22 September 2006

Summary Report

NSF EPSCoR Cyberinfrastructure Workshop

Nashville, Tennessee
May 10-12, 2006

Table of Contents

Table of Contents 2
Acknowledgements 3
Executive Summary 4
1. Introduction 6
  1.1 Process 6
2. The Concept of Cyberinfrastructure 7
3. Context for CI planning 9
  3.1 Assumptions of proposed CI planning model 9
  3.2 Occasions for CI planning 10
4. Elements of a model for statewide/regional CI planning and implementation 11
  4.1 Recruit research leadership 11
  4.2 Establish common goals and assumptions 13
  4.3 Develop Portfolio of Application Drivers 13
  4.4 Survey Physical Resources 14
  4.5 Organize good internal communications 15
  4.6 Identify obstacles to successful statewide CI 15
5. Illustrations of State CI Planning and Implementation 16
  5.1 Northern Tier Network 16
  5.2 Tennessee: The OneTenn plan 17
  5.3 Kentucky: Center for Computational Sciences 19
  5.4 Texas: Texas Internet Grid for Research and Education (TIGRE) 20
6. Conclusion 22
7. References 23
8. APPENDICES 24
  A-1 WORKSHOP AGENDA 24
  A- 2 Workshop attendees 26
  A – 3 List of Abbreviations and Acronyms 32

Acknowledgements

The Tennessee EPSCoR Workshop Organizing Committee consisting of Dr. Micah Beck, Dr. Terry Moore, Dr. David Hercules, Dr. Arlene Garrison, Dr. Peter Bridson, Dr. Greg Sedrick and Dr. George Garrison express their deep appreciation to all of the individuals that contributed to the planning and successful implementation of the Cyberinfrastructure Workshop.

First, we express our sincere thanks to Dr. Sherry Farwell and Dr. Julio Lopez-Ferrao of the NSF EPSCoR Office for encouraging and supporting this endeavor and to Dr. Miriam Heller of the NSF CI Office for valuable guidance in developing the workshop program.

The Steering Committee, Chaired by Dr. Beck, consisted of Dr. John Connolly, KY; Dr. Bonnie Neas, ND; Mr. Michael Grobe, Kansas; Dr. Doug Hurley, Tennessee; and Mr. Eric Cromwell, TN Director of Technology. The primary responsibility of the Steering Committee was to develop the details of the Workshop, including defining the general and breakout discussion areas and identify and invite the experts that provide the overview and experience to guide the discussions.

We were able to attract an outstanding group of experts to serve as speakers, panel members and discussion leaders, as follows:

Dr. Joe St Sauver, University of Oregon
Dr. Cherri Pancake, Oregon State University
Dr. John Cobb, Oak Ridge National Laboratory
Dr. Frank Williams, Director of Artic Region Supercomputing Center
Dr. Scot Lathrop, TeraGrid, University of Chicago/Argonne National Lab
Dr. Mary Fran Yafchak, SURA IT Program Coordinator
Mr. Jeffrey Scot Averbeck, President/CEO of SMARTech Corporation
Dr. William H Sanders, Director, Blacksburg Electronic Village
Dr. Miriam Heller, NSF Office of Cyberinfrastructure
Dr. Peter McCartney, NSF
Dr. Doug Hurley, University of Memphis, TN
Dr. Alan Sill, Texas Tech University
Dr. Bonnie Neas, North Dakota State University
Dr. Mel Ciment

We are indebted to each of these individuals for the significant contributions they made to the content and, therefore, the success of the workshop.

Finally, but certainly not least, we thank Becky Stines, Betty Bright and Penny Morris for providing administrative support during the planning and conduct of the workshop.

Executive Summary

Both inside and outside the scientific community, there is a broad and growing consensus that strategic planning for basic, shared computing and communication resources, now referred to collectively as "cyberinfrastructure" (CI), is critically important to the future success of organizations and individuals at nearly every level of society. Yet despite this agreement, planning in the United States for CI at the level of state government is, for the most part, only just beginning. Many, if not most states, especially in the EPSCoR community, are only now beginning to understand what advanced CI is and why developing a vision for it in their state is critical to their future in the era of globalization. To help address this situation, the ***NSF EPSCoR National Workshop on Cyberinfrastructure*** brought together a wide range of researchers, educators, and administrators from the EPSCoR states with a select group of leaders from the national CI community in order to develop a reasonable and effective model of state CI planning.

The first stage of the workshop included a series of talks and panel discussions focused on the nature of CI, the challenges confronting application communities who want to provision and use CI, and the benefits and costs of investing in advanced CI. During the second phase, the attendees divided into discussion groups to focus on the planning process. They considered the requirements for developing a successful state CI plan, the costs and benefits of CI deployment, and the technical and organizational requirements for successful implementation.

A significant part of the challenge of developing an adequate planning model for CI lies in formulating a clear conception of what CI is. Such a conception is needed in order help planners raise and frame the issues that are likely to arise in the design, deployment, and use of CI for research and education, which is the primary context for the EPSCoR community. Out of the various conceptions of CI presented, either explicitly or implicitly, a formulation that seems to capture most of the salient features for planning purposes can be given as follows:

> *Cyberinfrastructure is the common and persistent base of computing and communication resources — hardware, software, and people — shared by a community in order to facilitate the use of digitized information for the purposes of collaboration in highly distributed environments.*

Even with such a definition of the subject matter in hand, however, any planning process inevitably makes certain assumptions about the common goals and circumstances of the parties involved and about the expected circumstances of the plan's proposed implemented. In these regards, the academic research and higher education communities within states are especially well positioned for leadership. Common commitment to the goals of research and education, combined with the shared context provided by a state's geography, history, and government institutions, have historically enabled higher education to take the lead in developing new communication infrastructure (e.g., state networks based on Internet technology) to enable cooperation and collaboration in a wide range of areas. In the emerging era of cyberinfrastructure, a range of new opportunities and conditions, such as competition for major federal funding, participation in important national or international research efforts, and leadership in national CI efforts, can arise to motivate these communities to drive the state planning forward.

But from the point of view of CI planning, academic research and education communities in states across the country possess another and even more critical asset: more than two decades of experience in major IT infrastructure efforts, such as computing centers and state and regional networks. The lessons learned from the process of planning, developing and building such facilities, which have now been absorbed into the broader concept of CI, can be applied to advantage in the new challenge of creating a richer, more powerful, and more comprehensive CI at the state or regional level. In the workshop, this experience, combined with the ideas and information from presentations and panel discussions, brought out a number of crucial factors for successful CI planning, including the following:

- *Research leadership* – Leadership from scientists and engineers engaged in advanced research has historically been a critical component of earlier, proto-CI efforts, such as the development of the

Internet in the 1970s and 80s. Their leading edge research applications have proven to be an excellent way to prepare for the requirements of every day applications tomorrow. Successful plans have depended on the ability of research leaders to recruit high profile applications, leverage research funding to bootstrap deployment and expand educational impact, attract industry participation, and stimulate the development of essential human resources.

- *Common goals and assumptions* – While there are natural similarities among a state's network of colleges and universities, the value for the plan of making these shared, general goals explicit at the outset becomes evident as other potential constituencies (e.g., state government, health care and medicine) become engaged. The more generic the goals are, the easier it is to garner broad support.
- *Portfolio of application drivers* – Compelling applications are essential to successful CI plans. They help define the infrastructure requirements in a definite way, demonstrate new CI capabilities, and suggest clear examples of broader educational and economic impact.
- *Survey of current CI resources* – Providing an accounting of current physical CI resources within a state can be a useful part of the planning process. This is more easily accomplished for networking resources, which are typically shared across administrative domains, than it is for the processing and storage resources, which are often not centrally owned or managed.
- *Good internal communications* – Statewide CI planning processes can be expected to involve cross organizational communications outside the normal patterns that organizational hierarchy normally permits. Success may require creative use of new communication tools.
- *Common obstacles to state CI* – Typical obstacles to successful CI planning and implementation include a scarcity of human resources for research and technical support, establishing a stable and sustainable way to fund growing CI efforts over time, and building appropriate conditions for cooperation among the IT groups involved in CI, both within organizations and between them.

The workshop offered several illustrations of where different states and regions are in their CI planning and implementation. In technological terms, these plans range from more high performance networking efforts at the regional level (Northern Tier states), a combination of advanced networking and distributed storage (Tennessee), a focus on high performance computing (Kentucky), and a full blown computational grid deployed state wide (Texas).

The workshop concluded with widespread agreement that advanced cyberinfrastructure will provide the essential foundation for future progress in science and engineering generally. But there were also anxieties about whether all states or even most states would be able to remain competitive in this area. Beyond the normal concern about scarcity of available funding from state governments, several notable problems currently inhibit states from taking next steps in CI planning. These problems include lack of established standards for interoperability in key technologies, a perceived lack of entry level approaches for individual campuses to engage national CI projects, such as the TeraGrid, and a dearth of well known and widely adopted CI alternatives for aggregating *local* (i.e. campus) CI resources in order to deliver benefits to *local* research and education communities. The National Science Foundation is better positioned than other agencies at the federal level to formulate and invest in the strategic efforts necessary to overcome these challenges.

1. Introduction

The ongoing revolution in information technology, which continues to transform our society in all its dimensions, is expected to continue for the foreseeable future. In an era of increasingly global competition, public and private institutions at every level, especially in the areas of research and education, are focusing their attention on the investments essential to future success in this rapidly changing environment. Today, the term *cyberinfrastructure* (*CI*) has come to represent the constellation of shared and persistent IT resources required to effectively support the collective activity of large communities and organization. Consequently, strategic planning for CI has become a vitally important activity in which organizations of many types are required to engage.

Although there are notable national CI projects already underway (e.g., TeraGrid, OptiPuter, Open Science Grid), planning for CI at the state and regional level is, for the most part, still nascent. Projects like TIGRE in Texas and SURAgrid in the southeast region are vanguard efforts. Some states, especially in the EPSCoR group, have only recently completed their high performance networking capability and many, if not most, are only now beginning to understand what advanced CI is and what it might mean, in terms of benefits and cost, for their state or region. Against this background, ***NSF EPSCoR National Workshop on Cyberinfrastructure*** brought together a wide range of researchers, educators, and administrators from the EPSCoR states with a select group of leaders from the national CI community in order to start to develop a rational and well informed model of CI planning for use at the state or regional level.

