that enables scalable quality control. As an example, the IEEE Signal Processing Society has committed to begin using this post-publication peer-review system for reviewing Connexions’ signal-processing resources.

As more institutions engage more deeply in CI, some pertinent and often challenging questions arise. For example, institutions need to assess whether and how their students can access the digital libraries they need. Institutions need to assess whether major research projects are integrated with their own infrastructure for managing data. Similarly, institutions need to determine whether local infrastructures interoperate with national and international collections that faculty and students might want to access. Issues of data ownership, access rights, and administrative controls also pertain.

3.5.2 Outreach

In many cases, an investment in CI is also viewed as an investment in economic development. Universities can seek opportunities for businesses and regional communities to benefit from the CI programs it has developed for faculty, staff, and students. This can take many forms, such as leveraging computational assets to support business use of advanced modeling technologies, or educating future workers about computational science. Further, universities can stimulate the regional economy by leveraging their networks to expand access to broadband or to share higher education resources with business.

A number of universities have begun initiatives focused on industrial applications of supercomputing and overcoming the barriers to widespread adoption of advanced CI. The academic and research communities frequently create codes as part of funded research work and doctoral theses. Many of these codes can be adapted for industrial use, but this rarely happens because of the cost and effort involved in hardening such code to production quality.

The creation of tools and applications focused on specific industrial needs also is echoed in the Department of Defense Computational Research and Engineering Acquisition Tools and Environments (CREATE) program, which targets problems in aircraft, ship, and antenna design.

In addition, these steps are complemented by many universities' sustained investment in CI education and training programs for their clients and the national audience. These universities often survey key users to assess satisfaction with services and to learn what key applications are heavily used or needed.

Companies are sometimes unwilling to use shared resources because of privacy concerns. Implementing and demonstrating virtualization and virtual private network (VPN) technologies ensure confidentiality of a customer's data. Federal agencies should fund research into developing advanced networking and operating system environments that are designed to work in large networked communities. Current models for data protection and system security are rooted in an era of different access paradigms.

Another barrier to small and medium-sized businesses' adopting supercomputing is the perceived cost. This can be alleviated by clearly demonstrating the return on investment possible using a shared service model.

Lack of trained personnel and access to expertise also can be a barrier to adopting advanced CI. The training, outreach, and certificate programs being developed by universities can address this concern. In addition, the increased emphasis on STEM training that is emerging at the state and federal levels will help.

Finally, as industrial programs grow, the demands on university resources must be balanced against access for academic users. Projected growth in computing capacity and the possible revenues from industrial use might address this concern.

Academia can help industry with its large-scale computing needs and help improve the innovation cycle, thus making U.S. industry more competitive. Many small-to-medium-sized companies, and even larger ones, do not always have enough in-house expertise and resources for large-scale computations; as a result, they have not been able to use simulation and analysis tools with the frequency needed to innovate faster and become more competitive on a global scale. The academic community can partner more frequently with industry if CI resources are structured as a platform around which continuous interaction becomes easily possible.

Acknowledgments

The Organizing Committee gratefully acknowledges the support of both EDUCAUSE and the Coalition for Academic Scientific Computation. We would also like to thank the Indiana University IT organization for providing logistical and technical support during the workshop and the Ohio Supercomputer Center for hosting the workshop wiki. The committee would also like to thank Amy Apon for her careful final reading and suggestions for this report.

Endnotes

1. Thomas J. Hacker and Bradley C. Wheeler, "Making Research Cyberinfrastructure a Strategic Choice," *EDUCAUSE Quarterly*, vol. 30, no. 1 (2007), pp. 21–29, http://connect.educause.edu/Library/EDUCAUSE+Quarterly/MakingResearchCyberinfras/40013.
2. John H. Marburger, E. F. Kvamme, George Scalise, and Daniel A. Reed, "Leadership Under Challenge: Information Technology R&D in a Competitive World," 2007 report of the President's Council of Advisors on Science and Technology.

