University Research Cyberinfrastructure Committee

Interim Report

August 31, 2006

Committee Members

Joyce Mitchell, co-chair, Biomedical Informatics
Martin Berzins, co-chair, School of Computing
Kenning Arlitsch, Marriott Library
Tom Cheatham, Medicinal Chemistry
Steven Corbato, Scientific Computing and Imaging Institute
Julio Facelli, Center for High Performance Computing
Steve Hess, Office of Information Technology
Joyce Ogburn, Marriott Library
Wayne Peay, Eccles Health Sciences Library
Pierre Pincetl, Information Technology Services
Edward Rubin, Linguistics
Cassandra Van Buren, Communication
Greg Voth, Chemistry
Mark Yandell, Human Genetics

Merrell Patrick – Office of VP for Research
Shanna Erickson, Administrative Support, Office of VP for Research

Acronyms

ACS – Administrative Computing Services
CACGT – Center for Advanced Computational and Grid Technologies
CDG – Computational and Data Grid
CHPC – Center for High Performance Computing
CI – Cyberinfrastructure
CIAAC – CI Applications Advisory Committee
CICT – CI Coordination Team
CITT – CI Technical Team
INSCC – Intermountain Network and Scientific Computing Center
ITS – Health Sciences Center, Department of Information Technology Services
NLM – National Library of Medicine
NSF – National Science Foundation
OIT – Office of Information Technology
UCDG – University CDG
UCIC – University CI Council
USHE – Utah System of Higher Education

Table of Contents

Appendices

Executive Summary

"Campus cyberinfrastructure is not just about technology."[1] Beyond access to technology, Cyberinfrastructure (CI) defines a new information technology paradigm that includes people and their expertise, enabling technologies, software and tools, and provides a foundation for an integrated approach to research and education workflow. CI should facilitate application use and evolution, data analysis, collaboration and data management. Such model is at odds with the traditional model of investigators. The traditional model of the independent investigator and/or research team has historically been a problematic component of University technology planning and investment. However, the scale of the challenges and the expectations of the funding agencies are redefining the research environment to include interdisciplinary, multi-institutional collaborative projects. The National Institutes of Health Translational Clinical Medicine initiatives exemplify this new model of investigation. Advanced computing, networks, data storage technologies/resources and personnel – Cyberinfrastructure – are essential elements of this new research environment and of the University's success.

The Cyberinfrastructure Committee was constituted with representation from senior investigators and administrators with responsibilities that include infrastructure resources and services. The committee conducted a review of recent reports and publications, presenting national perspectives and priorities. Additional perspectives were offered through invited presentations and dialogues. Considerable effort was invested in the development and administration of a survey of the research community. The 114 responses provide the basis for a number of the committee's recommendations.

COMMITTEE RECOMMENDATIONS

Immediate Actions:

1. Establish a Cyberinfrastructure Council to provide co-ordination, institutional planning/budget recommendations and oversight. The Council will develop institutional priorities and be responsive to the opportunities provided by state and national funding agencies/programs.

2. Reconstitute the Center for High Performance Computing (CHPC) as a campus-wide Cyberinfrastructure Center (CIC) that is a user focused service provider. The Cyberinfrastructure Council will form a subcommittee including major faculty clients of the CIC to provide guidance and oversight. CHPC will transition research activities to extramural funding sources over time.

3. Submit a Utah System of Higher Education Disaster Recovery & Large Scale Data Repository Proposal to the 2007 Utah State Legislature.

4. Formulate a plan for the development of an Institute, with world-class leadership (possibly through U*), to provide campus-wide leadership,

---

[1] Final report: A workshop on effective approaches to campus research computing cyberinfrastructure. National Science Foundation. April 25-27, 2006. Arlington, VA.

encouraging research and collaboration in disciplines exploring Cyberinfrastructure opportunities, ex. Science, Medicine, Engineering, Humanities, Architecture. The plan will identify incentives the institution will provide to encourage participation and collaboration from existing and newly established research centers (Brain Institute, Scientific Computing and Imaging Institute, Huntsman Cancer Institute, Eccles Institute of Genetics, etc). The Cyberinfrastructure Council will be responsible for the formulation and communication of this plan.

High Priority Initiatives:

5. Secure earmarked funding for a large tera-scale class system in keeping with institutional needs in order to meet NSF Cyberinfrastructure initiatives. The Cyberinfrastructure Council will be responsible for the development of a plan for long-term hardware/software acquisition, development and support.

6. The University should provide the baseline of Cyberinfrastructure support expected of a research university for its current and potential investigators. The Cyberinfrastructure Council will develop guidelines and recommendations for Cyberinfrastructure connectivity, hardware, and support.

7. Seek state funding to establish a state-wide Grid activity to enable all the major research Universities in Utah to collaborate and to share resources. This development effort will provide the future framework for Cyberinfrastructure for all of higher education, public education and government agencies in the State of Utah. This Grid would also allow for researchers to lead research teams throughout the US and the world.[2]

8. Initiate the planning process for fund raising, design and construction of a state-of-the-art data center, with the goal of have the facility operational in less than four years. The Cyberinfrastructure Council will be responsible for providing oversight for this activity. This would include a campus-wide data grid.

9. Charge the libraries to provide basic to mid-level support and training for faculty research and data management.

"Cyberinfrastructure has become a key enabler for scholarly research."[3] The University needs to continue to invest in high-performance computing, networking grids, data repositories, disaster recovery, and associated support services in order to remain a leading research university in the 21st century. Senior administration must be responsible for, and invest in, the resources to support the continuing development of cyberinfrastructure.

---

[2] See Appendix A for additional information relating to Grid development.

[3] Final report: A workshop on effective approaches to campus research computing cyberinfrastructure. National Science Foundation. April 25-27, 2006. Arlington, VA.

Introduction

A principal finding in the 2005 report of the President's Information Technology Advisory Committee (PITAC) titled "Computational Science: Ensuring America's Competitiveness" was "Computational Science is now indispensable to the solution of complex problems in every sector, from traditional science and engineering domains to such key areas as national security, public health, and economic innovation."

The increasing complexity, scope, and scale of computational science requires the use of a more integrated infrastructure that takes advantage of the continuing rapid advancements in digital computing, communications and information technologies. A National Science Foundation (NSF) Blue Ribbon Panel notes that "the capacity of these technologies has crossed thresholds that now make possible a comprehensive 'cyberinfrastructure' on which to build new types of scientific and engineering environments and organizations and to pursue research in new ways and with increased efficacy." The NSF addresses this by implementing a new program based on the recommendations in *Revolutionizing Science and Engineering Through Cyberinfrastructure: Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure*, Daniel E. Atkins (Chair), January 2003 (http://www.nsf.gov/od/oci/reports/atkins.pdf).

NSF has further recognized the importance of CI in the conduct of research and education across all areas of science and engineering by creating an Office for Cyberinfrastructure (OCI) whose Director reports to the NSF Director. It, as well as other organizations, has sponsored numerous workshops addressing the importance of cyberinfrastructure in various areas of science, engineering, humanities, social sciences, libraries and education (see Appendix A).

Given the advancements and opportunities that are discussed in the above reports, the final report of the American Council of Learned Societies' Commission on Cyberinfrastructure for Humanities and Social Sciences states that "Cyberinfrastructure is being built much more quickly [than tradition infrastructure], and so it is especially important that humanists and social scientists actively engage with it, articulate what they require of it, and contribute their expertise to its development." This report outlines the need for "more advanced software applications, greater bandwidth, and more access to expertise in information technology. We also heard from many who spoke about the potential for cyberinfrastructure to enhance teaching, facilitate research collaboration, and increase public access to (and fair use of) the record of human cultures across time and space. (see Appendix A).

In the health sciences the Director of National Institutes of Health (NIH) appointed a committee of experts to investigate the needs of NIH-supported investigators for computing resources, including hardware, software, networking, algorithms, and training. A report titled the "Biomedical Information Science and Technology Initiative" (BISTI [2]) was submitted to the NIH Director in late 1999. Based on that report the NIH developed a bio-informatics roadmap for its funding programs. In 2003 the NIH developed the NIH Road Map [http://nihroadmap.nih.gov/overview.asp] that is currently being used to guide interdisciplinary research and funding; all of the Road Map initiatives rely on advanced cyberinfrastructure as the basic support for biomedical sciences.

Given the advancements and opportunities that are discussed in the above reports, as well as the need to examine whether University funds for CI resources are appropriated in a way that addresses University research priorities, the University has appointed a University Research Cyberinfrastructure Advisory Committee to investigate the challenges and opportunities these initiatives offer.

Committee Charge

Committee Charge as specified by the Vice President for Research:

"Assess how current high performance computing, networking and data storage needs for research are being met. Identify current gaps in existing infrastructure that inhibit the development of multi-disciplinary research projects that are a stated priority of the university administration.

Advise on the future needs for research computing, data storage and networking and whether a more integrated (cyber) infrastructure as described in the NSF report would better meet research and education needs and enhance multi-disciplinary research.

Advise on a strategy and an organizational structure for meeting the identified needs.

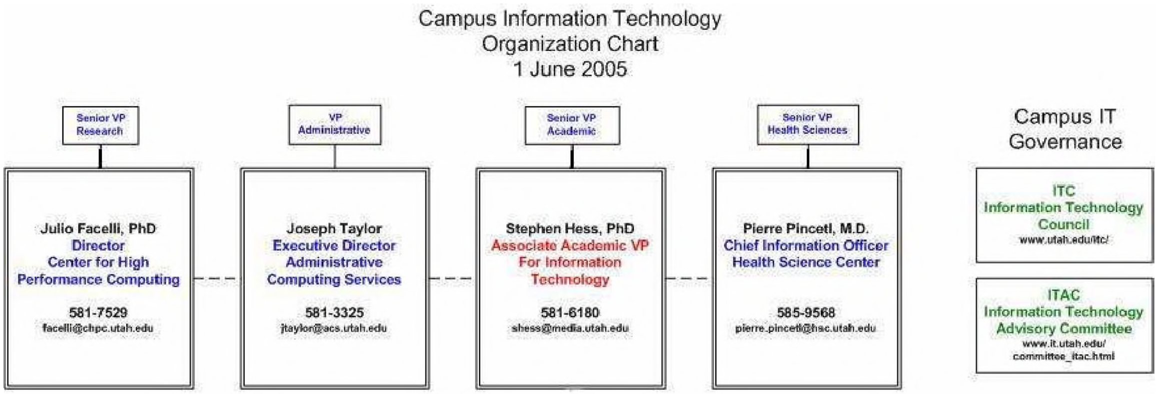
Look specifically at the area of high performance computing and the issue of many distributed clusters versus a more centralized mode, including the issue of the demands for power, cooling and maintenance and support staff. Try to assess the future of the current trend toward addressing research computing needs with the use of clusters. Note: DARPA currently has a program that supports the development of high productivity computers. Might such computers offer a better means for conducting large-scale multi-disciplinary research in the next 5 years?

Advise on strategy for developing additional external resources to support cyberinfrastructure and where future additional funding might be focused. E.g., should CHPC be transitioned to an institute that provides both service to the university community and conducts research to bring in external research funds? Should the university, in partnership with other state institutions, take the lead in developing a statewide cyberinfrastucture, that could meet broader state needs and lead to additional funding.

Advise on how we should be allocating our current central support for high end computing, networking, and related infrastructure activities."

Background

Four separate campus organizations address different aspects of the University's general CI needs. They are the Center for High Performance Computing, the Office of Information Technology, the Health Sciences Department of Information Technology Services and Administrative Computing Services. Each of these organizations reports to different University Vice Presidents – Academic Affairs, Health Sciences, Research and Administrative Services. The organization chart for University IT services appears at (www.it.utah.edu/images/leadership/campus_IT_org.jpg).