1.1 Process

To achieve this goal, the workshop followed a two phase process. Phase one was intended to clarify and broaden the understanding of the attendees concerning the nature of CI and its applications. Following this information sharing phase, the goal of phase two was to elicit the attendees' ideas and experiences regarding requirements for successful state CI plans.

Phase one of the workshop, the information sharing part of the process, consisted of a series of talks and panel discussions, divided into three major sessions. In each session, national and state leaders in CI described and analyzed, from a variety of different perspectives, several major CI projects, applications, and initiatives. The CI Workshop agenda (Appendix 7.1) and list of attendees (Appendix 7.2) are attached for reference.

The first session focused on the nature of cyberinfrastructure. Following an overview of the development and structure of the NSF's CI program(s), speakers from several organizations and projects discussed the construction and promulgation of cyberinfrastructure regionally (e.g., SuraGrid), nationally and internationally (e.g., the TeraGrid, Internet 2, National Lambda Rail). Talks covered not only basic concepts, architectures, strategies (especially in the area of optical networking), and deployed hardware and software resources, but also illustrated the powerful uses of this CI by describing a variety of high impact scientific applications and collaborations enabled, such as the Network for Earthquake Engineering Simulation (NEES), Linked Environments for Atmospheric Discovery (LEAD), and major collaborations revolving around the Large Hadron Collider in CERN and the Spallation Neutron Sources at Oak Ridge National Laboratory. The session concluded with a panel discussion of return on investment that such efforts were expected to deliver.

The second session examined application and community enabling challenges of advanced CI, offering further illustrations of domain specific collaborations, such as the National Ecological Observatory Network (NEON), the Virtual Tsunami Center, and more community based projects such as the Blacksburg Electronic Village. It concluded with a panel discussion of the similarities and differences that EPSCoR states might experience on this front.

The third session focused on how state and regional cyberinfrastructure plans can benefit states more broadly, addressing the value of the investment in CI to communities beyond the major research universities. State CI plans from Texas, Kentucky, and Tennessee were examined, along with a regional plan from the Northern Tier Networking Consortium, which includes Wisconsin, Minnesota, Iowa, North Dakota, South Dakota, Montana, Idaho and Washington.

During the second phase, the attendees were divided into four discussion groups, with each group focusing on a different aspect of the CI planning process. These breakout sessions, each of which had a facilitator, a scribe, and a set of guiding questions, covered the following topics:

- Requirements for the development of a successful state CI plan
- Realistic benefits and costs of CI deployment in a state
- Technical requirements for successful implementation of a state CI plan
- Organizational and political requirements for successful state CI plan implementation.

At the end of the first day, the facilitators and scribes for each group presented an overview, for question and discussion, of the results of their group's work. During the workshop's final plenary session, on the end of the second day, the detailed notes from all the breakout sessions were collected and given an initial integration so that they could be presented to and discussed by the attendees.

2. The Concept of Cyberinfrastructure

A basic goal of the workshop was to further promulgate the idea of cyberinfrastructure and increase the understanding of it within the EPSCoR community. This is not as simple a task as it might first appear. The word 'cyberinfrastructure' is of relatively recent origin, and the concept it represents is still in the process of being clarified through discussions taking place in the S&E community. In the absence of an adequate and widely accepted definition of CI, the idea is usually communicated using one or more of three different approaches: 1) An historical approach that summarizes the origins of CI in the past two decades of rapid growth in high performance networking and computing; 2) A anecdotal approach that describes its use in major research projects where CI is most prominently displayed; and 3) a basket of technologies, in which the major technologies that form its constituent elements are listed.

1) The term "cyberinfrastructure" was coined as part of a major program initiative started by the NSF in 2001 and carried out, in its first phase, by a blue ribbon commission [1]. A main goal of the initiative was to set the stage for future investments in national research infrastructure by consolidating the lessons of the preceding decades, a period during which modern science and engineering were utterly transformed by the digital revolution. Because of that transformation, every aspect of the scientific enterprise is now dominated by the flow of digital data and the technology necessary to manipulate it, from the instruments and simulations that supply its inputs, to the analytical tools and display devices necessary to digest its results. During the 1980's and 90's, the NSF made two major sets of investments in national facilities intended to enable scientists and engineers across a range of fields to make rapid progress in this new world: First, there were the NSF Supercomputer Centers, and the Partnerships for Advanced Computational Infrastructure that succeeded them; and second, there was the Internet, in the form of NSF Net, and the high performance research networks (such as Internet2) that followed. In large measure, the new concept of CI represents an effort to unify and build on the successes and experiences of those historic initiatives, including their broader economic and cultural impact.
2) Over the past decade, the catalog of major S&E projects in which the idea of CI has played a leading role, either explicitly or in latent form, has been impressive. Drawing examples from the current catalog, there are several projects explicitly dedicated to building up national CI as such, most notably the NSF's TeraGrid project and the DOE's Open Science Grid (OSG). Both are now operational and delivering services to hundreds of users at dozens of sites through "application gateways" in a wide range of fields, including Physics, Astronomy, Climate, Bioinformatics, Nanoscience, Neutron

Science, Geoscience, and many others. Notable efforts are also underway at the regional level, such as SuraGrid, and at the state level, such as the Texas Internet Grid for Research and Education (TIGRE). For federally funded projects, it is now expected that any major collaboration in virtually any field will have some plan for providing the CI necessary to enable the collaboration, either by leveraging current shared CI facilities or by proposing to build out and provision additional resources.

3) Building on NSF's initial CI vision and this wealth of community experience, a general, if somewhat open-ended, description of the technological elements of CI has developed. For example, this characterization from the recent PITAC report, *Computational Science: Ensuring America's Competitiveness* [2] is typical:

> *In the United States, NSF has adopted the term "cyberinfrastructure" to describe the complex, integrated IT tapestry of the future whose elements will include seamless networking, system software, and middleware providing the generic capabilities and specific tools for data, information, and knowledge management, processing, and transport. The NSF-commissioned report,* ***Revolutionizing Science and Engineering through Cyberinfrastructure****, characterizes cyberinfrastructure as that portion of cyberspace where scientists can "build new types of scientific and engineering knowledge environments and organizations and . . . pursue research in new ways and with new efficiency."*
>
> *The major components of cyberinfrastructure should include:*
>
> - *High-performance, global-scale networking, whether a hybrid of traditional packet switching or a more advanced model built upon high-bandwidth optical networks*
> - *Middleware enabling greater ease in applications building and implementation, secure communications, and collaborative research*
> - *High-performance computation services, including data, information, and knowledge management*
> - *Observation and measurement services*
> - *Improved interfaces and visualization services*

However, for CI planning at the state level, where resources tend to be scarce and goals are naturally more focused on education and broader impacts, a less expansive and *ad hoc* definition would seem to be called for. For instance, some accounts of the concept of CI include large instruments and sensors, like telescopes and colliders, on the list of basic CI elements. No doubt such facilities are a main source for the rising deluge of data that is beginning to swamp many fields. As such, they are powerful drivers of CI development. But the conditions under which it would make sense for a state government, or a state's academic S&E community, to expend its limited CI resources on such an investment would have to be extraordinary. The value of such instruments is too specific to a particular field, or small set of fields, and the communities that benefit are mostly national or international, not confined within any particular state. If states are to do adequate CI planning, their deliberations need to be informed by a concept of CI that is somewhat tighter and better suited to state goals and operating conditions.

The workshop provided adequate material for such an account, or for at least a plausible candidate. It can be stated as follows:

> *Cyberinfrastructure is the common and persistent base of computing and communication resources — hardware, software, and people — shared by a community in order to facilitate the use of digitized information for the purposes of collaboration in highly distributed environments.*

This account is intended to define "cyberinfrastructure" in a way that is not only easy to state, but also easy to explicate. While it could be made even more general by broadening the purpose of CI beyond the research community, it seems to contain all the key parts. Working through these parts brings out the following relevant points:

- *"…common and persistent base …"* – Cyberinfrastructure shares the general properties possessed by other types of infrastructure. Typically we think of infrastructure as an integrated set of resources or capabilities that address or facilitate the satisfaction of common needs (e.g. for mobility, power, water, communication, etc.). Because they often require major physical modifications to the environment, such needs are usually addressed via capital investments intended to implement them as persistently as possible.
- *"…computing and communication resources – hardware, software, people …"* – Processing power, storage, and network bandwidth are the fundamental physical resources on which all CI is based. The equipment and supporting facilities— supercomputers, storage clusters, machine rooms, fiber-optic cable, communication switches, etc. — represent significant capital investments. For the purposes of CI planning, however, it is important to remember that these hardware components are useless without essential software components (e.g., basic protocols, programming interfaces, libraries, etc.), as well as necessary human resources to keep this systems operating continuously and reliably. Software is the essential glue that unifies and holds CI together, even as the hardware changes. Since well designed protocols and libraries frequently outlast the hardware they initially run on, they might also be viewed as capital investments.
- *"… shared by a community…"* – The raw capacity of basic computational and communication resources that CI uses continues to increase at exponential rates, but the resource requirements of leading edge S&E research (e.g., the size of the simulations to be run, the amount of data to be analyzed) continues to outstrip what research teams, and even institutions, can afford to own for their private use. At all levels, sharing these basic resources within larger communities is the only practical way to make sufficient quantities affordably available to the vast majority of researchers in any given area. The need for resource sharing is a primary motivator for advanced CI.
- *"… facilitate the use of digitized information…"* – The concept of CI is a product of the digital revolution. By providing a universal way of representing all kinds of information – bits – digitization makes it possible to move, store, and process all kinds of information on a single type of infrastructure. Thus, CI is a medium in which information on media of (almost) all other types can be dissolved, transferred, preserved, and transformed for an enormous range of possible uses.
- *"…collaboration in highly distributed environments."* – Science today is becoming more and more collaborative and interdisciplinary. As such collaborations grow, all of their chief components – the data, the people and the research resources – tend to be increasingly distributed across geographic and organizational boundaries. A guiding purpose of CI is to provide the technology substrate necessary to make it easy for distributed teams to share data, human expertise, and other S&E resources as they would if they were all collocated.