References and Online Resources

- Scott Lathrop and Thomas Murphy, "High-Performance Computing Education," Guest Editors' Introduction, Computing in Science & Engineering, vol. 10, no. 5 (September/October 2008), pp. 9–11.
- Judy Baker, "Introduction to Open Educational Resources," http://cnx.org/content/col10413/latest.
- *Educational Technology*, vol. 47, no. 6 (November/December 2007), http://asianvu.com/bookstoread/etp/Educational_Technology_Nov-Dec2007_issue.pdf.
- Richard G. Baraniuk and C. Sidney Burrus, "Global Warming Toward Open Educational Resources," *Communications of the ACM*, vol. 51, no. 9 (September 2008), http://doi.acm.org/10.1145/1378727.1388950.
- Daniel E. Atkins, John Seely Brown, and Allen L. Hammond, "A Review of the Open Educational Resources (OER) Movement: Achievements, Challenges, and New Opportunities," February 2007, http://learn.creativecommons.org/wp-content/uploads/2008/03/areviewoftheopeneducationalresourcesoermovement_bloglink.pdf.
- The William and Flora Hewlett Foundation, "Open Educational Resources (OER)—Making High Quality Educational Content and Tools Freely Available on the Web," http://www.hewlett.org/Programs/Education/OER/.
- Connexions, http://cnx.org/.
- MIT Open Course Ware, http://ocw.mit.edu.
- *Opening Up Education: The Collective Advancement of Education through Open Technology, Open Content, and Open Knowledge*, Toru Iiyoshi and M. S. Vijay Kumar, Eds. (MIT Press, September 2008), http://mitpress.mit.edu/catalog/item/default.asp?ttype=2&tid=11309. (See also http://commons.carnegiefoundation.org/openingupeducation/opening_up_education.mov.)
- The Cape Town Open Education Declaration, http://capetowndeclaration.org/.
- IEEE-Signal Processing Society Connexions Project, http://www.ieeecnx.org/.
- Barbara Illowsky and Susan Dean, *Collaborative Statistics* (online Connexions course textbook, February 2009), http://cnx.org/content/col10522/latest/.
- Catherine Schmidt-Jones, *Understanding Basic Music Theory* (online Connexions course textbook, September 2008), http://cnx.org/content/col10363/latest/.
- Online Connexions course, instructor Richard Baraniuk, "Signals and Systems," modules by Benjamin Fite, Don Johnson, JP Slavinsky, Justin Romberg, Matthew Hutchinson, Melissa Selik, Michael Haag, Phil Schniter, Ricardo Radaelli-Sanchez, Richard Baraniuk, Roy Ha, Steven Cox, and Thanos Antoulas, http://cnx.org/content/col10064/latest/; Spanish translation by Erika Jackson and Fara Meza, "Señales y Sistemas," http://cnx.org/content/col10373/latest/.
- Harvey Blustain, with Sandra Braman, Richard N. Katz, and Gail Salaway, *IT Engagement in Research: A Baseline Study* (Research Study, Vol. 5) (Boulder, CO: EDUCAUSE Center for Applied Research, 2006), available from http://www.educause.edu/ecar.
- EDUCAUSE Cyberinfrastructure Resources, http://connect.educause.edu/Library/Abstract/CyberinfrastructureResour/48070.

Appendix: Workshop Materials

Workshop Announcement

The CASC and the EDUCAUSE CCI-WG have agreed to co-sponsor a 2-day workshop on the topic:

"Developing a Coherent Cyberinfrastructure from Local Campus to National Facilities: Challenges and Strategies"

Both CASC and the EDUCAUSE Campus Cyberinfrastructure Working Group believe there has been insufficient planning directed toward implementing a seamless cyberinfrastructure among individual principal investigators, central campus support organizations, and the national facilities. This is an opportune time for us to come together, both to discuss these CI challenges and to draft a strategic document to address these issues.

This relatively small working group will be charged with participating in the deep discussions and the drafting of the initial document in the spirit of contributing to the larger cyberinfrastructure community. The workshop format will be an in-depth 2-day set of working sessions rather than a conference-style meeting. The committee recognizes that due to summer scheduling conflicts, as well as financial and logistical constraints, it is not possible to accommodate all key leaders from across the full campus and national cyberinfrastructure community at this July meeting.

Although these constraints necessitate that the workshop size be capped at approximately 40 participants, the long-term success of the workshop will depend, in part, on the ongoing participation of the broader community in the program of activities that emerge from the Indianapolis meeting. The draft document from the workshop will be distributed to both the full CASC and EDUCAUSE CCI memberships with ample time for comments and thorough discussions. The members of the workshop organizing committee plan to attend both CASC and EDUCAUSE CCI meetings for open discussion and comment on the content of the workshop draft document.

Workshop Abstract

CASC and EDUCAUSE Campus Cyberinfrastructure Working Group will hold a workshop to discuss the emerging national requirements for a pervasive, coherent, tiered cyberinfrastructure.

A number of federal agencies have made, or are making substantial investments in key components of the national cyberinfrastructure, and cyberinfrastructure investments are also underway on several campuses. Campuses themselves may have anywhere from minimal cyberinfrastructure capabilities to large multi-Tflops computational systems. Leveraging these investments within the campus and understanding how campuses can integrate with federally-funded infrastructure including larger systems (500 Tflops to ~1 Pflops or more) requires a coordinated effort at multiple levels. Identifying possible options and implementations in order to build a coordinated cyberinfrastructure will both benefit individual researchers and their universities as well as position the U.S. to lead in interdisciplinary scientific discoveries, accelerate innovation, and drive economic development.

The workshop goal will be a working document to the memberships of both organizations with specific suggestions and recommendations that (1) lay out the basic arguments for a pervasive national cyberinfrastructure strategy, (2) include short-term and longer-term recommendations and actions that will enable CI implementation on the university campus, (3) promote funding agency, foundation, and university CI coordination, and (4) develop a draft set of building-block suggestions and recommendations for enhancing university and funding agency/foundation CI coordination and implementation.