Campus Information Technology
Organization Chart
1 June 2005

Below we give brief background information on 4 organizations included in the study. More detailed descriptions of the role and current activities of these organizations are given in Appendix C

The Center for High Performance Computing (CHPC)

CHPC evolved from the Utah Supercomputing Institute as a result of recommendations in the 1995 report of a Committee appointed by Research Vice President, Richard Koehn and chaired by Professor Carleton DeTar, of the Physics Department. It was officially formed by a resolution of President Arthur Smith in September of 1995. Since then CHPC has been tasked with carrying out activities that were not considered in the DeTar Report. In November 1996, President Smith signed a directive tasking CHPC with management responsibilities for distributed computing, security, advanced networking and infrastructure in the Intermountain Network and Scientific Computing Center (INSCC) building. The Security Office was moved along with its budget to the Office of Information Technology in 2001. With the reorganization of IT at the university in June of 1999, CHPC was given added responsibilities in institutional IT R&D, in particular testbeds for new technologies.

The High Performance Strategy Planning Committee of 2000 appointed by Vice President Koehn was asked to look at the appropriateness of these activities in relation to its role in high performance computing and to look at CHPC's role in the future. The Committee chaired by Merrell Patrick, Special Assistant to the Vice President for Research, submitted its report in 2000. The report contained three major recommendations:

a. the University should contribute $250K /year to a capital fund for hardware upgrades,
b. CHPC should move to establish a computational science research initiative, and
c. CHPC should assess the opportunities for advancing the use of high performance computers in the medical area and to assist medical researchers.

Recommendation (a) was implemented for three years but then was dropped in budgeting for 2006. Recommendation (b) was never implemented. Dr. Julio Facelli, Director of CHPC, took steps to implement recommendation (c) and has had some success (see summary of these in the CHPC section in Appendix C) but has been unable to make major advances.

In an attempt to increase the use of CHPC and its resources in advancing research in the Health Sciences, Merrell Patrick, with the encouragement of Research Vice President Ray Gesteland, spent several months meeting with 25-30 individuals in the Health Sciences. As result of these meetings, he wrote and submitted a 2003 report titled "Advancing Biomedical Computing at the University of Utah" to Vice President Gesteland. Dr. Gesteland distributed the report but most of the recommendations in the report have yet to be implemented. The 2000 and 2003 reports can be found on the CHPC website (http://www.chpc.utah.edu/~facelli/CI/).

The Director of CHPC, Julio Facelli, reports to the University Vice President of Research.

Office of Information Technology (OIT)

OIT was formed in 2002 by University leadership to address institutional IT challenges through central planning, policies, and operations under the Associate VP for Information Technology, Stephen Hess. OIT plans are developed based on their ability to assess the needs of the campus community, develop solutions to those needs that have broad campus support, justify the plan based on sound business cases, define project plans that will succeed, and communicate the solutions and services to the campus community to facilitate adoption

Stephen Hess, Associate Vice President for Information Technology, is responsible for the OIT and reports to the David Pershing, Senior Vice President of Academic affairs.

The Department of Information Technology Services (ITS)

ITS was formed in 1996 to provide IT solutions and services to the University of Utah Health Sciences Center. Its mission is to provide access to data in a secure, reliable, and timely manner, to enhance the outcomes of patient care, education, research, and community service and to offer excellent service by meeting and exceeding diverse customer needs. The Data Resource Center (DRC) is a division of Information Technology Services that provides data services and system integration support to all Health Sciences Center organizations as well as affiliated main campus entities. Clinical Information Services is responsible for the implementation of information services for University Hospital. ITS is also responsible for the managing security and complying with HIPAA regulations. The Utah Telehealth Network is a component of ITS providing videoconferencing, clinical services and education support statewide. ITS also manages the Health Sciences Center website with particular emphasis on University Hospital and information and services. An organization chart is at http://uuhsc.utah.edu/its/orgchart/.

ITS is headed by Pierre Pincetl, M.D, Assistant Vice President and Chief Information Officer for Health Sciences. He reports to the Lorris Betz, Senior Vice President for Health Sciences.

Administrative Computing Services (ACS)

The mission of Administrative Computing Services is to fulfill the institutional information needs of the University of Utah community by providing valuable information services. Administrative Computing Services is committed to the strategic use of technology for the continual improvement of the operation of the University of Utah. The major areas of responsibility for ACS are Financial, Employee and Student Systems. Of particular interest to the research community is the Grants Administration System, which is also a responsibility of ACS.

ACS is led by Joe Taylor, Executive Director. He reports to Arnold Combe, Vice President, Administrative Services.

State-of-the-Art at the University

The following are illustrations of advanced computing and networking initiatives at the University that illustrate the importance of Cyberinfrastructure development.

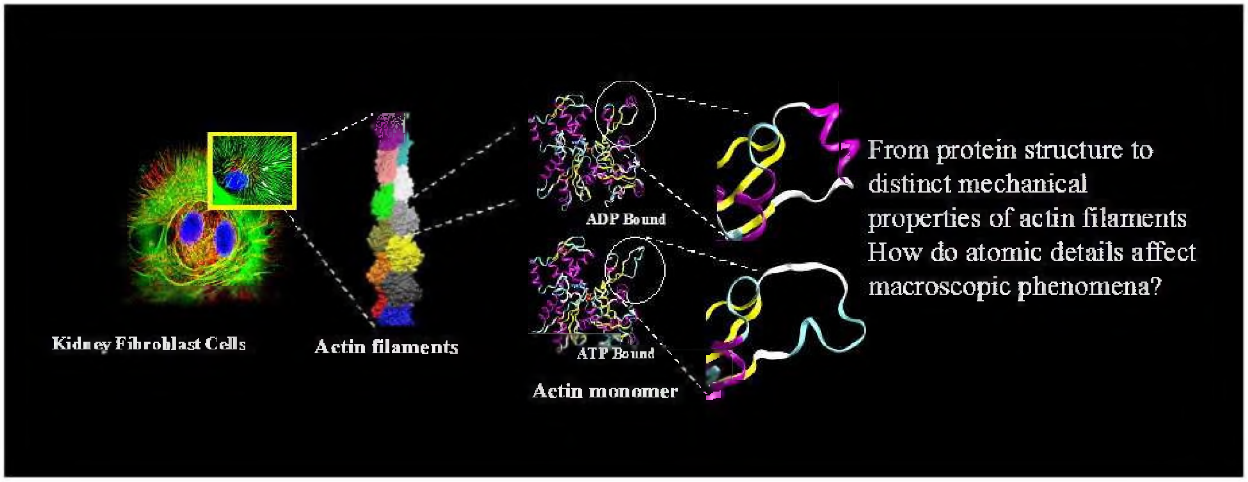
Computational Science and Engineering

Computational Engineering in Utah reflects both the recent National Science Foundation panel on Simulation-Based Science and the PITAC report when they make the case for the importance of modeling and simulation as key elements for achieving progress in science and engineering. Examples of such activities are the multi-disciplinary DOE-funded CSAFE project and DARPA Virtual Soldier Project which encompassed a computational approach to healthcare. These are examples of activities ranging across many departments. Combustion and energy, geophysical and atmospheric/weather simulations are a few of many other notable examples of activities making use of extensive computational resources and with the capacity to expand to accommodate almost any level of compute resources available. Such activities together with associated activities in Institutes such as EGI and the SCI Institute form a substantial part of research income generation. The research undertaken involves the use of perhaps thousands of processors as part of shared DOE resources to the dedicated use of smaller local clusters of processors. This trend will accelerate with new activities such as the Brain Institute and the new energy centers. These activities need to be seen in the context of a rapidly changing global research arena.

The present state of the art in computational science and engineering is that global competition in this area is fierce in both basic science and engineering and related applications. The first petaflop machines (10^15 operations per second), working on petabyte data sets are expected within the next three years. Such machines may well have as many as hundreds of thousands of processors if current IBM architectures are extended or may have a smaller number of more powerful processors if manufacturers such as Fujitsu are first. A key part of the large scale engineering and science undertaken on such machines is collaborative. The extensive use of the grid to promote virtual organizations and large scale collaboration in Europe and Asia is perhaps ahead of the US. For example high schools in Shanghai use the grid to collaborate and share resources. The UK escience program is a multi-hundred million dollar program aimed at getting cyberinfrastructure used in industry and evolving applications. At the same time the advances in simulation capability make it possible to solve industrial problems on a scale hitherto unthinkable. For example US car makers are concerned that the use of the Japanese Earth simulator gives Japanese automakers an edge in design that they do not have. NSF's vision is that in order to compete in this global race it will fund a petaflop machine. As will DOE and other government agencies. Equally importantly the NSF roadmap explicitly assumes that Tier One research institutions will house medium level resources having the order of thousands of processors. The first instances of such computers being funded are the Rensselaer Blue Gene which is a $100M project and Indiana's Big Red machine. The Top 500 list gives other examples the closest to home such as Brigham Young's MaryLou4 cluster ranked at 87 in the world. On a worldwide level, regional universities in Germany, such as Chemnitz, are acquiring machines with thousands of processors. While such rankings may be downplayed as an expensive status game the level of simulation possible with large scale architectures will define who can compete in 21st century engineering and who has to sit on the sidelines. Within this framework the computationally driven research in Utah is potentially well-placed to compete.

Computational Grand Challenge in Molecular Dynamics

The field of computer simulation has contributed significantly to the ongoing revolution in the biophysical sciences. Perhaps the best example is Molecular Dynamics (MD) simulation wherein Newton's equations of motion are integrated in time for an atomistic model of a biomolecular system of interest; for example a protein, usually surrounded by solvent (e.g., an enzyme) or embedded in a lipid bilayer (e.g., an ion channel). MD simulations can now be routinely carried out for systems with tens of thousands of atoms and for trajectories lasting tens of nanoseconds. However, while such simulations may seem both large and long at the atomic scale, at the biological scale they are in fact only a very small part of the overall picture. While MD simulations are without a doubt both valuable and insightful, it is hard to imagine that they can capture the true essence of the vast number of processes occurring in the living cell over a very wide range of length and time scales. To make the situation even more difficult, the computational "tricks" usually involved in MD simulations can introduce artifacts into the simulations that are not real and merely reflect the finite size and time scale of the simulation itself. Despite the remarkable (even heroic) efforts to date in the design and execution of MD simulations of biomolecular systems, real biology is simply more complicated and a new paradigm for the computer simulation of such systems is sorely needed. *This effort involves far more than just computational algorithms. It includes the development of whole new theoretical and methodological concepts, often even re-thinking the foundations of statistical mechanics and condensed matter dynamics.*



In order to address this problem, a computational and theoretical methodology that has the capability of bridging the multiple spatial and temporal scales present in biomolecular systems has been under development in the Voth group, with key results having been published in leading journals. These new concepts are being developed for biological membranes (including membrane-bound proteins), filaments (such as action as shown above), microtubules, nucleic acids, and viral capsids. It is noteworthy that the Voth group computations are featured as an actual required benchmark for bidders on the future $200 million NSF Petascale computer system (see:

http://www.nsf.gov/publications/pub_summ.jsp?ods_key=nsf06573).

Our multiscale methodology is a singular accomplishment coming from the University of Utah in the field of computational science and something upon which the University can build.

Cyberinfrastructure in the Humanities

The problem of the traditional model of the independent researcher mentioned in the Executive Summary is perhaps nowhere greater than in the Humanities, where a diversity of perspectives, intellectual histories, methodologies, and, perhaps most importantly, limited funding opportunities both internal and external to the U, has led to vast differences in the adoption of new research technologies. Despite increasing success in the acquisition of external monies, the Humanities continue to receive a disproportionately small share of internal research resources. Nevertheless, because of their longstanding commitment to interdisciplinary approaches to the study of the most complex of natural phenomena, the human being, researchers in the Humanities have made progress in areas including communication, data storage and dissemination, and the formation of virtual research communities. The University must closely attend to the recommendations of the August 2006 American Council of Learned Societies Commission on Cyberinfrastructure in the Humanities in order to become and remain more competitive for research funds. Establishing a first-rate Humanities Computing Center (whether stand-alone or as part of a larger initiative) should be carefully considered when planning cyberinfrastructure at the U. The following selection of projects highlights both successes and challenges faced by researchers in the Humanities.