All of these aspects of CI have been on prominent display during the growth of the Internet, which can be counted as a wildly successful prototype for all future CI. By offering a packet-switched communication network based on a generic common service for moving bits, the Internet made it possible for research and education communities to deploy massive amounts of network capacity in support of their own goals, and yet enable resource sharing and dramatically expand collaboration across geographic and administrative borders. An analogous case could also be made relative to the NSF supercomputing and PACI programs of the late 80's and 90's.

3. Context for CI planning

3.1 Assumptions of proposed CI planning model

Since the workshop was organized by and for the NSF EPSCoR community, there were some implicit assumptions about the focus and scope of the model of CI planning under consideration. Specifically, the desired model addresses the goals and general circumstances of the academic research and education institutions in a state or region, i.e., its universities, colleges and related organizations. Of course CI planning is essential in many kinds of institutions, including government, the military and private industry. With different goals and circumstances, however, CI planning and implementation models will tend to vary. Over the past three decades, the research and education community has consistently been at the forefront in the development and use of new CI. The workshop brought out important factors in the planning process that suggest that this trend is likely to continue, so that leadership from higher education is an asset that states should leverage in this area.

Academic research and education communities within a state have certain advantages when it comes to leading statewide CI planning efforts. First, while a state's universities and colleges are normally spread out across the state geographically, their goals, assumptions, and operating procedures are similar enough to provide a common frame of reference and orientation for planning purposes. Second, although institutions of higher education within a state usually compete at some level, they also have and recognize many reasons to collaborate in both research and education; better CI within the state can often increase the number and value of opportunities for doing so. Third, the coincidence and natural interaction between research and education at colleges and universities means that thinking about advanced CI for research and for education will tend to go hand in hand. Finally, leaders in a state's higher education community are in a natural position to help develop a plan to maximize the value of CI for the states P-12 community as well, which is a high priority for every state government.

But though the goals of scientific research and higher education formed the background of the workshop's discussion of CI planning, it was clear that a successful model will have to engage with all a state's major stakeholders. Developing and articulating such broader opportunities and impacts for other commercial, governmental, and civic interests within a state is an important part of the planning process.

3.2 Occasions for CI planning

Allusions to conditions that tend to trigger or encourage state CI planning efforts were sprinkled through the workshop's presentations, discussions, and breakout sessions. These conditions help to occasion productive CI planning because, when they occur, especially in combination with one another, they tend to mobilize the various factors that are important to creating and implementing a good state CI plan, such as quality research leadership and strong advocacy from government or industry CI partners. The most commonly mentioned trigger conditions include the following:

- *Establishment of common standards to support interoperability* – As the explosive growth of network infrastructure over the past two decades has demonstrated, the broad acceptance of common and reasonably generic standards (such as IP) for the shared use of highly desirable CI resources (such as network bandwidth) can stimulate immense amounts of investment in the infrastructure necessary to make those resources widely available. Such standards are necessary in order to give investors and stakeholders confidence that the infrastructure to be deployed will support interoperability among an ever expanding pool of applications over time, even as innovation changes the underlying technologies. The absence of such standards inhibits progress by prompting fears that the value of the infrastructure investment will be too soon degraded through obsolescence or balkanization. By contrast, when such standards are in place, the consensus necessary for planning and investment can coalesce much more quickly. This fact suggests that the current relative lack of such standards for storage and processor resources is a significant constraint on the growth of advanced CI today.
- *Support for coordinated planning by the state's academic community* – Even small amounts of funding to facilitate coordinated planning within a state's academic research and education community can help catalyze state CI planning. EPSCoR funding has had this effect within several EPSCoR states, mostly in the area of networking. More recently, an EPSCoR planning grant for Tennessee produced a plan that embraced a complete contemporary CI vision for the state. From the nature of the case, whenever a state's research and education community come together to discuss collaboration on common needs and interests, planning for improved statewide CI will tend to be high on the agenda.

- *Competition for major federal funding* – Since CI is so crucial to leading edge research today, the decision to locate a major research facility, such as a big instrument or a supercomputer, in a given place can be directly influenced by the quality of the CI that surrounds it. When a university or laboratory is in a reasonable position to compete for such a project, state government may be encouraged to invest to help it. Competition with other states in the region may also be motivational in this regard, as was shown by the effect that Louisiana's $40M investment in the Louisiana Optical Network Initiatives (LONI) in 2003 had on its neighbors.
- *Participation in national and international research efforts* – Participation in national or international research efforts can place demands on CI that requires local leaders to confront current limitations. One well known current example of such a case is the desire by physicists at many institutions to become tier 2 sites for Large Hadron Collider projects, such as ATLAS and CMS. Qualifying for such participation requires that certain non-trivial levels of networking, storage, and computational power be available at a proposed tier 2 site.
- *Direct funding for participation in national CI efforts* – Programs such as NSF's *Connections to the Internet* and *Extensions to the TeraGrid* have, in numerous cases, been effective in eliciting state support and catalyzing statewide planning for CI. In the case of the latter, the competition to join the TeraGrid energized statewide collaborations in Texas and in Indiana that have vaulted them to national leadership in state CI.
- *Surplus state funding* – Obviously the availability, or the potential availability, of state government funding to support statewide CI can be highly motivational. But since taking advantage of such opportunity in a timely manner requires that some kind of plan, if only a vague one, needs to be ready, this factor often comes into play only in combination with one or more of the others.

4. Elements of a model for statewide/regional CI planning and implementation

Given the concept of CI and the context for CI planning outlined above, academic research and education communities in states across the country possess another and even more critical asset: more than two decades of experience in broad IT infrastructure efforts, such as computing centers and state and regional networks. The lessons learned from the process of planning, developing and building such facilities can be applied to advantage in the new challenge of statewide CI planning. Workshop participants combined this with ideas and information from presentations and panel discussions to bring out the crucial factors for successful CI planning outlined below.

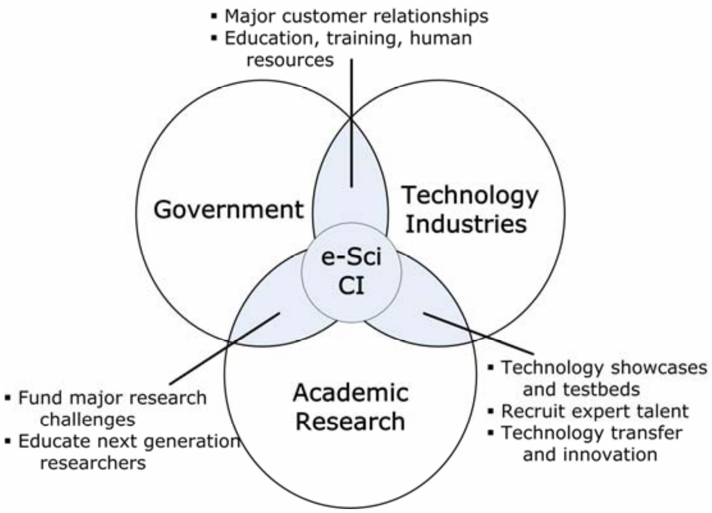

Figure 1: Academic research community builds on relationships with technology industries and the federal government to drive the development of advanced CI. Adapted from the OneTenn report [3].

4.1 Recruit research leadership

A key component of successful CI planning is research leadership, which needs to be established as early as possible in the planning process. The belief in the importance of this factor is largely based on experience. Historically, the use of new and advanced information technologies, such as the Internet, high performance computing, and the World Wide Web, received its earliest development within the research and education community. One leading reason for this is that the unbounded nature of scientific inquiry tends to push the working requirements of the research community well beyond the capacity of whatever technology happens to be available at a given time. For example, the size of the simulations and the data sets that groundbreaking science uses continue to increase exponentially. Today's research problems can require extraordinary amounts of processing power, require the ability to move data thousands of times faster, and store files thousands of times larger than most non-research applications. Planning and building for the needs of leading edge research today have proven to be a good way to prepare for the requirements of every day applications tomorrow. In consequence, researchers studying problems that put extreme demands on CI have proven to be ideally positioned to leverage to maximal effect important synergies between the academic community, government, and private industry (Figure 1). Important impacts of research leadership for the multifaceted CI planning process include the following:

- *Ability to attract meritorious applications* – A key factor in the success of early CI efforts, both in high performance networking and supercomputing, was the existence of a group of meritorious applications that could make good use of the resources that the given CI plan proposed to provide. When researchers head up the CI planning process, it enhances the planning group's ability to recruit other researchers, encouraging them to contribute their knowledge, resources, prestige, and, most importantly, their application requirements to the effort.
- *Research funding and advanced CI are mutually reinforcing* – The presence of advanced CI both enables and is enabled by computational intensive and/or data intensive research. For example, successful participation in national and international research communities often requires access to high-performance networking or large computational clusters. At the same time, research funding provides independent support for the essential human resources, in the form of students and research staff, usually required to make good use of new CI. The availability of such additional support for the personnel outside normal institutional budgets enhances the credibility of CI plans led by top researchers.
- *Educational impact of research driven CI plans* – As members of academic departments, researchers are in an excellent position to make sure that the educational impact that advanced CI can have is well articulated and communicated. Various common kinds of benefits can be expected for each particular state's plan. For example, research funding for student participation in CI intensive research increases the size of the pool of the skilled individuals that new CI plans need. This system helps prepare the next generation of researchers for the next round of CI leadership. Moreover, such pools of talented and well trained students and staff are major attractors for potential commercial partners in major CI projects.
- *Ability to engage industry participation* – Major CI efforts in the service of advanced research, such as SDSC in California, NCSA in Illinois, and MCNC in North Carolina, have proved to be powerful attractors for industry technology partners looking for opportunities to develop and demonstrate their newest technologies in the context of significant government markets. Research led CI plans are well positioned to make effective use of this fact.
- *Development of necessary human resources* – One of the most direct and critical measures of a CI plan is its potential to attract talented people to a state or region. Leading edge research tends to generate the kind of activity and excitement that is magnetic for the most talented people — scientists, engineers, expert research staff, and students. CI plans that can build in a credible way on enabling new scientific discoveries offer the best prospect of yielding this result.