There will be 3 workshop breakout groups that will be asked to articulate both a near-term and long-term coherent national strategy for universities that can be applied across the spectrum from the individual user through the national level.

1. Computational systems

This group will address high-performance and high-throughput computing, networks/ communications, visualization, advanced instrumentation interconnected to cyberinfrastructure, and other similar systems.

2. Information management

This group will address data creation, storage, handling, retrieval, distribution interpretation, security, policies on research data, long-term preservation, metadata, etc. including partnerships and opportunities with libraries and repositories and interfaces with funding agencies (grants and contractual). Also included in this breakout are identity management, security, authorization, and authentication.

3. Human/social aspects of cyberinfrastructure

This group will address:

- Campus communities, including outreach to nontraditional computing groups
- Education and training
- Education of professionals who develop, deploy, and support current and emerging CI
- Educational programs to make CI accessible to faculty/researchers, graduate students, and especially undergraduate students
- CI enabled learning—professional development for faculty and teachers so that they can include that in their classroom
- CI partnership strategies for faculty projects using CI
- Virtual organizations

Workshop Participants

Name	Institution	Position
Adams, George	Purdue University	Associate Director for Programs, Network for Computational Nanotechnology
Agarwala, Vijay**	Penn State University	Senior Director, Research Computing and Cyberinfrastructure
Ahalt, Stan**	Ohio Supercomputer Center	Executive Director
Almes, Guy**	Texas A&M University	Director, Academy for Advanced Telecommunications & Learning Technologies
Apon, Amy	University of Arkansas	Director, Arkansas High Performance Computing Center
Bedford, Virginia	University of Alaska	Associate Director, Arctic Region Supercomputing Center
Boisseau, Jay	Texas Advanced Computing Center	Director
Bottum, Jim	Clemson University	Vice Provost and CIO
Butler, Randy	NCSA	Co-Director, CyberSecurity Directorate
Connolly, John	University of Kentucky	Director, Center for Computational Sciences
Dougherty, Maureen	University of Southern California	Director, HPCC
Dreher, Patrick*	Renaissance Computing Institute	Director, Advanced Computing Infrastructure and Systems and Adjunct Professor, North Carolina State University
Facelli, Julio	University of Utah	Director, Center for High Performance Computing
Fratkin, Susan	Fratkin Associates	Washington Liaison, CASC
Hauser, Thomas**	Utah State University	Director, Center for High Performance Computing
Hillegas, Curtis	Princeton University	Director, TIGRESS HPC Center and Computational Science and Engineering
Hobby, Russ	UC Davis	Cyberinfrastructure Program Architect
Jackson, Sally	University of Illinois	CIO
Jelinkova, Klara	Duke University	Senior Director, Systems Group
Katz, Daniel	Louisiana State University	Director for Cyberinfrastructure Development, Center for Computation and Technology
Kerian, Dorette	University of North Dakota	Interim CIO and Director of ITSS
Lassner, David	University of Hawaii	Vice President for Information Technology and CIO
Lifka, David	Cornell University	Director, Center for Advanced Computing
Livny, Miron	University of Wisconsin	Professor of Computer Science and Biochemistry
Luker, Mark	EDUCAUSE	Vice President
Lynch, Cliff	Coalition for Networked Information	Executive Director
McMullen, D. F. (Rick)	University of Kansas	Director, Research Computing

Moore, Reagan	San Diego Supercomputer Center	Director, Data and Knowledge Systems
Moore, Richard	San Diego Supercomputer Center	Division Director Production Systems
Moses, David	Pittsburgh Supercomputing Center	Executive Director
Muzio, Paul	CUNY-College of Staten Island	Director, CUNY High Performance Computing Facility
Neeman, Henry	University of Oklahoma	Director, OU Supercomputing Center for Education & Research
Odegard, Jan E.**	Rice University	Executive Director, Ken Kennedy Institute for Information Technology
Pearce, Michael	University of South Florida	Chief Technology Officer
Pepin, James**	Clemson University	Chief Technology Officer
Punch, Bill	Michigan State University	Director, High Performance Computing Center
Schaffer, Henry	North Carolina State University	Professor Emeritus, Genetics and Biomathematics
Sern, Garret	EDUCAUSE	Government Relations Officer & Staff Liaison
Shephard, Mark	Rensselaer Polytechnic Institute	Director, Scientific Computation Research Center
Smith, Philip	Texas Tech University	Senior Director, High Performance Computing Center
Stalzer, Mark	Caltech	Executive Director, Center for Advanced Computing Research
Stewart, Craig**	Indiana University	Associate Dean, Research Technologies
Waggener, Shelton	UC Berkeley	Associate Vice Chancellor & CIO
Wallman, Marc	North Dakota State University	Director, IT Infrastructure Services

*Organizing Committee Chair

**Organizing Committee Member