The NSF-sponsored Shoshoni language project (PIs Mauricio Mixco and Marianna Di Paolo, both from the Linguistics Department) exemplify one sort of project seen throughout the humanities, in which large amounts of data (sound recordings of spoken language, here) need to be made accessible to a broad community of researchers. The digitization of degrading older media (reel-to-reel tape, here) is a step preliminary to the primary analysis that interests Linguists, Historians, Anthropologists, Sociologists, and others in their creation of dictionaries, grammars, histories, ethnographies, etc. Many other such projects will arise from the NSF and Smithsonian sponsored Center for American Indian Languages here at the U under the direction of Presidential Professor Lyle Cambell, Linguistics.

The Upper Tigris Archaeological Research Project (PI Bradley Parker of the History department) is another example of the strides being made in the use of Cyberinfrastructure in Humanities research. It uses several web-based applications to catalog, store and share all of the information gathered during excavations at the archaeological site of Kenan Tepe in southeastern Turkey. The main database already contains approximately 90% of the data gathered after eight years of excavation including photographs, measurements, plans, journals etc. and works as a kind of electronic notebook system. This system not only archives these data but allows team members to access them remotely and thus permits continued, normal, remote collaboration. To aid further analysis and publication we use an FTP server to move around large files and a project website (www.utarp.org) where we organize publications and conference papers. Unfortunately, because of firewall issues and limited resources at the U, all of this infrastructure is housed at the institution of the project PI's assistant, USC. The PI wants to move the project's equipment (donated by Microsoft) and technical support to Utah, and this may become possible with a pending NEH grant (and the resolution of security issues).

The Speech Acquisition Lab (PI Rachel Hayes-Harb of the Linguistics department) faces similar infrastructure challenges. Primary data in this field consists of high quality sound files, analyzed acoustically and studied using statistical analysis. All data is gathered at a computer terminal, often with specialized equipment (e.g. a sound-attenuated booth), but web-based data-gathering tools are becoming increasingly attractive. The need for software development and for equipment and technical support for data storage and backup are subsequently becoming more urgent. So far, the PI has had to outsource some of these concerns: hiring a programmer, and buying a domain name (www.speechacquisitionlab.net) on a server that can collect online data in the appropriate format (the available university servers apparently could not). Data storage is undertaken on two lab-purchased 250GB hard-drives (added to the college server), which capacity will eventually be exceeded.

The College of Humanities is a leader in integrating the research and teaching missions of the University. With the College of Fine Arts, it will currently require increased computing resources (hard money renewing budget for hardware, software, support staff) to support faculty research and creative development tied to the Minor in Animation; these needs will only grow as the University pursues plans for a Major in Animation. In addition the Department of Communication has grown to include 4 tenure track faculty lines in new media technologies, signifying substantial growth in the computing needs of research-oriented faculty and their graduate students.

These projects represent a sample of the work already being done in Humanities using cyberinfrastructure, but it should be noted that many researchers are not making use of the new technologies because the college still does not receive adequate attention to its requests for resources. It is clear that a baseline standard of research support established at an institutional level would do much to promote broader access (in all colleges) to the increasingly necessarily cyber-tools whose use is flourishing at other institutions, would allow a economy of scale for many specific needs, and would help to protect our institution's RU1 status.

Personalized Medicine and Cyberinfrastructure

The University of Utah Health Sciences Center has as one major goal to become a worldwide leader in Personalized Healthcare. Personalized health refers to using methods of molecular analysis to identify predispositions to diseases and thereby to prevent, diagnose, better manage or treat patients. Personalized health aims to achieve optimal medical outcomes by helping physicians and patients select the best therapeutic approach in the context of a patient's genetic and environmental profile.

The Health Sciences Center is working to develop a broad-based program that takes advantage of a molecular understanding of disease mechanisms to direct preventive measures and therapeutic approaches to the right population of people while they are still well. The foundation upon which the program will be built includes the extensive databases characterizing the Utah population (e.g. Utah Population Data Base, the Utah Genetics Reference Project, and associated linkages to data from the Utah Department of Health, basic research laboratories, and the Electronic Medical Records from multiple institutions), the informatics expertise to capture this knowledge in ways that allow it to be used for patient management purposes, the unique expertise at the University in the identification of genetic determinants of human diseases, the use of mouse models to uncover disease mechanisms and therapeutic targets, and the strengths in pharmacology and drug development including expertise in drug metabolism, toxicology and pharmacogenetics. These elements span the gamut from prevention to treatment, and provide a platform upon which to address the variability in individual patients that is fundamental to the concept of personalized medicine.

This ambitious project must have an advanced and fully functional cyberinfrastructure to succeed. The HSC has larger and more numerous data resources than most other places in the world, but these resources need work to make them fully available to the researchers at the University of Utah. This involves more coordination of services and infrastructure than is currently available. Furthermore, to make the results of molecular analyses available to clinicians will require a level of integration of data and decision support that extends from the molecular laboratories to the electronic medical record. This is one aspect of Translational Medicine Research (spanning from bench to bedside). Many aspects of cyberinfrastructure need attention to realize this goal. The basic science laboratories have need for machine learning and visualization techniques to be able to assist in the discernment of patterns from large data sets. Grid computing is considered standard for the collaborative research projects emerging in this area (see next section for an example in the cancer domain) and will be required to be considered leaders in the field and also to compare our results with those of other research teams. Expertise for constructing, merging, analyzing, maintaining and distributing complex databases and developing clinical decision intelligence is essential for moving forward in this area, especially when considering the scope of the resources that include extensive health and genetic and genealogical records for the entire population of Utah and their relatives. Finding genes in this data set requires extensive processing power. Finding correlations between genotypes and phenotypes and health outcomes requires new analytical approaches, multiple processors, new data models, semantic and syntactic harmonization, controlled vocabularies. All of these research threads require secure and extensive long-term storage. Combining all of these analyses with pharmacogenetic data to find new approaches to treatment or new drugs further dictates excellent cyberinfrastructure that extends far beyond the boundaries of this institution and throughout the government laboratories and into the private sector of the pharmaceutical industry. Most of all, these projects involve moving beyond the technology and engaging the research and clinical community to bridge cultures and enhance collaborative relationships. Cyberinfrastructure is truly the key to realizing our research goals in this arena.

Cyberinfrastructure and Grid Computing for Cancer Research:

The Cancer Bioinformatics Grid (caBIG) is a grid initiative undertaken by the National Cancer Institute (NCI) to share data and tooling across cancer centers. NCI's grid is an interoperable data sharing infrastructure that supports the building of common ontologies, terminologies and data elements for sharing data. It does this work in the domains of clinical trial management systems, integrated cancer research and bench to bedside translational research. It undertakes the difficult task of insuring that the semantic and syntactic definitions of clinically relevant variables are consistent across institutions. Initiated under the directorship of Dr. Andrew von Eschenbach, he stresses its importance for NCI's strategic plan, "Nearly every facet of NCI's strategic plan for 2015 is predicated on the potential of caBIG." This is evidenced by The Cancer Genome Atlas (TCGA) building upon caBIG and requiring compliance for their Biospecimen Repository pilot project. A strong cyberinfrastructure that can support grid architectures is critical for the University of Utah to be competitive for current and future NCI funding.

Parallel Genetic Algorithms to Discover Structures of Atomic Clusters and Molecular Crystals (NSF TeraGrid Award MCA05S018)

This project uses TeraGrid computational resources to continue the development and application of our MGAC (Modified Genetic Algorithm for Crystal and Cluster structures) in the topics described below:

i) Computational GRID implementation of the MGAC method (GRID MGAC) to allow for multiple levels of parallelization and improvement of its load balancing capabilities over the NSF TeraGrid (http://www.teragrid.org/).
ii) Study of the structures and properties of large Si, Si-H and Si-coinage metal clusters using the MGAC/CPMD method to overcome present limitations imposed by methods that use either limited searches and/or very approximate QM methods.
iii) Application of the MGAC to the study of the crystalline structures of flexible molecules (a field in which MGAC is the only technique available), with emphasis on its applications to high energy materials and pharmaceutical drugs.
iv) Study of the convergence properties of parallel GA for determining structures of atomic clusters and crystals, with the goal of developing better and more efficient genetic operators. We also will explore the use of recent techniques developed by the computer science community, like co-evolutionary capabilities, particle swam optimization (PSO), ant colony optimization (ACO), artificial immune systems (AIS), etc.

The Libraries: an essential part of research infrastructure

The three libraries offer many resources and services to support diverse research activities. They supply traditional underpinnings - journals, databases and books – and also e-text, data, statistics, multimedia, images, and the like. The libraries focus on applications, tools, and information services rather than advanced computational support or networking. They offer support for equipment configuration and access to internet resources.

For faculty to perform their research, the libraries manage access to licensed and purchased digital information. The libraries also help faculty convert analog materials to digital formats conducive to advanced research methods. On request the libraries may digitize items in their collections. As rapidly as feasible, the libraries are digitizing collections for incorporation into research.

The libraries use multiple avenues to create access to collections of unique resources to incorporate into research. They are leaders in the West in digital library development and creation of high-use content. With other research libraries in the West, they are creating the Western Waters Digital Library, which contains documents and information regarding water rights, law, policies, and natural history. The Marriott is acquiring recorded natural "soundscapes" of the West that will aid the study of environments in addition to individual species. The Eccles Library has partnered in the development of the Neuro-Ophthalmology Virtual Education Library - a collection of images, video, lectures and other digital media.

Increasingly more advanced services are being requested by library users, tapping library skills such as creation, organization and description of primary research sources, interpretation of copyright law, hosting content, and creating, editing and streaming media. Users have requested computing support such as software access and training. The libraries track and employ standards for creating and preserving digital media and data. Plans are underway to create an Advanced Technology Studio at the Marriott Library to facilitate the creation of new kinds of multi-media and discussions are underway with the Digitlab about expanding support for use of Geographic Information Systems (GIS). The libraries will increase their involvement in using and developing specialized software, tools and applications for research. As data and statistics grow in importance, the libraries will acquire them and facilitate their use.

The traditional role of libraries to archive the results of research in all fields and make them accessible for the long term has been enhanced by instituting a digital archive for knowledge produced at the university - the Institutional Repository. In addition to articles, the IR will contain theses and dissertations, working papers, simulations, data sets, learning objects, images, media, data, and more. As more federal agencies will be requiring aggressive data management plans, and the IR should be a crucial piece of these plans. The libraries role also includes sharing research results through formal publication and other means. The University Press is a case in point, as is a partnership with others to develop open source software for digital publishing

These services allow faculty to integrate digital resources into their research and teaching. The IR provides a place where research results can be accessed and referenced perpetually. The libraries also offer a place for experimentation with new applications. They also are a center for information about activities across departments, an intersection between research and teaching, and a home of interdisciplinary research.

The survey shows a high demand for:

- Access to e-journals, databases and e-text;
- statistical packages and analysis;
- archiving, preservation, and dissemination of digital text, data, video, and images;
- developing and editing multimedia;

- training students to use software and work in a technology rich environment
- training students in data management, visualization, and presentation;
- access to digital resources from many places;
- implementing vocabulary standards;
- GIS support;
- equipment maintenance and trouble shooting;
- staff support for all of these activities.

Instructional support was also mentioned that included digital media, course design, and incorporation of electronic resources into course sites. Many of these services were listed under the general question of the needs that are critical to the success of their research program and the training of their students. These issues also arose in the question about their desire for centralized facilities and resources. These are services that the libraries already offer to some degree and can evolve to a new level to match contemporary computational research methods.

Committee Process

The Cyberinfrastructure Advisory Committee was formed on November 2, 2005, and has met fifteen times. The Committee invited three national leaders to spend a day on the campus:

(1) Dan Atkins, Director, Office of CI, NSF
(2) Donald Lindberg, Director of National Library of Medicine, NIH
(3) Clifford Lynch, Executive Director, Coalition of Networked Information.

During the day they met with faculty and staff from Engineering, Health Sciences, Physical Sciences, Earth Sciences, Humanities, Social Sciences, and the University Libraries. These meetings were conducted using a “town-hall” meeting format. Each of the visitors met with the Committee at the end of the day to discuss their findings. The Committee also studied reports from CI related workshops and other documents (see Appendix A). The Committee also arranged for Dan Reed to meet via the Access Grid with Senior Vice Presidents Betz, Pershing, and Associative Vice President for Research Pugmire to review what is happening with cyberinfrastructure at the University of North Carolina, Chapel Hill.