4.2 Establish common goals and assumptions

It is natural to suppose that the chances of producing a state CI plan that is both coherent and mutually beneficial to all the stakeholders are greatly improved if the participants have common and/or compatible goals and assumptions. The similarities that exist between a state's network of colleges and universities put this group in a good position to establish such a common basis for planning. Undoubtedly differences of emphasis and capability will exist, e.g., as regards the balance between the goals of research and scholarship and the goals of education. But experience has shown that, aside from issues of scale, the kinds of CI requirements this community has are generically similar enough to allow for a common vision to be developed. [4]

The value of establishing shared, general goals for the plan at the outset tends to become more evident as the effort is made to reach out for support to other important constituencies in the state. State government, medicine and health care, and commercial and industrial organizations may benefit greatly from the generic infrastructure that a state's academic research and education community requires, but they often have goals and requirements (e.g., in the areas of security and intellectual property protection) that are not universally shared. While more generic infrastructure can often be augmented to accommodate these more specialized needs, integrating them in as essential goals for the overall plan will tend to increase costs, reduce scalability, and make it less coherent. In general, the more generic the goals of the plan are, the easier it will be to garner broad support.

4.3 Develop Portfolio of Application Drivers

A common theme of our workshop presentations and discussions, and a widely accepted idea within the CI community generally, is that compelling applications are the key to building effective CI. A fundamental part of a good state or regional CI plan is a collection of descriptions of leading applications to be deployed within the state or region – what they would do, who would use them, what they would enable, why they need advanced CI, what their broader impact would be, and so on. These motivating applications help translate a plan's descriptions of raw CI power and capacity – gigabits, petabytes, and teraflops – into activities and effects that can be grasped and appreciated by the broader audience to which such plans must inevitably appeal. To appreciate the role such applications play in CI planning, consider some of the main reasons that make an application a good candidate for inclusion in the planning process:

- *Resource intensive applications help define requirements* – Advanced applications in many fields of S&E today push the envelope in terms of CI capabilities, most often because they generate unparalleled amounts of data, require exceptional levels of processing power, involve widely distributed multi-party collaborations, or all of these factors at once. Numerous illustrations of such applications were offered at the workshop. New instruments and facilities (such as the Spallation Neutron Source at Oak Ridge) can pump out a terabyte (=1000 gigabytes) of data per day. Large simulations can occupy supercomputers with thousands of processors for hundreds of hours. And sharing the data generated or using this processing power from a remote site requires networks capable of moving data at hundreds of gigabits per second. The existence of a number of such applications within a state establishes the need for the CI, which the plan proposes to provision and deploy.

  Since history has shown that the requirements for leading edge science today quickly become routine for commonplace applications tomorrow, these research applications have proven to be excellent vehicles for prudent CI planning. Even in today's environment, there are non-scientific applications already emerging that show that resource intensive research applications provide a good model for planning purposes. The use of high definition television for education and the need for regular, fast, remote backup of massive amounts of administrative data for disaster preparedness are two obvious examples.

- *Distributed teaming shows impact of collaboration* – Advanced CI provides the connective tissue that makes possible the kind of highly distributed, interdisciplinary scientific teams that are now characteristic of leading edge research in every field. Such infrastructure is necessary not just for communication, but also for sharing a range of critical resources that are fundamental to such cooperative efforts – storage, processing power, people, data, and software. The patterns of collaboration described in the application scenarios of a state's CI plan can show vividly how the proposed infrastructure will open up a wide range of new opportunity spaces for people to work together across traditional organizational boundaries.

  This new potential for cross-boundary collaboration has both external and internal aspects. Externally, advanced CI can involve the state's research and education community in significant, and often high profile, national and international science and engineering projects. For example, there are now TeraGrid "science gateways" in Nanotechnology, Bioinformatics, Environmental Modeling, and Particle Physics. A state's ability to join the TeraGrid and become not just an isolated user, but also real contributor to one of these gateways can mean joining research groups at the leading edge of their respective fields. Internally, within a state, advanced CI can make it possible to structure innovative research or educational projects in which schools from across the state can jointly participate as full partners. By giving faculty and students at smaller schools access to much higher levels of computing power and storage, for example, they can enable them to start doing data intensive research at a level that would otherwise be unattainable in the near term.
- *Application narratives can show broader educational and economic impact of CI* – Applications that show the impact of next generation CI are not restricted to advanced research applications, but can involve new educational programs and economic development as well. For example, new combinations of high-performance networking, computational resources, and distributed storage can give a state's higher education community a foundation for multimedia production and distribution that far outstrips anything that is typically available today. Colleges and universities can collaboratively develop media-intensive, high-quality course content and then easily distribute that content virtually anywhere, including the technology centers and P-12 schools. Similarly, more pervasive CI can increase the potential to involve commercial and industrial partners within a state or a region to participate in and benefit from (e.g., through federal SBIR programs) the latest scientific and engineering research projects. Application narratives of this kind can help recruit and energize support for the CI plan from a broad group of stakeholders and other constituencies within the state.

It is important to note that there is a close relationship between having good research leadership, as described above, and having the kind of application portfolio that a successful CI plan requires. One of the main reasons for this is that real application drivers are needed to deliver on the expectations that the CI planning process inevitably creates. This becomes apparent if and when the plan is accepted, receives funding, and implementation and deployment begins. At that point, the new cyberinfrastructure has to start being used and start delivering at least some of the benefits and at least some of the impact that the plan described or suggested. The applications described in the plan, along with the commitment and resources of the readership leaders who recruited or provided them, are usually essential for making good on these expectations in a timely manner.

4.4 Survey Physical Resources

It is natural to suppose that a state's CI planning process should begin by surveying the CI resources that the state already possesses. In order to know what we need to build, we should first learn what we already have. But such studies are not simple undertakings, even when they are confined to just the major networking, storage, and processing resources within a state. It is true that in the case of the networking, where the rise of the Internet and the build out state networks has established networks as a statewide resource, taking such a survey can be relatively straightforward process, at least for the major links. Accounting for the storage and processing resources within a state, however, is more problematic.

In the first place, it is typical to think of major storage and processing resources as shared within the organizations that own them, not across those domains. This is true even at the campus level, where different colleges and departments may have their own computing and storage resources that are independent of any other computing or data resources that may exist within the organization. There often is no central point of control keeping track of what is available and to whom. Moreover, there is a question about what to list. Hardware that was state of the art four years ago may be barely worth talking about today. Finally, the willingness of organizations to describe their storage and computing resources can be affected by their concerns about the kind of cross-institutional resource sharing that such statewide CI is intended to produce. These resources are often already committed to serving certain user communities, and administrators and researchers can be reluctant to list them as parts of a possible statewide CI when it is unclear how this would affect those users.

All these points were reflected in the experience of the Tennessee CI planning group, which was presented at the workshop. Their experience shows that the planning process can proceed with a survey of available resources that covers only the broadest outlines.

4.5 Organize good internal communications

A statewide CI planning process confronts the same general problems of cross-organizational collaboration that CI itself is supposed to address on a broader scale. Each of the participants starts from some position in their local administrative environment, with their own particular set of requirements and constraints. Establishing good communication is essential for bridging any gaps that this situation may produce. In EPSCoR states, the existing state committee may be able to provide significant support, especially by way of organizing at least some face to face meetings of the leaders of the planning process. But the quality and quantity of communication required to achieve success in such a challenging undertaking is likely to require the aggressive and creative use of all the tools that modern telecommunications and the Internet now make available.

4.6 Identify obstacles to successful statewide CI

The enthusiasm surrounding the potential benefits of statewide CI should not obscure the major obstacles that often lie in the way of achieving it. Identifying such obstacles is not only critical to getting them addressed, it is also essential to setting appropriate expectations for the different stakeholders. Some of these problems, such as those arising from disparities between administrative environments of the different organizations involved have already been mentioned. But the community is familiar with other difficult and important challenges that CI planners need to take into account, including the following:

- *Scarcity of human resources*: People who work in technical and administrative areas associated with CI are familiar with the fact that the scarcest resource is often people. Human resources, including both research support and technical support (e.g., networking, computing, and storage/database specialists), tend to be both expensive and difficult to recruit, even when the funding is available. This is often true of IT professionals in general, but the problem is exacerbated in the field of CI, where new technology is being deployed and the focus is on supporting new forms of cooperative work. To some degree, the CI community confronts a chicken-and-egg problem: It is hard to develop and keep good people until you have good CI, but it is hard to build up good CI until you have the right complement of people to support the effort.
- *Establishing a sustainable funding mechanism*: Finding a stable way to fund growing CI efforts is clearly a challenge, as it has been historically throughout the current era of IT development. The builders and operators of computer centers and IP networks have a long history with this issue and have considered a wide variety of charge back models, user fees, and other solutions. CI plans cannot be expected to propose solutions to this problem. However, suggestions for possible or plausible strategies for finding such solutions might increase acceptance for the plan.
- *Cooperation with internal and external IT groups*: Development of advanced CI impinges on, and will inevitably require cooperation from, established IT organizations, both inside and outside the

university. But establishing such cooperation can often require significant effort and/or resources. Inside the university, tensions between major research groups and administrative or other traditional IT departments have been a familiar part of academic life. The advent of advanced CI, with its larger vision and its emphasis on cross-institutional collaboration, may revive or exacerbate such stresses. At a statewide level, developing good relationships with telecom providers has often been crucial to the success of research networking; the broader scope of CI may similarly broaden the set of commercial providers that CI will affect. CI plans that take these factors into account in advance will tend to make faster progress.

5. Illustrations of State CI Planning and Implementation

The workshop offered several illustrations of where different states and regions are in their CI planning and implementation. In technological terms, these plans range from more high performance networking efforts at the regional level (Northern Tier states), a combination of advanced networking and distributed storage (Tennessee), a focus on high performance computing (Kentucky), and a full blown computational grid deployed state wide (Texas). Brief overviews of these efforts are presented below.

5.1 Northern Tier Network

In some areas of the country, the challenges of geography are so imposing that high performance networking continues to be the primary focus of CI planning. In particular, the states of the “Northern Tier” — ten states primarily along the Canadian border between Chicago and Seattle — have long been overlooked by our nation’s high performance Research and Development (R&D) networks. This puts the states of the Northern Tier at a significant competitive disadvantage in many fields of R&D in today’s digital age. For example, to support the important work of North Dakota’s Research Corridor, it is essential to connect the fiber-optic telecommunication assets of NDSU’s Centers for Nanoscale Science and Engineering and High Performance Computing to the broader national fiber optic backbone and high performance research networks. The Northern Tier Network project is viewed as imperative if such

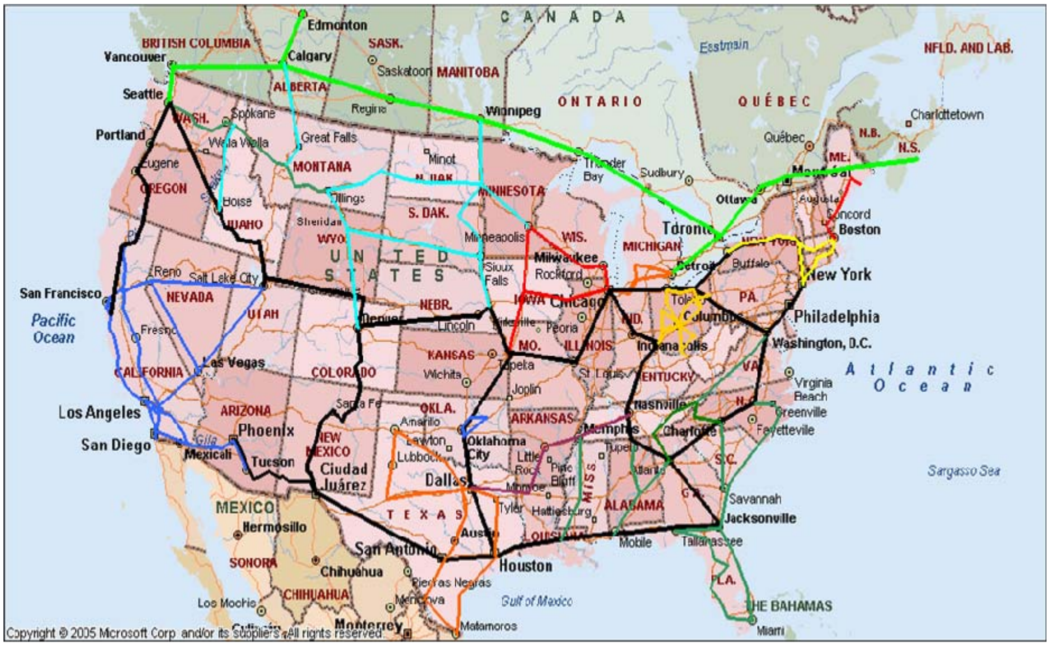

Figure 2: North American research networks with the proposed Northern Tier backbone (light blue).

projects are to be able to thrive, or even remain viable, in the states of the Northern Tier.

The Northern Tier Networking Consortium (NTNC) initiative includes the states of Wisconsin, Minnesota, Iowa, North Dakota, South Dakota, Montana, Idaho and Washington - large rural states with relatively low populations. Historical records and today's national R&D network maps show that a significant area of the northern United States has long been overlooked from participating and/or contributing fully to this country's R&D efforts enabled by advanced networking.

Data-intensive applications increasingly rely on broadband networking and advanced computing. Applications include modeling and simulation for the development of virtual and animated environments as well as collecting, mining and sharing the simple to the most complex scientific data sent from RFID tags, microprocessors, microscopes, telescopes, satellites and databases located world-wide.

As a result of the telecommunications dot.com meltdown starting about 2000, higher education institutions around the country have been acquiring fiber optic assets to be in charge of their own networking destiny. The NTNC states are making coordinated efforts to acquire fiber-optic assets and these efforts will position the NTNC to provide critical links toward the ultimate goal of connecting the region to national and international advanced networks (Figure 2).

The White House has clearly stated that the U.S. government has a vested interest in seeing that leading-edge networking and advanced computing thrive throughout our country to maintain our global leadership. Expanding the advanced computing network is essential in efforts to keep America preeminent at the frontier of research and education and to keep the U.S. economy thriving. This project is specifically aligned to produce results supporting these efforts.

5.2 Tennessee: The OneTenn plan

Introduction and background

OneTenn is a proposed CI for research and education for the state of Tennessee. The plan for OneTenn, which was originally contained in a report/proposal to the Tennessee state EPSCoR committee [3], is noteworthy for offering one of the first state CI plans that takes a comprehensive point of view, encompassing not only networking, but storage and computation as well. It describes the creation of a state-of-the-art technical infrastructure based on the newest developments in optical networking and distributed data storage. The plan incorporates the three leading features of cyberinfrastructure — it provides access to large scale essential computation and storage resources for shared use, it enables collaboration across organizational and geographic boundaries and barriers, and it is designed to both stimulate and incorporate innovation in a way that maximizes the benefits of this statewide facility for all Tennesseans.

The planning process leading to the development of the OneTenn proposal was very broad based. It was led by a state wide Cyberinfrastructure Commission (CIC), charged by the Presidents of the University of Tennessee (UT), the Chancellor of the Tennessee Board of Regents (TBR), and the Chancellor of Vanderbilt University to create a long-range, higher education cyberinfrastructure plan for Tennessee. CIC was co-chaired by senior officials from the UT and TBR systems.

Representatives from UT Knoxville and the University of Memphis co-chaired the two working teams, which included representatives from across higher education in the state. An optical network team researched the network infrastructure and fiber assets across Tennessee and developed the technical architecture, implementation schedule, and cost estimates. A research computing team (which also included a representative from the Tennessee Department of Economic and Community Development) surveyed all four year public institutions and Vanderbilt University to identify the principal research activities and promising statewide research collaborations, catalogued existing or planned CI resources at the different sites, and discussed and analyzed alternative methodologies for sharing these assets.

The plan

OneTenn proposes a focused, phased statewide cyber-infrastructure implementation for all higher education institutions in Tennessee. Of principal interest, the OneTenn cyber-infrastructure would interconnect with the Oak Ridge National Laboratory and its advanced FutureNet infrastructure as well as other regional and national infrastructures, such as Southern Light Rail, Internet2, and National Lambda Rail. Founded on a strategy of acquiring affordable 20-year dark fiber leases and ownership of carrier-grade optical equipment, OneTenn would deploy a super-fast network essential for advanced research and education purposes. Its core optical infrastructure could simultaneously support production network services, such as web, e-mail, video conferencing for course sharing, consolidated systems/services and general Internet access as well as a cutting-edge network to serve scientists and network researchers. This scaled network infrastructure could also provide the core backbone for future growth and expansion to a variety of statewide connectivity and resource sharing needs. The OneTenn network would also provide for seamless interconnectivity with the State's managed network solution, allowing all state institutions/entities to share resources and communicate in the most effective manner possible.

The other main element of the OneTenn plan is distributed storage. OneTenn's CI would unify high performance networking and a wide area deployment of high capacity storage infrastructure in one system, creating a pool of resources to enable collaboration in ways that would otherwise be impossible. Along with its state-of-the-art communications backbone, its key innovation is a new form of network storage technology based on the Internet Backplane Protocol (IBP). Developed at UT with support from the NSF, DOE, and UT's Center for Information Technology Research (CITR), IBP is used by research and education groups internationally. IBP extends the Internet design for interoperability to storage resources, making it possible to aggregate and use the resources of widely scattered storage nodes, called "depots," as if they were one giant storage pool. By integrating a state-of-the-art optical networking infrastructure with high capacity, fully interoperable data depots deployed to every public college and university (as well as Vanderbilt U), OneTenn would enable resource sharing, data intensive collaboration and rapid innovation across Tennessee's research and higher education community.

Goal

The goal of this OneTenn strategy is (in phases) to deliver a carrier class, very fast multi-gigabit backbone to eventually connect all higher education institutions in Tennessee, including five campuses of the University of Tennessee and the 19 institutions in the Tennessee Board of Regents (Figure 3). Each school would have a high speed connection to the backbone (e.g. 10 gigabits for the research campuses or 1 gigabit for smaller campuses). The 10 gigabit core backbone provides necessary capacity for campus-to-campus connectivity while also allowing for pooled, extremely low cost Internet access for all institutions. Phase 1 would connect all four year campuses, and Phase 2 would connect the two year campuses. Private institutions would also be able to connect to this network.

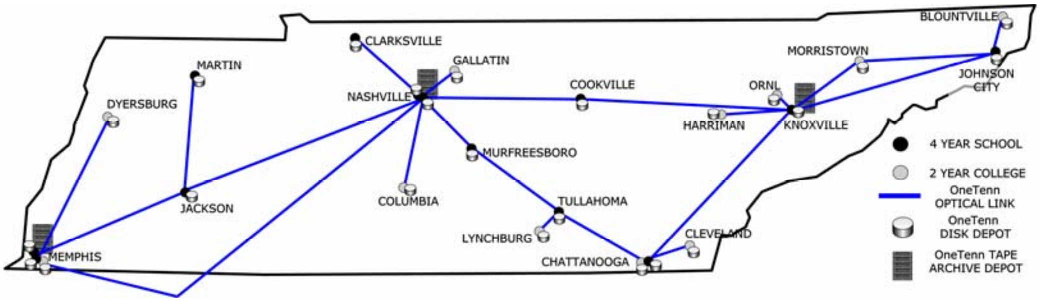

Figure 3: The proposed plan for the OneTenn cyberinfrastructure. All of the four year and two year institutions are shown.