In an effort to assess the current and future needs of the University, the Committee prepared and issued an e-survey. The survey can be found at the website (http://websurveyor.net/wsb.dll/9849/CyberInfrastructure.htm). One hundred fourteen (114) responses were received from twelve different Colleges and the School of Medicine. A summary of the survey results can be found in Appendix D.

A summary of Committee findings from all of these campus visits and the faculty infrastructure survey appears below.

Committee Findings

- Cyberinfrastructure includes high performance computing in all disciplines, advanced networking services, very large scale data storage, data management, security, visualization systems and associated support for these systems. Various disciplines utilize computing in different ways, thus what is considered advanced varies across research domains.

- Multidisciplinary/Interdisciplinary education and research is a stated institutional priority, offering significant opportunities and challenges for the computational research infrastructure which has mostly developed in single discipline silos.

- Cyberinfrastructure is an essential component of institutional competitiveness.

- Cyberinfrastructure does not include commodity technologies, desktop support and software, although all of these are used daily by the same individuals who use the cyberinfrastructure components for their research.

- 90% of the Cyberinfrastructure Survey responses referred to infrastructure needs as critical for their success. The top three categories of needs were physical infrastructure, staff support and software.
  - Physical Infrastructure
  - Staff Support – Includes all levels of education/expertise/training to allow research to effectively use emerging technologies
  - Software

- Cyberinfrastructure has not been specifically considered or addressed in institutional technology planning and budgeting.

- Distributed computing is congruent with an institutional culture that values local autonomy and generates significant resources through an extraordinary level of entrepreneurial energy. However, more coordination of the distributed computing environments could limit redundancy and allow the available resources to concentrate on more advanced projects.

- While originally conceived in the context of science and engineering research, Cyberinfrastructure provides an institution-wide framework in support of advanced research and discovery. This would result in an institution-specific blend of distributed and centralized resources to fit the needs of the individual researchers and make them more competitive for research funding.

- The Center for High Performance Computing constitutes one component of essential Cyberinfrastructure, providing advanced resources and expertise in support of the research enterprise.

- The institution has a robust research community but as the Cyberinfrastructure survey demonstrates, there are real needs that should be addressed. As an example, backup and disaster recovery constitutes a critical institutional need. In a recent internal audit, the following observation was made: "We found that most of the departments within the college are not adequately storing their computer

> back-up information. Most departments are storing backups either in the same room as the computer or in the same building. We found that one department was not backing up their computers at all."

Recommendations of the Committee

The following provides additional details and specifics relating to the recommendations provided in the Executive Summary.

1 University Governance

The traditional research model of independent investigator and/or research team has not been easily incorporated in campus IT planning. However with the increasing role of multi-disciplinary and multi-institutional research initiatives, representation of the research community, development of priorities and investment in cyberinfrastructure is now an imperative. Planning, implementation and management of the institution's Cyberinfrastructure is essential for the University in the competitive research environment, the recruitment of high-quality faculty and defining the development direction of IT services for the larger institution.

1.1 Establish a Cyberinfrastructure Council chaired by Associate Vice President for Information Technology. Co-chairs of the Council will be the Assistant Vice President and Health Sciences Center Chief Information Officer and Director, Center for High Performance Computing. The chair and co-chairs will function as an executive committee for the council. The charge to the council will include:

1.1.1 provide oversight and direction for Cyberinfrastructure development;
1.1.2 approve Cyberinfrastructure components of the annual update of the Office of Information Technology's Integrate Information Technology Strategic Plan;
1.1.3 responsible for maintaining campus-wide inventory of significant computational and network resources available for research;
1.1.4 advocate Cyberinfrastructure investment.

1.2 The Council will consist of Principal Investigators on current research grants and contracts and other project leaders that rely on Cyberinfrastructure or provide Cyberinfrastructure resources/services.

1.3 Cyberinfrastructure support should be explicitly addressed in the planning and budgeting done by the Office of Information Technology and the Health Sciences Center Information Technology Services.

2 Cyberinfrastructure Support

In the Draft Report of the American Council of Learned Societies' Commission on Cyberinfrastructure for Humanities and Social Sciences, it is observed that "Humanists and social scientists have much to gain through the collaboration with technologists, possibly creating interdisciplinary labs and research groups that include both technical and subject expertise." The University should take action to pursue the ACLS's recommendation within the humanities, arts, and social sciences as well as in sciences and engineering. To facilitate growth in research-related faculty IT knowledge and skills, innovative IT outreach, training, and support personnel configurations should be considered critical and integral to the cyberinfrastructure planning and budgeting process. Currently, basic to mid-level research computing training opportunities, support staffing levels, and support staff expertise are unevenly distributed across departments, colleges, and units. For faculty and organizational units requiring advanced research computing services, CHPC has provided support for and access to staff with advanced technical expertise. This critical resource has been particularly effective in supporting network initiatives, Access Grid Development, large-scale computing services, and it functions as a critical component of the University's current and future Cyberinfrastructure.

2.1 Reconstitute the Center for High Performance Computing (CHPC) as a campus-wide Cyberinfrastructure Center (CIC) that is a user focused service provider The CIC should aggressively partner with research initiatives to partially offset operational costs. CIC IT staff will be accountable for the salary that they receive to support active research projects following appropriate policies and guidelines provided by the CI Council. CIC is well situated to promote multi-disciplinary research initiatives. Considering the previous commitments for desktop and network support to the INSCC occupants, the administration may want to re considered this free support in order to bring equality among researchers in other areas of the campus. CHPC will transition research activities to extramural funding sources over time

2.2 The Cyberinfrastructure Council will form a subcommittee including major faculty clients of the CIC to provide guidance and oversight. The Director of the CIC will be an ex officio member of the subcommittee.

2.3 Given the traditional role of libraries in supporting faculty research, the campus libraries will be charged to provide innovative basic to mid-level research-related training, support, and outreach programs should be developed to maintain and expand the IT-enhanced research productivity of faculty across lower and upper campuses.

3 Data Center and Disaster Recovery

Data storage and disaster recovery were identified in the Cyberinfrastructure Survey as critical needs by the research community. More than half of respondents indicated that they had no disaster recovery plan. The deployment of a very large scale data center addresses both an immediate need and presents an immediate opportunity to advance Cyberinfrastructure development. There is a synergy between the universal needs of disaster recovery and Cyberinfrastructure.

3.1 Develop a Utah System of Higher Education (USHE) collaborative legislative proposal for a very large scale data center, serving all USHE institutions, managed by CHPC, with libraries providing metadata support and selectively including institutional assets in the respective institutional repositories. This very large scale data repository would function as resource, archive and laboratory.

3.2 The CI council working with the OIT and campus planning should immediately initiate the planning process for fund raising, design and construction of a state of the art data center, with the goal of have the facility operational in less than four years.

4 Cyberinfrastructure Institute

The University should formulate a plan for the development of an Institute, with world-class leadership (possibly through U*), to provide campus-wide leadership, encouraging research and collaboration in disciplines exploring Cyberinfrastructure opportunities, ex. Science, Medicine, Engineering, Humanities, Architecture. The plan will identify incentives the institution will provide to encourage participation and collaboration from existing and newly established research centers (Brain Institute, Scientific Computing and Imaging Institute, Huntsman Cancer Institute, Eccles Institute of Genetics, etc). The Cyberinfrastructure Council would be responsible for the formulation and communication of this plan.

5 Computational Resources, Software, Networks and Grids

The University should develop and deploy a University Computational and Data Grid (UCDG) as the underlying architecture for its Cyberinfrastructure. The UCDG should have state of the art network connections to national and international resources such as the NSF TeraGrid, not only for gaining access to additional resources but also for encouraging collaborations and partnerships with other researchers and institutions. Major elements of the UCDG should be state-of-the-art networks, computational facilities, and extensive data repositories that are needed to meet the goals of University research priorities. Other elements may be group, department, college, or college-to-college subGRIDS for those who choose to collaborate and partner with others in meeting their Cyberinfrastructure needs or sharing resources such as computing facilities, experimental devices or sensors and the data collected from them. These subGRIDs may be connected to the UCDG to access resources not available on the subGRIDS. A principal responsibility of the Cyberinfrastructure Council will be to provide oversight for the planning, deployment and management of the UCDG.

5.1 Initiate a campus-wide planning initiative for the design and deployment of the University Computation and Data Grid (UCDG). The goal of the UCDG should be state-of-the-art networks, computational facilities and extensive data repositories, supporting multi-disciplinary, collaborative research initiatives. The UCDG should function as both infrastructure and laboratory. As a campus-wide or a statewide initiative, the UCDG will encourage investment from investigators, the institution and external funding sources.

5.2 Seek state funding to establish a state-wide Grid activity to enable all the major research Universities in Utah to collaborate and to share resources. This development effort will provide the future framework for Cyberinfrastructure for all of higher education, public education and government agencies in the State of Utah. This Grid would also allow for researchers to lead research teams throughout the US and the world

5.3 The Office of Software Licensing should survey investigators in order to determine potential site licensing opportunities that would benefit the research community.

5.4 Investments should be made in acquiring and deploying collaborative software tools and technologies, e.g., Access Grid, Content Management Software.

5.5 Develop funding proposal to the Utah State Legislature to establish a Grid program to enable all of the major research universities to collaborate and share resources.

6 Funding

As is the case for most University-wide initiatives, there is no single "silver bullet" solution to funding Cyberinfrastructure planning, deployment and management. However, there are multiple sources of support that should be explored in the development of Cyberinfrastructure.

6.1 Develop a plan for the allocation of the Indirect Cost funding to be allocated to support Cyberinfrastructure.

6.2 Tuition income formula should be revised to include support for Cyberinfrastructure.

6.3 Collaborative funding proposals with the USHE have proven to be an effective strategy with the legislature and should be pursued for system-wide investments that would contribute to the development of Cyberinfrastructure.

6.4 Utah Education Network investments should be explicitly directed toward the goals identified in the UCDG implementation plan.

6.5 Pursue extramural funding to support planning and Cyberinfrastructure development, e.g. NSF, NLM.

6.6 Funding generated by student computing fees should be accessible for investments in UCDG.

6.7 Major infrastructure investments may be made with federal ear-marked funds.

Appendix A – CI Related Reports

American Council of Learned Societies' Commission on Cyberinfrastructure for the Humanities and Social Sciences. Final Draft July 26, 2006. http://www.acls.org/cyberinfrastructure/

Building a Cyberinfrastructure for the Biological Sciences; workshop held July 14-15, 2003 http://research.calit2.net/cibio/archived/CIBIO_FINAL.pdf

CHE Cyber Chemistry Workshop; workshop held October 3-5, 2004 http://bioeng.berkeley.edu/faculty/cyber_workshop

Commission on Cyberinfrastructure for the Humanities and Social Sciences; sponsored by the American Council of Learned Societies; seven public information-gathering events held in 2004; report in preparation□http://www.acls.org/cyberinfrastructure/cyber.htm

Cyberinfrastructure for Environmental Research and Education (2003); workshop held October 30 - November 1, 2002□http://www.ncar.ucar.edu/cyber/cyberreport.pdf

CyberInfrastructure (CI) for the Integrated Solid Earth Sciences (ISES) (June 2003); workshop held on March 28-29, 2003;, June 2003□http://tectonics.geo.ku.edu/ises-ci/reports/ISES-CI_backup.pdf

Final Report: NSF SBE-CISE Workshop on Cyberinfrastructure and the Social Sciences, F. Berman and H. Brady□http://vis.sdsc.edu/sbe/reports/SBE-CISE-FINAL.pdf

Geoinformatics: Building Cyberinfrastructure for the Earth Sciences (2004); workshop held May 14 - 15, 2003; Kansas Geological Survey Report 2004-48□http://www.geoinformatics.info/

Geoscience Education and Cyberinfrastructure, Digital Library for Earth System Education, (2004); workshop held April 19-20, 2004□http://www.dlese.org/documents/reports/GeoEd-CI.pdf

Identifying Major Scientific Challenges in the Mathematical and Physical Sciences and their CyberInfrastructure Needs, workshop held April 21,2004 http://www.nsf.gov/attachments/100811/public/CyberscienceFinal4.pdf

IT Engagement in Research. Roadmap. EDUCAUSE Center for Applied Research. July 2006. http://www.educause.edu/ir/library/pdf/ECAR_SO/ers/ers0605/ECM0605.pdf

Materials Research Cyberscience enabled by Cyberinfrastructure; workshop held June 17 - 19, 2004□http://www.nsf.gov/mps/dmr/csci.pdf

An Operations Cyberinfrastructure: Using Cyberinfrastructure and Operations Research to Improve Productivity in American Enterprises"; workshop held August 30 - 31, 2004 http://www.optimization-online.org/OCI/OCI.pdf

Cyberinfrastructure for Education and Learning for the Future: a Vision and Cyberinfrastructure for Education and Learning for the Future: a Vision and Research Agenda (170 KB PDF).**Research Agenda** (170 KB PDF).

Appendix B Notes on the Grid

Taken from the Gridcafe website http://gridcafe.web.cern.ch/gridcafe/

What is the Grid? One answer is that, whereas the Web is a service for sharing information over the Internet, the Grid is a **service for sharing computer power and data storage** capacity over the Internet. The Grid addresses needs such as:
Ten years ago, **biologists** were happy if they could simulate a single small molecule on a computer, now they want to simulate thousands of molecular drug candidates to see how they would interact with specific proteins.***Earth scientists*** keep track of the level of atmospheric ozone with satellite observations. For this task alone, they download, from space to ground, about 100 Gigabytes of raw images per day.

Unlocking the secrets of the **human genome** would be impossible without the computerized analysis of massive amounts of data, including the sequence of the three billion chemical units that comprise our DNA, which is the genetic blueprint of our species.

There are perhaps five big ideas behind the Grid, none of them being unique in this respect: The sharing of resources on a global scale is the very essence of the Grid. Ssecurity is a critical aspect of the Grid, since there must be a very high level of **trust** between remote resource providers and users. If the resources can be shared securely, then the Grid really starts to pay off when it can balance the load on the resources, so that computers everywhere are used **more efficiently**, and queues for access to advanced computing resources can be shortened. For this to work, however, communications networks have to ensure that distance no longer matters – on a global scale.

Finally, there is the issue of open standards, which are needed in order to make sure that R&D worldwide can contribute in a constructive way to the development of the Grid, and that industry will be prepared to invest in developing **commercial Grid services** and **infrastructure. There are hundreds of grid projects going on at the moment in a number of areas:**

- Grid-tech Projects - primarily involved in development of Grid-enabling technology, such as middleware and hardware
- Testbeds Projects - devoted to developing and maintaining a working testbeds using existing Grid technology
- Field-specific applications - projects devoted to explore and harness grid technology in the context of specific fields of scientific research
- Grid Fora Projects - devoted to catalyze, stimulate and foster collaboration on grid related projects
- Grid Portals - Internet portals to grid related activities
- Commercial Grid initiatives - Grid solutions and initiatives by commercial vendors
- ...@home - distributed computing projects Internet computing projects
- Grid Outreach initiatives - educational and informative websites on Grid computing
- Grid Consulting companies See http://gridcafe.web.cern.ch/gridcafe/gridprojects/projects.html

Appendix C - Current Campus CI Organizations

The Introduction above summarizes background information on 3 campus organizations engaged in meeting general University research cyberinfrastucture needs, namely the Center for High Performance Computing Center (CHPC), the Office of Information Technology (OIT), and the Health Sciences Department of Information Technology Services (ITS). In this Appendix we present more detail information on the activities of these organizations.

CHPC

CHPC activities can be categorized into 4 main areas (1) Large Scale Computing (LCS), Advanced Networks (AN), Visualization Lab and INSCC AV, and INSCC Networking and Desktop Support.

Large Scale Computing requires approximately 50 % of CHPC's FTE effort. It includes operating and maintaining the parallel computing systems Arches (Opteron64), ICEBox (I32), and Sierra (COMPAQ). It also provides Statistical Servers, a BLAST server , SEQUEST Cluster Server and an NMR Analysis System for approximately 200 students. The architecture of the most heavily used system, the ARCHES meta-cluster, is described in **Appendix E**. Note that during 2005 thirty-three faculty had accounts on one or more of the above systems. In the last 5 years more than 172 researchers have acknowledge the contribution of CHPC in their published papers. Faculty users and their usage are listed in **Appendix F**.

Advanced Networks requires approximately 5 % of FTE effort. It includes providing OC12 to Internet2, Access Grid for teleconferencing at INSCC, Eccles Library and the New Media Wing. In addition, it coordinates R&D for OIT, including IPV6 deployment, multicast deployment, the wireless working group, and optical networks.

The Visualization Lab and INSCC AV require approximately 5 % FTE effort. This includes operating and maintaining the new 3D visualization wall and editing facilities, production of videos, posters, etc., technical support for the INSCC AV, testing of video technologies for campus including Eccles Library and the new Media Wing Access GRID, the Art and Technology Telematic Projects. It participated in the design of the new Medical Education Video Servers of the new Medical Education Video System.

INSCC Networking and Desktop Support require approximately 30 % FTE effort. This includes operating, maintaining and upgrading INSCC networks with full service to wall plates (~ 600 connections), providing e-mail for most people in INSCC, maintaining 200 desktops systems, 160 of them for research groups in INSCC, 30 file servers with total backups of approximately 30 Tbytes, teleconference facilities, and group compute servers. It is also responsible for the physical plant of INSCC. Ten different research groups in INSCC take advantage of these services. These research groups are listed in **Appendix G**.

CHPC's Bioinformatics Initiative - As noted in the Background section, CHPC took steps to implement one of the majors recommendations in the 2000 Strategic Plan through it's Bioinformatics Initiative. These included collaboration with Genetic Epidemiology to develop scalable parallel software, developing a SEQUEST Cluster for Proteomics, participation in several Bioinformatic Planning committees and co-PI (Julio Facelli) in several NIH proposals with one funded seed grant (JCF), and development of a BLAST cluster.

Office of Information Technology (OIT)

OIT is organized into 8 departments that report to the Associate VP of Information Technology.
They are charged with maintaining the IT infrastructure and ensuring the accessibility of core IT resources. They are:
Network and Communication Services (NetCom) - phones, networks and cable tv services
Information Security Office - network security: audits, incident reporting, network monitoring
IT Architecture - campus-wide IT project research, design & support
IT Systems - web hosting, DNS, email systems maintenance and support
Instructional Media Services - classroom media equipment and services
Office of Software Licensing - affordable software for campus & home use
Media Solutions - websites, videos, and multimedia services
U Webmaster - resources for campus webmasters, oversight of the U home page

OIT policy is developed when necessary to ensure compliance with laws, regulations and best practices, or to protect the assets of the University, including its people. OIT policies will empower, not deter, the adoption of new technologies and the development of centrally provided and distributed client services. Information Technology policies are developed to mesh seamlessly with official University policies.
Plans are developed based on the ability of OIT to:
assess the needs of the campus community,
develop solutions to those needs that have broad campus support,
justify the plan based on sound business cases,
define project plans that will succeed, and
communicate the solutions and services to the campus community to facilitate adoption.
Evaluation of plans and resulting projects takes place at several steps in the process, not the least of which is the determination of end-user satisfaction with the results.

The Information Technology Council (ITC), as authorized by the Senior Academic Vice President, is the legislative driver of IT policies and plans. It's purpose is to facilitate the development of the University's Information Technology and e-Commerce infrastructures, resources, and applications. The ITC is comprised of members from most colleges and administrative departments. The ITC receives technical advice from the **Information Technology Advisory Council (ITAC)**. Its purpose is to advise the Office of Information Technology, ITC and Campus IT managers on technical issues that have campus-wide impact. It is responsible for recommending allocation of scarce core IT resources and recommends the direction of core technology implementations.

The October 10, 2005 Integrated Information Technology Strategic Plan developed by the ITC can be found at
www.it.utah.edu/leadership/policies/Campus_Strategic_Plan10102005.pdf.

Health Sciences Department of Information Technology Services (ITS)

ITS's role is to advance Health Sciences Center goals through quality information technology services and resources. The goals are met by implementing action items in the IT Strategic Plan that were developed by over 40 stakeholders from various HSC missions in a series of meetings held from January to May, 2001. The Plan has a set of objectives :

Develop an information technology infrastructure that will enhance clinical access and streamline clinical process

Improve clinical documentation tools

Implement the Orders Entry and Decision Support functions of the EMR to improve clinical outcomes project

Fully implement the Data Warehouse and associated query tools

Enhance educational offerings through use of information technology

Provide the technical assistance and infrastructure required to offer high quality education programs

Coordinate investments in support of education

Establish benchmarks and evaluate the impact of technology

Coordinate database applications and development with Main Campus

Provide Electronic Research Administration at increase research revenue by improving the administrative processes of identifying, applying for and managing grants

Provide a "Research-Enabling" Network Infrastructure Strategically Manage Information Use Integrate Research into the Data Warehouse

Enhance enterprise-wide information technology systems

Promote web-enabled systems Streamline services through electronic transactions

Improve administrative management through increased information accessibility

Establish state-of-the-art IT healthcare application benchmarks to assist HSC leadership with enterprise-wide resource planning

Provide a secure, yet open and network architecture to create an environment that will facilitate the missions of the Health Sciences Center

The action items for each of the above objectives and their state of implementation can be found at IT Strategic Plan.

ITS organizational areas are

Business Services/Administration

Clinical Information Services

Data Resource Center

Financial and Ancillary Information Systems

Information Security and Privacy

Network Operations

Utah Telehealth Network/Telemedicine

Web Resource Center and Customer Services .

ITS's organizational chart can be found at http://uuhsc.utah.edu/its/orgchart/.

Appendix D – Summary of Survey Results

CURRENT AND FUTURE NEEDS

1. Identify perceived cyber-infrastructure needs and specify the ones that are critical for the success of your research program and the training of your students.
RESULTS: The top three categories of needs critical for success were physical infrastructure, software, and staff support. About 90% of the responses referred to physical infrastructure needs as critical for their success. The responses were reviewed and categorized into the top five categories according to the number of times an item was mentioned. The summary follows.

1. Physical infrastructure (96)
    - Networks (33)
    - Storage (24)
    - Cluster (10)
    - Servers (10)
    - Other: data center, grid, PCs, videoconferencing, video, handheld devices
2. Software (28)
    - Email (8)
    - Collaboration (5)
    - Database warehouse (5)
    - Programming (3)
    - Other: student software, bio informatics, CAD, collaboration with other universities, software purchases, instructional, information simulation, search.
3. Staff support (15)
    - Statistical analysis (5)
    - Training (4)
    - Video, survey, electronics (2)
    - Other: training, bio informatics, cluster, desktop support, security
4. Connection to digital library resources (11)
5. Back-ups (3)

2. Identify the top three infrastructure needs of your research that could be provided by <u>centralized</u> facilities/resources.
RESULTS: The top three categories of needs that could be provided centrally were physical infrastructure, staff support, and software. About 50% of the responses referred to physical infrastructure needs as critical for their success; about 40% listed staff support. The responses were reviewed and categorized according to the number of times an item was mentioned. The summary follows.

1. Physical infrastructure (59)
    - Networks (20)
    - Storage (14)
    - Cluster (11)
    - Wireless (5)
    - Other: servers, data center, computer upgrades, PCs, AV equipment, printing
2. Staff support (46)
    - More staff knowledgeable in software and hardware (8)
    - Training (7)
    - Programming (6)
    - System administration (4)

Other: security, database, more staff, hyper speed internet, web, backup, informatics, survey help, statistics help, PC/Macs, workstations, vocabulary standard, GIS tech, grant requirements and accounting.

3. Software (28)
    - Email/FTP (5)
    - Database warehouse (5)
    - Statistics (3)
    - Collaboration (2)
    - Other: student software, system server, staff software, searches, NATLAB, implicit/explicit tools, mesh generations tools, data analysis, firewall
4. Backups and remote backups (13)
5. Digital library (2)

3. Identify the top three distributive services needs of your research that could be provided by centralized facilities/resources.
RESULTS: There was a lot of confusion with this question; Twenty-nine respondents said they weren't sure or didn't know what distributive services were. Other responses included portals and access/storage and retrieval, networking, and parallel computing.