On the storage side, the research intensive universities, which correspond with the network POPs in Knoxville, Nashville and Memphis, would form the major storage backbone. These sites would have storage systems, archival (long-term) storage systems, and computing clusters for high performance buffering services. By the end of the third year, all four year universities would have storage/compute clusters, all two year colleges would have high capacity storage systems, and some test storage systems would be deployed at select P-12 and tech centers. Upon completion the total capacity of the OneTenn infrastructure would be approximately 10 PB.

The three stage OneTenn implementation plan would capitalize on the historical role that research universities across the country have played in leading advanced cyberinfrastructure development in other states. UTK, Vanderbilt, UTHSC and the University of Memphis, are already well positioned to immediately take advantage of the significant new resources that OneTenn would make available. These institutions already have substantial research communities involved in data intensive research, well developed campus networks, and high capacity external connectivity to the nation's research networks. Building on the foundation of these principal research universities, the OneTenn plan would progressively strengthen the capacity of all Tennessee universities and colleges at different levels of cyberinfrastructure development to participate in, utilize, and benefit from the OneTenn cyberinfrastructure.

The first major services for OneTenn would focus on data intensive research collaborations across the state, including those which exploit the instruments and supercomputing capabilities of Oak Ridge National Laboratory, and on the movement and staging of very large multimedia files for distributed learning. In the longer term, OneTenn would create a platform for innovation that will produce new services that empower all the members Tennessee's higher education community to exploit the benefits of data intensive collaboration in all dimensions of their respective missions.

The OneTenn proposal was unanimously endorsed by the Tennessee EPSCoR committee in 2005. The plan was subsequently reviewed and embraced within the higher education community and various state government entities including the Governor's office. In spring, 2006 at the request of the Secretary for Finance and Administration, an infrastructure which could support the operating requirements of OneTenn was included in a state network initiative led by the office of Information Resources in state government. At the same time, the principal research institutions in Tennessee also began working with ORNL to determine the feasibility of using a state wide fiber backbone that ORNL has acquired to serve as a research network to connect the principal research facilities in Knoxville (ORNL, UTK), Nashville (Vanderbilt University), and Memphis (University of Memphis, UTHSC). At the mid-point of 2006, there are several viable initiatives underway in Tennessee to provide the critical network backbone proposed in OneTenn. Although the storage component of the OneTenn vision was considered by some readers of the report to be its most innovative element, there are currently no plans in place or underway to implement it.

5.3 Kentucky: Center for Computational Sciences

The focus of cyberinfrastructure in Kentucky is the Center for Computational Sciences on the campus of the University of Kentucky in Lexington. This is an Interdisciplinary research center, independent of college/department structure, which encourages innovative/non-traditional uses of computation. It is connected with, but not part of, the UKCC (UKy Computing Center) which is one of the largest Hewlett-Packard academic facilities. It is part of the National Computational Science Alliance (NCSA), and was funded by the NSF PACI (Partnerships for Advanced Computing Infrastructure), the predecessor to the new NSF Cyberinfrastructure program. The center has a total of 3.3 Tflops consisting of two supercomputers, a single memory HO Superdome and a 256 processor HP Linux cluster. Fifteen of each machine will be dedicated to the Computational Chemistry Grid (CCG) project (.5-.63 Tflops). The Center has a Gigabit connection to the Internet.

The CCG is designed to provide a collection of grid-based resources to routinely run chemical physics applications. It is a prototype of a distributed infrastructure for open scientific research. It focuses on an application space not requiring a high-speed network in its infrastructure and integrates a desktop environment into an infrastructure for a specific community of users, which will include computational chemists with both small and large scale needs and experimental chemists who occasionally need simulation capabilities to verify experimental results.

The CCG is funded by a NSF Grant under the NMI (National Middleware Initiative) program for $3 million. The lead institution is UKy, as a follow-on to its role of the Chemistry hub in the NCSA . The partners are the U. Illinois (NCSA), Louisiana State U. (CCT), U. Texas (TACC), and the Ohio Supercomputing Center. Technical teams to deploy applications, develop middleware, initiate training/education have been established. The ultimate goal is “point and click” access for computational scientists to the national grid.

The CCG provides production infrastructure to an amenable community of researchers, and lowers the barrier to use of significant computational resources for entire community. Large center resources are often difficult to use due to policies, which exclude computational chemistry applications that typically run on relatively few processors for extended periods. It leverages extant technologies, such as Condor, GridFTP, and GSI, and integrates commonly used computational chemistry codes such as Gaussian 98/03, GAMESS, MolPro, NWChem, and QMC. The plan is to eventually include molecular dynamics, nanotechnology, bio informatics, and other materials/biomedical areas.

Another promising project is the SURA Coastal Ocean Observing and Prediction (SCOOP) Program. The vision is providing community-wide information services and technologies that advance the sciences of prediction and hazard planning for our nation's coastal populations. The objective of the program is integrating diverse efforts and empowering a virtual community of scientists with the tools, resources, and ideas that lead to discovery. The purpose is to promote the effective and rapid fusion of observed oceanographic data with numerical models and to facilitate the rapid dissemination of information to operational, scientific, and public or private users.

In a related effort, Kentucky has established the Center for Resilient Information Systems (CRIS). This is a collaboration of UKy, UL eCavern, IBM & Cisco, and is funded by the state’s Treasury department. It will design and install a state-of-the art asynchronous backup system and evaluate system performance to identify limitations and bottlenecks and vulnerabilities of current systems. In these systems, network throughput, not storage is the bottleneck. The purpose is to increase security by “geographic diversity”, - i.e., increase the distance between primary and backup machines, and increase consumer confidence by enhancing the security of financial records.

These examples are the beginnings of the kind of partnerships that the states research and education leaders expect to be the primary component of future cyberinfrastructure.

5.4 Texas: Texas Internet Grid for Research and Education (TIGRE)

The Texas Internet Grid for Research and Education (TIGRE) is a project to provide a grid computing infrastructure that enables integration of computing systems, storage systems and databases, visualization laboratories and displays, and even instruments and sensors across Texas. The overall goal of the project is to facilitate new modalities for academic-government-private research partnerships, fostering collaborations and partnerships among universities and industries by dramatically enhancing both computational capabilities and collaborative research infrastructure. The TIGRE project is targeted to address research areas of interest to the state of Texas, where manifold increase of computing power, data access, and collaboration would be necessary. The TIGRE brings together leading researchers and TIGRE-funded personnel at each of five universities; Rice University, Texas A&M University, Texas Tech University, the University of Houston, and the University of Texas at Austin.

TIGRE’s high-level objective is to deploy a collaborative grid infrastructure and set up a working grid services architecture, specifically designed to create a software stack that initially supports a minimum of three broadly based driving applications of interest to the state of Texas, but is easily extensible to a variety of applications. As a further goal, TIGRE brings researchers together at multiple institutions into a coordinated grid application development community, able to demonstrate new and enhanced and storage handling capabilities offered by a statewide grid infrastructure.

The TIGRE member institutions have provided access to a limited number of computational and storage resources for the initial users so that they can evaluate software, services, and operational approaches. It is a priority of the project to ensure that the software, including grid middleware and client tools, support center tools, and other developed resources are capable of handling large operational loads when deployed.

Operationally, the project is organized as a group of cooperating developers active across the institutions under high-level control of a steering committee. Functions of the project explicitly under the control of the steering committee include establishing procedures regarding usage allocations; policies and usage priorities; selection of the initial project application areas; and working interactively with researchers and educators to stimulate, formulate, and guide applications development. The developers group has control of the specific technical and software choices needed to achieve the defined high-level goals. This consists of a range of technical tasks such as selecting, testing, and deploying specific software, defining and carrying out operational policies, providing documentation, providing user support and working with the pilot application deployment. Either of these organizational components may choose to elect an internal lead member or chair to coordinate its activities, convene and call meetings, and serve as a central point of contact for further activity.

The responsible lead of the activity or sub-activity in question is responsible to oversee the management of the task and to update web pages pertaining to reporting the progress of his or her group’s activity. Each activity is designed to fulfill a well defined area of need that arises either out of project milestones or from emerging work identified to lay the background needed to meet future milestones and project goals.

Specific choices made by the project so far include the selection of a desired set of functionality for grid job submission, identity certificate management, and support data transfer using Globus web services supplied from the Virtual Data Toolkit in its recent versions. A user information portal has been created using GridSphere and the GridPort tools, with activity underway to add functionality including job submission through the portal in the future. Command-line tools for client use and server features have been identified and are available for distribution to TIGRE sites. This software set has already been installed successfully at institutions beyond the TIGRE core team, which is a good indicator of future uptake of the tools at institutions beyond the initial developers.

Further formal reporting of progress on the project is done on a quarterly basis to the state. Completed milestones within each quarter are assembled by the developers group and forwarded to the steering committee. The steering committee then meets via telephone or video-conferences and/or exchanges e-mail as required to review the progress of the project to approve and complete the report for delivery to the State Department of Information Resources as well as to provide feedback regarding the progress if needed to the developers group. Internal TIGRE web pages are also updated as needed to reflect the progress and upcoming activities for common reference. Their web site provides information about milestones and deliverables for the project, activities that are ongoing or have been completed, and serves as a repository for meeting minutes.

The developers and, optionally the steering committee, periodically hold face-to-face meetings as required when there are discussions needed or tasks to perform that are best accomplished in person. Potential users are sometimes involved in these meetings to gather their feedback as well as to identify new applications that could gainfully exploit the TIGRE grid services architecture. Strong feedback between members of the TIGRE project and the user community is used at all times to ensure that the services being developed are responsive to the needs of actual applications and users.

TIGRE serves as a model for state-wide infrastructure development and works closely with its partner project, LEARN (the Lonestar Education and Research Network), which is charged with deploying high-speed fiber networking to TIGRE institutions as well as more widely to other higher education institutions throughout the state. By growing grid and related application software, TIGRE hopes to serve the research and education needs for collaborative, efficient use of storage and computing resources and will provide a strong foundation for further cyberinfrastructure work in Texas.