UNDERLYING DETAILS

Data access and storage

1. How are your data access and storage needs currently being met?
107 responses
Desktop (45)
Servers (40)
External media (11)

2. In meeting your data requirements what are the limiting factors? (See Figure 1.)
Almost half of the respondents selected STORAGE CAPACITY as a limiting factor. Transferring data, data management software/frameworks, and cost were also listed as most limiting factors by more than one third of the respondents. The tabulated results are as follows:
Storage capacity (55)
Transferring data from storage to desktop or cluster (38)
Data management software/frameworks (38)
Cost (37)
Data privacy/security requirements (30)
Transferring experimental data to storage facility (28)
Software compatibility (22)
Access to national repositories (21)
Data/format compatibility (21)
Data integrity (17)
Other (15) (5 responded as having no limits; other factors included backup costs, secure/speed/fidelity of transfer, cheap storage, technical support)
Lack of data in digital form (12)

3a. What is your current disaster recovery plan?
Other (38)
Informal plan (21)
RAID (13)

External tape (10)
Mirror site (9)
Tape (7)
Mirror site – real time (1)
The other (38) category included 12 respondents who reported their plan as none, unknown, and even "prayer." Other responses also included backups to CD, DVDs, optical form, and combinations of RAID, tapes, external hard drives, etc.

3b. What is your future disaster recovery plan?
Other (31)
Informal plan (13)
RAID (12)
Mirror site (11)
External tape (9)
Mirror site – real time (4)
Tape (3)
More than half of the other responses included none, unsure, or unknown and "pray harder." Other responses also included "same as our current plan" and "we need a plan" and external drives (RAID, LaCie, and network backups.

4. What are your greatest data access and storage needs?
About one third of responses referred to large data sets or specific amounts of storage space needed, ranging from 1 TB to 10 petabytes. Room for multimedia files (video, audio, electronic lab books, maps, images, etc.) was also listed in 15% of the responses. Accessibility was an issue (off campus, math server, national software centers, centralized location to share across other university assets) in 15% of the comments. People also mentioned data loss and recovery, speed or performance, knowledge and training, and safety and security.

5. Estimate your current and future storage requirements.
Most people have 10-99 GB right now and anticipate needing 2-100 TB in the future.

Size	Current	Future
10-99 GB	40	19
100 GB – 1 TB	33	31
2-100 TB	26	41
> 100 TB		8
Other	8	7

Software
1. What software barriers do you encounter? (See Figure 2)
More than half listed costs and upgrades as their greatest insufficiencies. Other problems included software incompatibility, accessibility, and incompatibility.
Software costs (62)
Software upgrades (60)
Software incompatibility (31)
Software accessibility (25)
Software portability (24)
other (13).

Other included installation, software support, software development, having to pay for uprades by myself, low software quality, time to train on new software. Comments included problems like needing software from a previous project that is currently unavailable, waiting for an administrator to install from my desktop, writing our own software, and multiple operating systems.

2. What are your greatest software needs?
83 responses
Discipline-specific programs (28)
Statistics (20)
Database and DB management (19)
Repositories, collaboration tools, s/w development tools and environment, compilers, visualization software (11)
Support (Mac, OSL, Linux, Office, PDA) (5)

Networking

1. Where are the perceived networking bottlenecks?
Within your department/bldg (27)
Within your college (21)
Exterior to your dept/college but within the university (21)
Other (15)
Security requirements (14)
Within the region/state (10)
With national connections (9)
With international connections (5)
Of the other responses, half did not know where the bottleneck is;
7 don't know and 3 say there isn't a bottle neck; other bottlenecks mentioned include the firewalls at HSC and Hospital, problems with big databases and concerns for constant security attacks.

2. What are your greatest networking needs?
The top 3 needs mentioned were fast connections and transfers, reliability, and wireless networking. Respondents also mentioned needs for specific links between labs, university and national networks and between certain buildings and labs here (such as PCMC and the University or INSCC and SP and JFB) were needed. Other responses included being able to videoconference beyond the firewall, accessing very large files on the server, a desire to work more effectively with student records, and a need for 1-10 gigabit/s on every desk.

3. Estimate your current and future bandwidth requirements.
61 responses
One third of the respondents did not feel comfortable making this estimate. 12% said their current situation was fine. The low end of the estimates ranged from 10-100 megabits. About a third of the respondents expressed a need for at least a gigabit connection, with the high end at 200 gigabit connections needed.

Computing Hardware

1. How are the computing needs for your research being met? (See Figure 4)
Desktop system (83)
Group or individual cluster (28)
Department owned cluster (20)

College cluster (19)
CHPC cluster (16)
National systems (15)
Other campus systems (13)
Other off-campus systems (7)

2. What are some of the systems you use?
Several hundred were listed, including: CHPC clusters, National Center for Atmospheric Research, SCI Institute clusters (inferno) Los Alamos, Livermore machines, MACs and PCs,,unix.fcs.utah.edu, Various NIH-sponsored tools, BLAST etc., NCAR/UCAR, Maui system, Berkley system, GFDL system, College of Mines and Earth Sciences unix boxes, GEON Server, OTSS within the college of ed., UUHSC ITS systems - PACS, EDW. NLM Medline, NSF Teragrid, NSF PSC, SDSC DOE BNL QCDOC (SciDAC) DOE NERSC, office desktop, web based genomics software, Math, NERSC, SOC and research group machines and clusters (various SGI Altix's in SCI, the Corvus cluster in SoC, etc.), CADE lab linux cluster. ITS, Uhosp applications, NCBI server (national), Wormbase (national) Blast, google, gene sifter, pub med, OMIM. fluorescence microscopy core, databases in Santa Cruz/NCBI/ENSEMBL, Pfam Wulfpack nodes (St. Louis), Cardiovascular Genetics, Eccles Med Library for electronic journals. PubMed, our own computer facilities within the Utah Center for Advanced Imaging Research UCAIR, C-SAFE cluster (inferno) for C-SAFE SCI clusters (muse, ray) C-SAFE LLNL Linux clusters (ALC, Thunder, Purple) C-SAFE Wharton Unix machines, HMBG, SBCC Structural Biology Computing Center in Biochemistry, VA, ASCI platforms at: Los Alamos National Laboratory Sandia National Laboratory Livermore National Laboratory, HCI, Laurie McMillan, NASA supercomputer, National Network of Libraries of Medicine located at the University of Washington University of Utah Washington University, Sequence analysis programs (like Clustal W) provided at various websites. Most of our computing is small-scale and performed on desktops; College of Nursing Open Access Student Computer Lab, Health Sciences Campus, HSEB Student Computer Labs, systems in foreign countries where the databases reside (Russia, Germany) JPL Supercomputining (astro-theory), LRAC (large resource allocation committee, NSF centers, NCSA/PSC), all NSF sites and some DOD sites.

2. What are your greatest computing hardware needs?
The needs reported seemed to vary greatly, but more power (faster machines, more RAM, more storage, faster connections, more processing power) was mentioned most frequently. This was an expressed need for desktops as well as servers, clustering and networking. They also wanted their desktops and laptops to be more current and to have a way for regular hardware and software updates. Another hardware need was the capacity to handle and serve multimedia (video server storage, 3D projects and other visualization projects).

Staff Support

1. What are your greatest staff needs? (See Figure 5)
Software maintenance (61)
Desktop maintenance (48)
IT administrator (43)
Hardware maintenance (42)
Program development (37)
Software parallelization (14)
Other (12)

Porting codes (7)

2. What is the size of your support staff?
The average size reported was 3, with 235 people being identified as staff support.

3. Do you include staff support in your research request?
No (59)
Yes (42)
Comments were added by 14 respondents; nine people said that staff was not likely to be funded (and would be inappropriate to ask) or that the staff was not needed in the research request.

Users
1. Indicate the number of users included in your response.
Faculty (77)
Post Docs (159)
Graduate students (832)
Research staff (143)
Undergraduates (15,000)
Total: 16,211

Estimate of future costs
1. Please estimate future costs for your departments' cyberinfrastructure needs; include possible funding sources.
52 responses of 113 respondents (12 gave no dollar figure)
A total of about $3M was estimated. Some of the possible funding sources identified included, NSF, DOE, NOAA, grants, student fees, College, F/A, corporate and NIH grants, None, DOD, return of indirect costs, NASA, NIH R-01, P-01, the ususal federal agencies,

Responses by college/school:

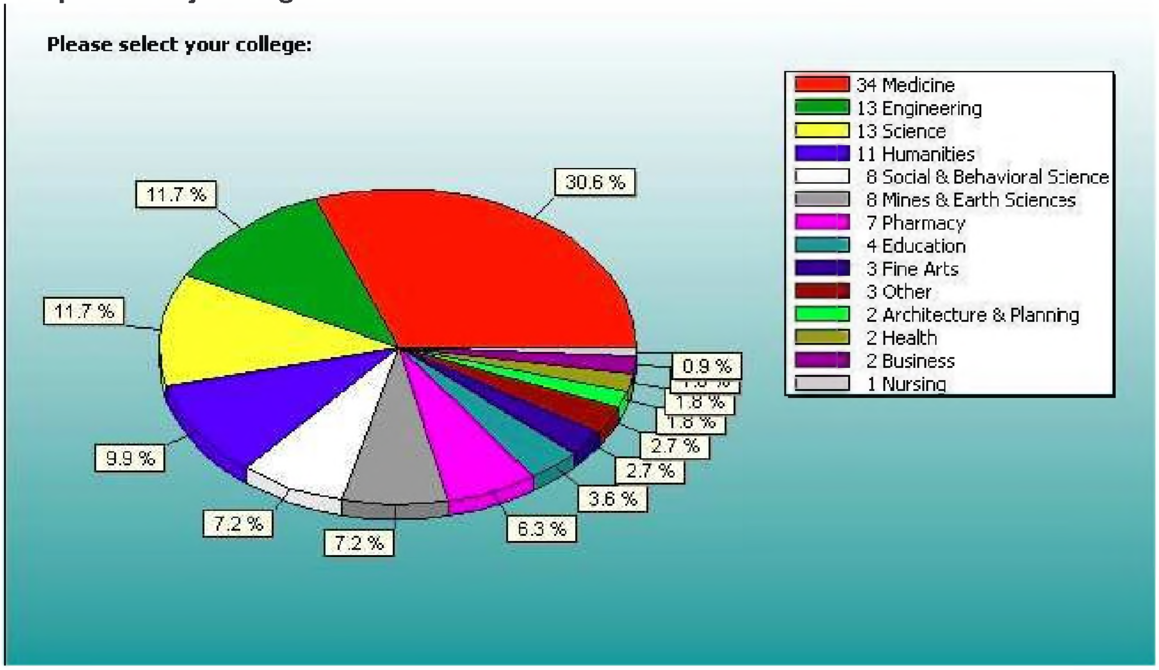


Figure 1: Limiting Factors for Data Requirements

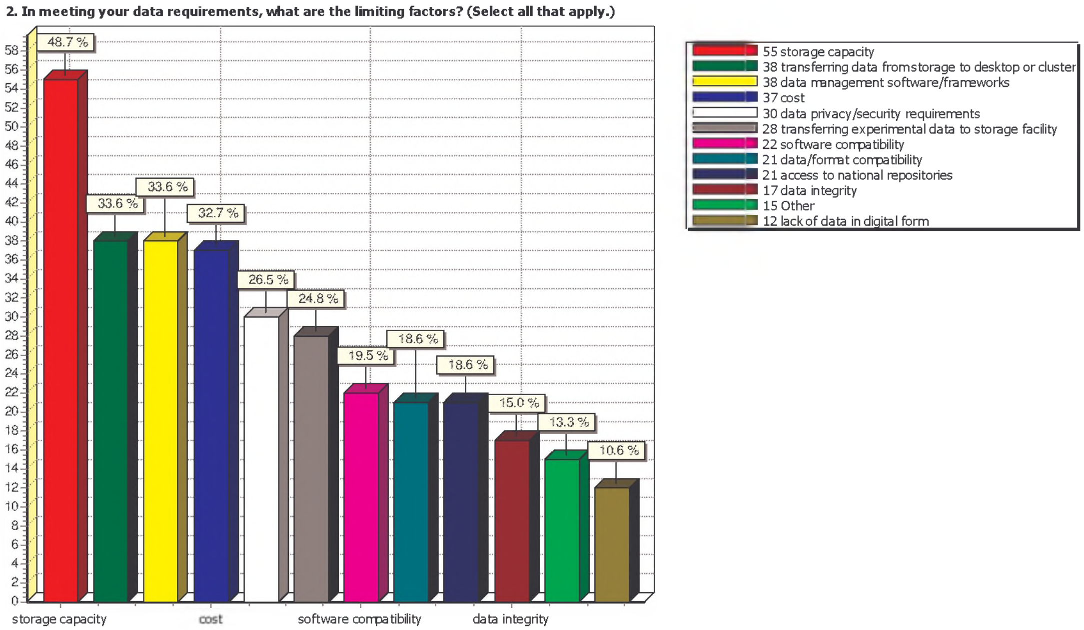


Figure 2 Software Insufficiencies

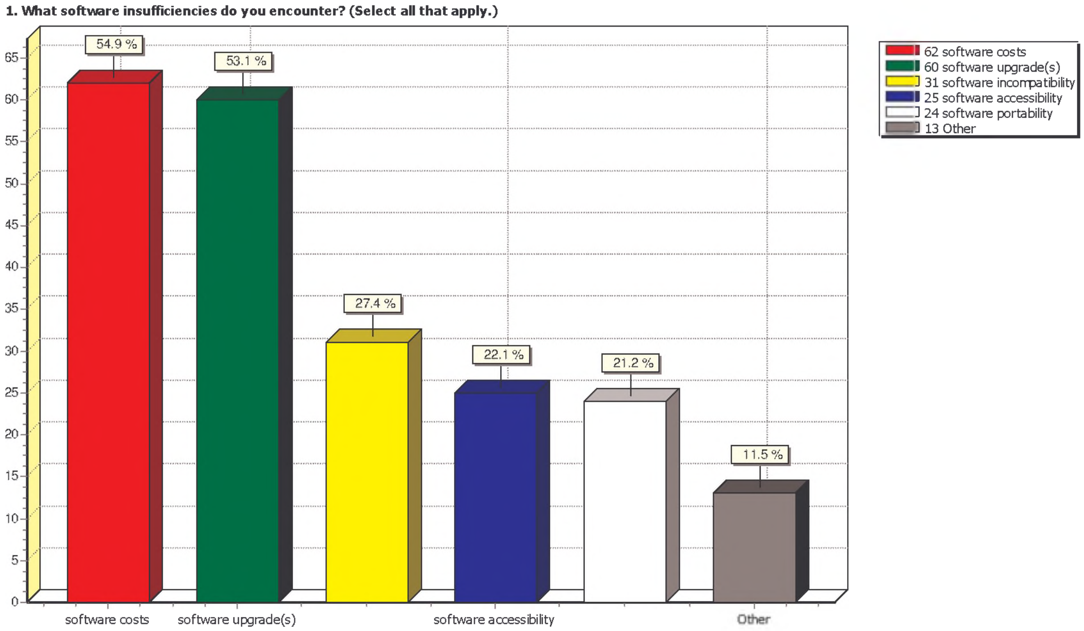


Figure 3 Perceived Networking Bottlenecks

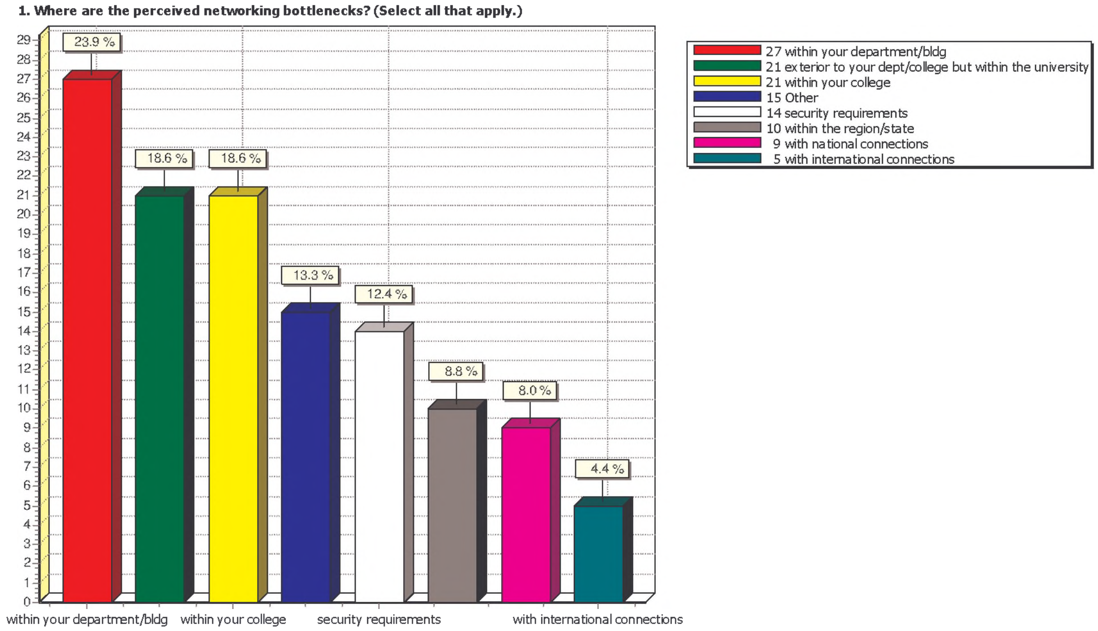


Figure 4 How Research Computing Needs Are Met

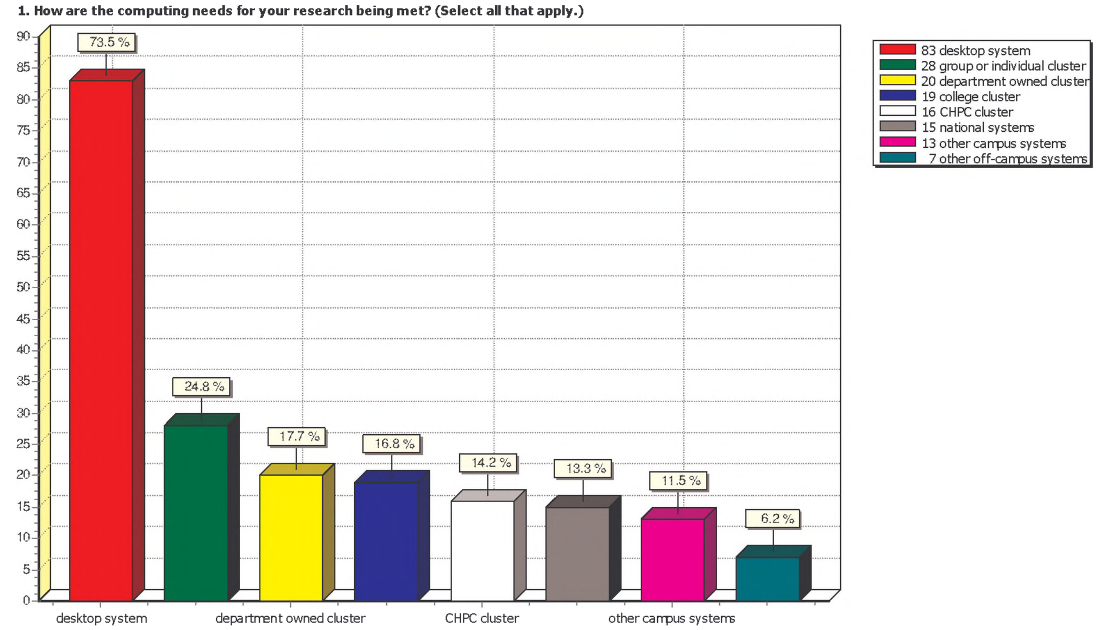
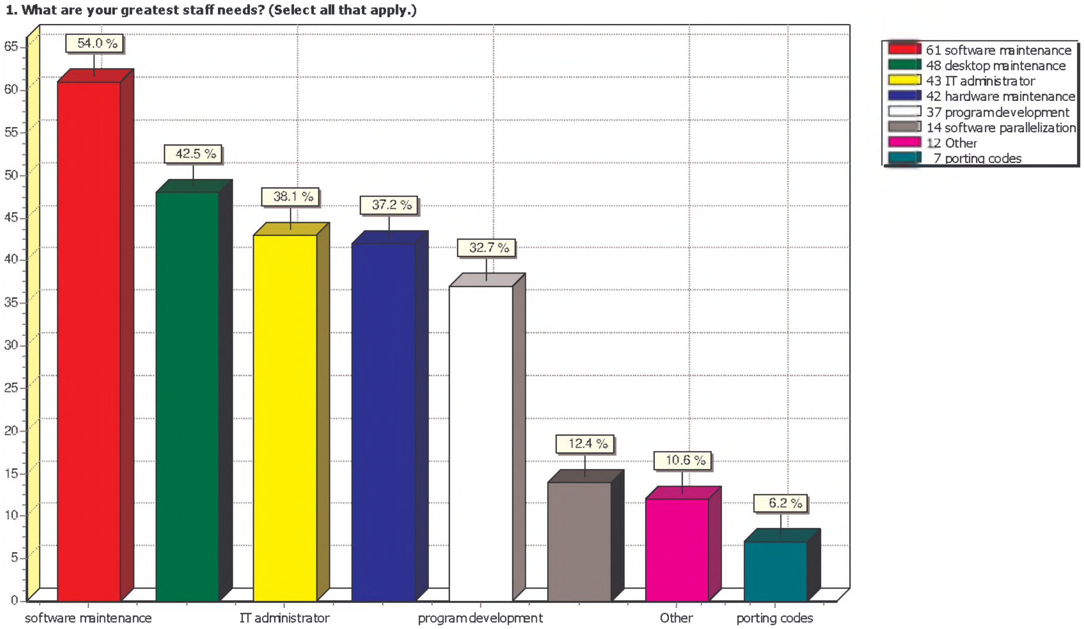


Figure 5 Greatest Staff Needs



Appendix E – Arches meta-cluster Architecture (1.4-2.0 GHz OPTERON CPUs)

DA: 256 dual nodes, 2 Gbytes connected by Myrinet
MM: 184 dual nodes, 2 Gbytes connected by GigE
TA: 48 dual nodes, 4 Gbytes connected by GigE
LA: Condominium style cluster funded by research funds from Voth, Schuster, Liu, Zhdanov, and Simons.

Appendix F – Arches Usage in 2005

Gregory A. Voth(voth)	4,278,097
Thomas Cheatham(cheatham)	2,350,408
Julio C. Facelli(facelli)	712,759
Feng Liu(liu)	622,273
Thanh Truong(truong)	360,277
Carleton DeTar(detar)	307,464
Jeff Weiss(weissj)	159,652
Phil Smith(smithp)	101,179
Joel S. Miller(millerjs)	93,889
David Grant(grant)	82,013
Peter B. Armentrout(armentro)	77,857
CHPC(chpc)	69,137
Thomas Reichler(reichler)	45,956
G. B. Stringfellow(stringfe)	41,986
Jack Simons(simons)	36,483
Gerard Schuster(schuster)	30,004
Grant Smith(smithg)	25,139
Michael Zhdanov(zhdanov)	17,626
Chris Ireland(ireland)	16,910
Alejandro Sanchez(sanchez)	7,892
Zhaoxia Pu(zpu)	7,850
Mary Ann Jenkins(jenkins)	4,325
Raymond F. Gesteland(gestelan)	2,944
Jon Rainier(rainier)	2,602

Fred Adler(adler)	1,802
Aaron Fogelson(fogelson)	1,366
Michael D. Morse(morse)	1,342
Chris Hill(hill)	767
Ilya Zharov(zharov)	545
Cuiye Chen(cchen)	411
Edward Zipser(zipser)	17
Charlie Jui(jui)	11
Cynthia Furse(furse)	5
Ed Trujillo(trujillo)	0
Total SU's	**9,460,987**

Appendix G – Research Groups in INSCC

Laser Institute.