6. Conclusion

The workshop participants came away united in the belief that advanced cyberinfrastructure will become more and more essential as a foundation for research and education in all fields of science and engineering. Moreover, there seems to be little doubt that this fact will produce powerful collateral effects that will ripple through other social spheres. Consequently, societies which seek to remain competitive by pushing across new frontiers of knowledge and understanding must plan to invest in building up this critical form of infrastructure. At the same time, however, and for the same reason, individual states that are already struggling with resource challenges are understandably anxious about whether they will be able to keep up. More prosperous states, like California and Texas, are already hard at work on next generation CI for their research and education community. Concern on the part of less prosperous states about falling further behind or spiraling downward in this area cannot be taken lightly.

Of course resource limitations, which can be endemic, are not the only problematic factor. Although the workshop showed that the elements of a successful CI planning process can be extracted from the community's cumulative experience and be laid out in steps that individual states and regional organizations can follow, various obstacles still impede progress along that path. In order to make it feasible for states across the board to begin to engage and invest in the next major phase of CI development, various challenges that surfaced during the meeting will have to be addressed:

- *Lack of established standards for interoperability in key technologies* – As the success of the Internet demonstrated, engineering for interoperability is fundamental to fostering and supporting cross domain collaboration, and doing so in a way that can be sustained in the face of relentless innovation. Unfortunately, for critical CI resources like storage and computing, the broadly accepted standards necessary to build interoperable infrastructures are largely missing. Industry, left to its own devices, is highly unlikely to deliver such standards, and at community wide organizations, like the Global Grid Forum, progress is measured. A strategy to move forward more quickly on this front would provide a major boost to CI planning.
- *Need for entry level approach for individual campuses to engage national CI* – Well funded national CI efforts, such as the TeraGrid, are clearly moving forward, but at present participation by individual campuses, at least in the form of the deployment of dedicated hardware, software and human resources, appears to be relatively limited. A possible reason for the slow rate at which campuses and research groups are joining is that the perceived return on investment is not high enough. Various approaches might be considered for incentivizing and jumpstarting this process, such as a "Connections to the TeraGrid" program to provide small amounts of start up funding or a scaled down TeraGrid (or OSG, DataGrid, etc.) node kit that improves the perceived cost/benefit ratio.
- *Absence of widely known and well understood CI alternatives in the local area* – The previous two points highlight, in different ways, the fact that the CI movement has largely been a top-down phenomenon, carried out on a national scale. Unlike the rise of the Internet, which occurred during a time in which local area networks were already proliferating rapidly, grid technologies that focus on aggregating *local* CI resources in order to deliver mostly *local* benefits have not been widely adopted and deployed. This tends to limit grass roots participation in the growth of advanced CI and thereby inhibits its spread at the state level. Since there are such CI technologies with a local focus, and since strictly local deployments (i.e. single campus) can avoid some of the complications and costs of infrastructures that span administrative domains, strategies which encourage their use would likely

Deleted: ance

strengthen grass-roots understanding and support for advanced CI. It seems clear that such increased intellectual and political investment at the local level would provide a much more robust foundation for CI planning efforts at the state level.

This set of challenges for CI planners and advocates at the state level is far from exhaustive. However, all of them are of a more general nature that individual states can do little to affect. They require strategic action at a national or perhaps even an international level. The National Science Foundation is well positioned to formulate and invest in such a strategic effort. In many cases, all that may be necessary is increased funding for, or minor adjustments of emphasis to, existing CI-related programs. New initiatives explicitly targeting these issues might also be appropriate. In any case, the workshop made it evident that there is a significant reservoir of pent up desire among a large group of state research and education communities to move to the next stage in cyberinfrastructure creation and deployment.

7. References

[1] D. Atkins, et al., "Revolutionizing Science and Engineering through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Panel on Cyberinfrastructure," Panel Report, pp. 1-84, January, 2003. http://www.nsf.gov/od/oci/reports/atkins.pdf.

[2] D. Reed, et al., "Computational Science: Ensuring America's Competitiveness," Computational Science Subcommittee (PITAC), National Coordination Office for Information Technology Research and Development, pp. 117, 2005. http://www.nitrd.gov/pitac/reports/20050609_computational/computational.pdf.

[3] B. Bible, et al., "OneTenn: A 21st Cyberinfrastructure for Tennessee," Tennessee EPSCoR Project, Nashville, TN, pp. 47, March, 2005. https://my.tennessee.edu/pls/portal/docs/PAGE/EPSCOR/ONETENN/ONETENN_CI_PLAN_06APR05.PDF.

[4] Computer Science and Telecommunications Board (CSTB) of the National Research Council, "Realizing the Information Future: The Internet and Beyond." Washington, DC: National Academny Press, 1994, pp. 320.

8. APPENDICES

A-1 WORKSHOP AGENDA

Wednesday, May 10, 2006

6 pm Workshop Reception

Thursday, May 11, 2006

7:00am Continental Breakfast

8:00 Welcome and Introductions
Dr. Micah Beck, Steering Committee Chair
Mr. Matt Kisber, TN ECD Commissioner
Mr. Dan Marcum, Chair of the TN EPSCoR State Committee
Dr. Sherry Farwell, NSF EPSCoR Director

Session 1: The State of Cyberinfrastructure
Session Chair: Dr. Micah Beck, University of Tennessee
8:30 Dr. Miriam Heller, NSF Office of Cyberinfrastructure
The NSF CI Perspective; Concept and Opportunities
9:15 Dr. John Cobb, Oak Ridge National Lab
TeraGrid Overview: National Scale Cyberinfrastructure for Science
9:45 Dr. Joe St Sauver, University of Oregon
Understanding National Optical Networks
10:15 ***Break***
10:30 Dr. Cherri Pancake, Oregon State University
Making CI Accessible and Appealing to Users: Lessons Learned from Early Efforts
11:00 Panel Discussion: CI Definition and Status; Dr. Micah Beck, Moderator
Dr. Scott Lathrop, Argonne National Lab, *Communities Using TeraGrid*
Dr. Mary Fran Yafchak, SURA IT Program Coordinator
12:00 ***Lunch***

Thursday, May 11, 2006

Session 2: Applications and Communities
Session Chair: Dr. Miriam Heller, NSF
1:00 Dr. Frank Williams, Arctic Region Supercomputing Center
Building a Broad Base for CI
1:30 Dr. Peter McCartney, NSF
CI Applications to Long Term Ecological Research
2:00 Panel Discussion: Community Impact – Moderator, Dr. Greg Sedrick,
Dr. William Sanders, *VT Blacksburg Electronic Village*
Dr. Jeffery Scott Averbeck, *Connecting the Corridor*
2:45 ***Break***
3:00 Breakout Session 1:
What are the prerequisites for successful implementation of Cyber-infrastructure in your state?
What are the realistic benefits of Cyber-infrastructure deployment in your State?
5:00 Reports from Breakout session

5:30 Adjourn ----- Evening is free

Friday, May 12, 2006

7:00 Continental Breakfast – **Cumberland South Foyer**

Session 3: Planning and Impact in EPSCoR States

Session Chair: Dr. Terry Moore

8:00 Panel Discussion: CI Plan Development and Implementation
- Dr. John Connolly, University of Kentucky
  - **CI Planning in Kentucky**
- Dr. Doug Hurley, University of Memphis
  - **The OneTenn Plan**
- Dr. Bonnie Neas: North Dakota State University
  - **Northern Tier Network**
- Dr. Alan Sill, Texas Tech University
  - **TIGRE, Implementing Cyberinfrastructure Standards Statewide**

9:45 ***Break***

10:00 Breakout Session 2 :
What are the prerequisites for development of a successful CI plan?
What is required for successful plan implementation?

11:30 Report from Breakout session

12:00 Closing Remarks – Dr. Sherry Farwell

12:30 Adjourn Workshop

A- 2 Workshop attendees

Babette Allina
University of Rhode Island/RI EPSCoR
elissia@uri.edu

Asai Asaithambi
University of South Dakota
asai.asaithambi@usd.edu

Jeff Averbeck
SMARTech Corp - Tennessee
jsa@smartechcorp.net

Puri Bangalore
University of Alabama at Birmingham
puri@cis.uab.edu

Alexander Barzilov
Western Kentucky University
alexander.barzilov@wku.edu

Brian W. Beck
University of Nevada, Reno
beckbw@unr.edu

Micah Beck
University of Tennessee - Knoxville
mbeck@cs.utk.edu

Nathan Beemer
Kansas State University
beemern@telecom.ksu.edu

Mark Bengel
State of TN
mark.bengel@state.tn.us

Brice Bible
University of Tennessee - Knoxville
brice-bible@tennessee.edu

Stephen Borleske
DBI/Univ of Delaware
borleske@dbi.udel.edu

Kristin Bowman-James
Kansas NSF EPSCoR
kbjames@ku.edu

Dennis Brewer
University of Arkansas
dbrewer@uark.edu

Peter Bridson
University of Memphis
pbridson@memphis.edu

Jack Buchanan
University of Tennessee Health Science Center
jbuchanan@utmem.edu

Diane Buehre
South Dakota EPSCoR
diane.buehre@sdstate.edu

Dave Bullard
Clemson University – South Carolina
dave@clemson.edu

Doug Byers
Kansas NSF EPSCoR
dbyers@ku.edu

Bryan M. Carson
Western Kentucky University
bryan.carson@wku.edu

Jeffrey Carver
Mississippi State University
carver@cse.msstate.edu

Kathryn Cataneo
NH EPSCoR; University of New Hampshire
k.cataneo@unh.edu

Terrilani J. Chong
IMUA NSF Hawaii EPSCoR
admepscr@hawaii.edu

Fred Choobineh
Nebraska EPSCoR
fchoobineh1@unlnotes.unl.edu

Frank C. Clark
Medical University of South Carolina
clarkfc@musc.edu

John W. Cobb
Oak Ridge National
cobbjw@ornl.gov

John W. D. Connolly
University of Kentucky
JWDC405@aol.com

Guy Cormier
University of Puerto Rico
guy@hpcf.upr.edu

Gary Crane
SURA – Washington, D.C.
gcrane@sura.org

Peter T. Cummings
Vanderbilt University
peter.cummings@vanderbilt.edu

Robert Dalton
University of New Hampshire
robert.dalton@unh.edu

Gayle Dana
Nevada NSF EPSCoR
Gayle.Dana@dri.edu

Jenny Q. Du
Mississippi State University
du@ece.msstate.edu

Paul Duong Tran
University of Wyoming
qduongtr@uwyo.edu

Datasha Edwards
Nevada EPSCoR
edwards5@nevada.edu

Joseph B. Evans
University of Kansas
evans@ku.edu

Sherry O. Farwell
National Science Foundation
sfarwell@nsf.gov

Bill Figg
Dakota State University
william.figg@dsu.edu

Lillian Gamache
Vermont EPSCoR
lillian.gamache@uvm.edu

Arlene Garrison
University of Tennessee
garrison@utk.edu

George Garrison
University of Tennessee Space Institute
ggarriso@utsi.edu

Jim Gershey
Louisiana Board of Regents
gershey@laregents.org

Dave Goetz
State of Tennessee
Dave.Goetz@state.tn.us

Manuel Gomez
University of Puerto Rico
mgomez@hpcf.upr.edu

Peter Goodwin
University of Idaho
ruthsb@uidaho.edu

James Gosz
NSF EPSCoR
jgosz@nsf.gov

Les Guice
Louisiana Tech University
guice@latech.edu

George Happ
University of Alaska, Fairbanks
fndmm@uaf.edu

Sandra H. Harpole
Mississippi State University
sharpole@research.msstate.edu

Tomasz Haupt
Mississippi State University
haupt@cavs.msstate.edu

John Hehr
University of Arkansas
jghehr@uark.edu

Miriam Heller
National Science Foundation
mheller@nsf.gov

David M. Hercules
Vanderbilt University
david.m.hercules@vanderbilt.edu

Tom Hickerson
State of TN
Tom.Hickerson@state.tn.us

Paul Hill
West Virginia EPSCoR
hill@wvepscor.org

William F. Hogue
University of South Carolina
hogue@sc.edu

Douglas E. Hurley
University of Memphis
dhurley@memphis.edu

James C. Iatridis
Vermont EPSCoR, University of Vermont
james.iatridis@uvm.edu

Gary Johnson
University of North Dakota
garyejohnson@mail.und.nodak.edu

Kenneth Kaneshiro
University of Hawaii
kykanesh@hawaii.edu

Richard Keller
Sinte Gleska University
Richard.Keller@sinte.edu

Dorette Kerian
University of North Dakota
dorettekerian@mail.und.edu

Michael Khonsari
Louisiana Board of Regents
khonsari@laregents.org

Barbara Kimbell
New Mexico EPSCoR
bkimbell@unm.edu

Barbara L. Kissack
Wyoming EPSCoR
bkissack@uwyo.edu

Sarah Koerber
University of Idaho
skoerber@uidaho.edu

Tina M. Koopmans
University of the Virgin Islands
tkoopma@uvi.edu

Barbara A. Kucera
University of Kentucky
bakuce2@uky.edu

Karl Lalonde
South Dakota School of Mines and Technology
karl.lalonde@sdsmt.edu

John C. Lankford
University of Tennessee
jcl@utk.edu

Scott Lathrop
TeraGrid - U. Chicago
lathrop@mcs.anl.gov

Jo-Ann Leong
Hawaii Institute of Marine Biology
joannleo@hawaii.edu

Randy Lewis
Wyoming EPSCoR
silk@uwyo.edu

Seth Lilly
West Virginia EPSCoR
lilly@wvepscor.org

Dennis Lindle
Nevada
lindle@unlv.nevada.edu

Yunkai Liu
University of South Dakota
Yunkai.Liu@usd.edu

Julio Lopez-Ferrao
National Science Foundation
jlopezfe@nsf.gov

Carol Lushbough
University of South Dakota
Carol.Lushbough@usd.edu

Richard Machida
University of Alaska, Fairbanks
rm@alaska.edu

Doug MacTaggart
NSF EPSCoR
dmactagg@nsf.gov

Dan Marcum
TN EPSCoR
dmarcum@marcumcapital.com

Joann McCafferty
DBI/Univ of Delaware
mccafferty@dbi.udel.edu

Peter McCartney
National Science Foundation
pmccartn@nsf.gov

Gail McClure
Arkansas Science & Technology Authority
Gail.McClure@arkansas.gov

Keith McDowell
University of Alabama
keith.mcdowell@ua.edu

Mary Alice Mixon
University of Arkansas
mmixon@uark.edu

Terry Moore
University of Tennessee, Knoxville
tmoore@cs.utk.edu

Juana Moreno
University of North Dakota
juana.moreno@und.nodak.edu

Jeffrey W. Mossey
University of Kentucky
epscor@uky.edu

Bonnie Neas
North Dakota State University
bonnie.neas@ndsu.edu

Henry Neeman
University of Oklahoma
hneeman@ou.edu

Vicki Nemeth
Maine EPSCoR
vicki.nemeth@umit.maine.edu

Patrick Nichols
University of North Dakota
pnichols@chem.und.edu

Jerry Odom
University of South Carolina
odom@sc.edu

Cherri M. Pancake
Oregon State University
pancake@nacse.org

Barbara Paschke
Kansas NSF EPSCoR
paschkeb@ku.edu

Lori Phillips
University of Wyoming
lphil@uwyo.edu

Valerie Pogue
Oklahoma EPSCoR
vpogue@okepscor.org

Sheri D. Powell
Alabama EPSCoR
sheri.powell@ua.edu

Donald Price
University of Hawaii
donaldp@hawaii.edu

Predrag V. Radulovic
University of Tennessee
predrag@utk.edu

Kameswara Rao
Wichita State University
kamesh.namuduri@wichita.edu

Carl L. Reiber
University of Nevada in Las Vegas
reiber@ccmail.nevada.edu

James Rice
South Dakota EPSCoR
james.rice@sdstate.edu

Stephen Roberts
University of Nevada Las Vegas
stephen.roberts@unlv.edu

Karen Sandberg
NSF EPSCoR
ksandber@nsf.gov

William H. Sanders
Blacksburg Electronic Village
sandersw@vt.edu

Rick Schumaker
Idaho EPSCoR
rschumak@uidaho.edu

Pete Schweitzer
University of Alaska, Fairbanks
fndmm@uaf.edu

Gregory A. Sedrick
University of Tennessee Space Institute
gsedrick@utsi.edu

Jeff Seemann
University of Rhode Island/ RI EPSCoR
elissia@uri.edu

Jean'ne M. Shreeve
University of Idaho
jshreeve@uidaho.edu

Alan Sill
Texas Tech University
Alan.Sill@ttu.edu

Richard Sincovec
University of Nebraska - Lincoln
sincovec@cse.unl.edu

Henry H. Smith
University of the Virgin Islands
hsmith@uvi.edu

Randy Smith
University of Alabama
rsmith@cs.ua.edu

Joe St. Sauver
University of Oregon computing center
joe@uoregon.edu

Vickie Stanfill
State of TN
Vickie.Stanfill@state.tn.us

Wilbur Stolt
University of North Dakota
wilburstolt@mail.und.edu

Mike Strauss
University of Oklahoma
strauss@nhn.ou.edu

Kevin Streff
Secure Banking Solutions – South Dakota
kevin@protectmybank.com

Anne Sudkamp
University of Alaska, Fairbanks
fndmm@uaf.edu

David Swanson
University of Nebraska-Lincoln
dswanson@cse.unl.edu

Anne W. Sylvester
Wyoming EPSCoR
annesyl@uwyo.edu

Alan Tackett
ACCRE/Vanderbilt Univ
alan.tackett@accre.vanderbilt.edu

Jan Taylor
West Virginia EPSCoR
jtaylor@wvepscor.org

Karen Theodosopoulos
South Dakota EPSCoR
karen.theodosopoulos@sdstate.edu

Fred Tompkins
University of Tennessee Research Foundation
tompkins@tennessee.edu

Ms. Maria Vargas
University of Puerto Rico
m_vargas@crci.uprr.pr

Kai Wang
The University of South Dakota
Kai.Wang@usd.edu

Frank Waxman
University of Oklahoma
fwaxman@osrhe.edu

Brad Weiner
University of Puerto Rico
brad@hpcf.upr.edu

Meri Whitaker
Virgin Islands EPSCoR
mwhitak@uvi.edu

James Wicksted
Oklahoma EPSCoR
james.wicksted@okstate.edu

Frank Williams
University of Alaska Fairbanks
williams@arsc.edu

Warren Wilson
South Dakota Board of Regents
wjwilson@sdbor.edu

Mary Fran Yafchak
Southeastern Universities Research Association
maryfran@sura.org

Yifeng Zhu
University of Maine
zhu@eece.maine.edu

A – 3 List of Abbreviations and Acronyms

ATLAS	A Toroidal LHC ApparatuS
CI	Cyberinfrastructure
CMS	Compact Muon Solenoid
CRIS	Center for Resilient Information Systems
EPSCoR	Experimental Program for Stimulating Competitive Research
IP	Internet Protocol
LEARN	Lonestar Education and Research Network
MCNC	Microelectronics Center of North Carolina
NCSA	National Center for Supercomputing Applications
NDSU	North Dakota State University
NMI	NSF Middleware Initiative
NSF	National Science Foundation
NTNC	Northern Tier Networking Consortium
OneTenn	One Tennessee
ORNL	Oak Ridge National Laboratory
OSG	Open Science Grid
PITAC	President’s Information Technology Advisory Committee
SBIR	Small Business Innovation Research
SCOOP	SURA Coastal Ocean Observing and Prediction
SDSC	San Diego Supercomputer Center
S&E	Science and Engineering
SURA	Southeastern University Research Association
TIGRE	Texas Internet Grid for Research & Education