Cosmic Rays: HiRes, Auger, Veritas.

CROMDI (Center for the Representation of Multi-Dimensional Information).

High Energy Physics Group.

CSEO (Computational Science and Engineering on Line).

CRSIM (Combustion and Reaction Simulations).

Center for Biophysical Modeling and Simulation.

CIRP (Cooperative Institute for Regional Weather Prediction).

UTAM (Utah Tomography and Modeling/Migration Consortium).

Appendix H – Utah Cyber Infrastructure Plan (DRAFT)

Importance of Cyber Infrastructure for 21st Century Science and Technology

Computational and network resources are a critical component of the modern research infrastructure and economic development. This has been recently recognized by the National Science Foundation (NSF) in the Cyber Infrastructure report (http://www.cise.nsf.gov/sci/reports/toc.cfm), describing how advances realized in information technology over the last two decades will create new paradigms for scientific research and engineering by integrating experimental and simulation approaches to scientific discovery and engineering design. The importance that the NSF is giving to cyber infrastructure becomes apparent when realizing the NSF has created a new office, reporting to its director, to lead the deployment of a pervasive cyber infrastructure for the US research enterprise (http://www.nsf.gov/div/index.jsp?div=OCI). As more researchers become dependent on advance information technology resources to acquire, analyze and simulate their data, the broad deployment of data repositories and computational facilities integrated by high performance networks will define the research and engineering environments of the 21st century. While the National Science Foundation is developing the guiding principles for the establishing the National Cyber Infrastructure, many States are making significant investments in cyberinfrastructure to enhance their competitiveness to attract research and foster economic development based on the emerging enterprises that develop products and services derived from academic research.

The development and deployment of cyber infrastructure can be effectively accomplished by deploying computational and data GRIDS, which as their electric counterparts, promise pervasive access to information and simulation resources needed for the modern research enterprise. Three key elements are necessary for the deployment of computational GRIDS: state of the art networks, computational facilities and extensive data repositories. A detailed review of the emerging modalities for performing science in the 21st century has been presented in a recent Science article by Ian Foster (http://www.sciencemag.org/cgi/content/short/308/5723/814),describing how remote access to disparate instruments and simulation platforms will make science a global enterprise.

The State of Utah has been a pioneer in state networks and high performance computing. UEN (Utah Educational Network) is an exemplar on the deployment of shared network infrastructure in support of education and research across the state. CHPC (Center for High Performance Computing) is one of the leaders among the state high performance computer centers (http://www.ncsc.org/casc/index.html). Recently Utah State University has also created the new center for high performance computing, recognizing the importance of this activity in support of the modern research enterprise. These three organizations working in a close partnership have the technical expertise required to deploy a statewide computational GRID, but they will need additional resources from the State of Utah.

In order to support a statewide grid successfully, the State must decide now to makesignificant investments in the four critical components needed to support its cyber infrastructure.These components are: data centers, optical networks based on University leased fiber, advancecomputational facilities and data repositories.*Economic Development implications of cyber infrastructure:* State, Education and Research leaders recognized the economic development implications of scientific research many years ago. Research centers like Silicon Valley have been economic engines for the country and region. Recently, the work done in the Council of Competitiveness (http://www.compete.org/hpc/) has strongly demonstrated, in greater detail, the growing importance of high performance computing and advanced networking for maintaining a vibrant economy. The new research modalities used in science today require cyber infrastructure support for the simulations, which nowadays are made possible through large scale data analysis and advance network applications. These methods have not only transformed science but also the design and engineering process for launching new products into the market place. For example, auto manufacturers now simulate collisions on high performance computers, saving millions of dollars in development costs (40%) and substantially shortening design cycle times. The fuel of the new economy is new technology with university trained personnel bringing new and improved products to market. In a Gartner study completed for the state of California it was shown that increased network capacity and connectivity can have a significant impact on increasing the domestic product per capita. Providing research centers, with broadband connectivity, cyber infrastructure and university trained people will speed Utah in achieving scientific and economic goals. This reality has not escaped the attention of many other states in the nation and elsewhere. A brief list of selected state based cyber infrastructure deployment in support of research as an engine for economic development is:

Ohio: (http://www.osc.edu/oarnet/)

SURA: (http://www2.gsu.edu/~wwwacs/suragridconf/).

Louisiana: (http://www.lsu.edu/highlights/051/loni.html/).

In the following we discuss recent developments in four key cyber infrastructure components, optical networks, high performance computing facilities, data storage repositories and data centers, and define appropriate action items necessary in the short term to start the development of a comprehensive cyber infrastructure plan for the state of Utah.

Optical Networks:

Research institutions or regional academic networks have been steadily aggregating into what are commonly known as GigaPops. These GigaPops have started to obtain long term IRU (irrevocable rights to use) of both metropolitan and long haul optical fiber plants formerly or currently owned by private carriers. These GigaPops have started to utilize this private fiber to connect various entities for research based needs in advanced networking and cyberinfrastructure. The term Regional Optical Networks (RONs) describes these build-outs of private fiber infrastructure. By utilizing equipment that multiplexes various light frequencies on the same pair of fiber, these RONs are able to create multiple high-bandwidth connections with traditional or experimental protocols. The need for these new types of facilities has been clearly demonstrated, for instance, in the recent paper by Corbató and Cotter (http://www.educause.edu/apps/er/erm05/erm0538.asp), the CENIC planning reports (http://www.slac.stanford.edu/grp/scs/trip/cottrell-cenic-may02.html) and Richard Katz EDUCAUSE report (http://www.educause.edu/LibraryDetailPage/666?ID=ERM0547) among many others. The importance that States are giving to this new type of regional optical networks can be realized by the cursory inspection of the map bellow, where the states in which optical networks based on IRUs have been deployed are colored in red.

While the technical details of RONs are well beyond the scope of this paper, perhaps we can provide an example on how these networks can impact research. For optical networks that are deployed by research entities the marginal cost of provisioning additional dedicated high bandwidth for a particular application (a dedicated lambda using the RON's jargon) is quite low once that the infrastructure has been deployed. Therefore it is possible to build, on demand and for relatively short period of time, self contained networks that researchers can use for transmitting large amounts of data or executing high end simulations using remote distributed computer resources. An example of this emerging trend of network usage by real scientific problems can be found in the NSF TeraGRID projects. These projects support improved storm forecast capability (http://www.teragrid.org/news/news05/0705.html), seismic modeling and oil reservoir simulations (http://www.teragrid.org/news/news05/seismic_model.html) as well as computational nanotechnology (http://www.teragrid.org/news/news05/nanohub.html).

Optical Network for the State of Utah:

In order to develop the necessary research cyber infrastructure, UEN will have to provide, at a minimum, redundant optical network connectivity between the three major research Universities in the State (UofU, USU and BYU). UEN should provide this connectivity via extended IRUs of fiber and via UEN owed/operated optical electronics. The fiber and optronics allow the provisioning of additional services on demand that projects such as the Hybrid Optical and Packet Infrastructure Project, (http://networks.internet2.edu/hopi/), are developing. Note that, due their experimental nature, optical networks on demand are not services that commercial providers will offer for many years to come and it is imperative that they are provided by UEN for use of the research community. Depending on design requirements and participation, UEN can connect the remaining Universities and Colleges in the system as spurs of the Utah Optical Network or as fully redundant nodes. UEN should establish additional connectivity between the University of Utah and international Cosmic Ray Observatory site in Millard County to provide high end network connectivity for this world class research facility.

Actions:

☐ UEN and CHPC will work on securing an IRU between the UofU campus and Hinckley (location of the Cosmic Ray Observatory) using the ATT fiber donated to SURA.☐ UEN and CHPC will issue a series of RFIs in order to carefully assess the availability and cost of the IRUs necessary to construct the first phase (R1 institutions) and second Phase (remaining Colleges and Universities) of the Utah Optical Network.
☐ UEN will develop a plan for incremental deployment of the necessary optical equipment to operate the Utah Optical Network.
☐ The cyber infrastructure planning committee will brief the Utah congressional delegation on the special challenges that we face in deploying RONs in the intermountain region. Note that a similar initiative is being carried on by the northern tier consortium (http://www.ntnc.org/default.htm), which represents the northern states of the US, which are facing similar challenges.

High performance Computing Facilities:

Large distributed systems provide the increased level of performance that HPC facilities require in today's computational environment for simulations. These systems encompass top national facilities, regional facilities and local facilities. In general, the cost, complexity and performance of these systems decrease by an order of magnitude for each category. The researchers in the State of Utah can make use of the national facilities by utilizing the local networks, the networks that link our Universities, and the research networks that link with the different national centers. The National Science Foundation (NSF), Department of Energy (DoE), National Aeronautics and Space Administration (NASA) and the Department of Defense (DoD) are some of the entities that manage the different national centers. The State of Utah must develop a sustainable plan to provide regional access to HPC facilities for a much broader community, including industry in need of simulation sciences support. An example on how such access can be structured can be found in the very successful Cluster Ohio project (http://www.osc.edu/hpc/cluster_ohio/). With the support of the State of Ohio OSC (Ohio Supercomputer Center) has developed a hierarchical and distributed system of advanced computational and simulation resources, by which Ohio researchers and engineers, in public, private and commercial entities have access to the most advanced simulation tools.

Typically, due to the rapid technology changes, HPC facilities tend to last 3 years before becoming obsolete. National caliber systems cost between 20M$ to 40M$, while regional facilities cost 10 times less and local facilities 100 times less. Following this model we propose to develop a HCP infrastructure that will locate regional size facilities at both the University of Utah and Utah State University and local facilities at the remaining institutions in the Utah System of Higher Education. Institutions receiving these systems will be responsible for their operation, will coordinate their operation, access and usage policies by all the participants in the Utah GRID, and establish outreach and educational programs to facilitate access to the HPC facilities by their own faculty and local industry in need of access to HPC resources for simulation. This goal of providing HPC access for the wide research community in the State can be achieved with an annual appropriation of $2,000,000. In a three year cycle this fund will be sequentially used to purchase a new regional size system for the UofU, USU and 10 small local systems to be distributed among the rest of the institutions. A special oversight committee from he Board of Regents and the Office of Economic development will oversee this program.

Action:

☐ Initiate the process to include this budget request in next year budget.

Distributed Storage Facilities:

Increasingly, research Universities depend on extremely large datasets. Research groups, library groups and other entities need to store this data and make it available electronically to users inside and outside of the University. The data includes digital collections, scholarly communications and curated scientific data. The Utah library coalition is already working on this problem and is requesting funds for a prototype system that will be developed jointly with CHPC. The prototype system will allow immediate access to unique digital collections from all he libraries in the state.

Modern HPC storage systems typically have a very distributed nature, making extensive use of local caches to minimize network usage and increase performance for the delivery of the material. We propose to develop a distributed storage system that follows the scheme used for the HPC systems including two large systems at UofU and USU, respectively and smaller systems at the rest of the colleges and universities in the State. While both research institutions share experience in distributed HPC, they have less experience in distributed data storage facilities, which is a much less developed field across the nation. Therefore before presenting a comprehensive plan for data storage we will work closely with the library community to develop a prototype system on which a final design can presented.

Actions:

☐ Continue working with the Library coalition to refine the proposal for a prototype distributed storage system that will be proposed to the legislature.
☐ Secure Legislative funding for the prototype system
☐ Develop final architecture for the distributed storage system

Data Centers:

The proposed cyber infrastructure facilities as well as other IT assets of the Universities in the State of Utah are housed in data centers that were designed for dated computer technologies. If the State is going to make a serious investment in cyber infrastructure it will also need to provide the necessary physical facilities to house and power the different cyberinfrastructure components. Modern data centers are needed for the two research institutions in the State of Utah. These data centers will connect via the dedicated high bandwidth optical connectivity that the Utah cyber infrastructure uses as its backbone. This network will provide the services necessary for the research enterprise, the redundancy for critical IT services and other services for all the higher education system.

Actions:

☐ Hire a consultant to provide pre-design documents for requesting formal architectural proposals for the construction of the major data centers at UofU and USU.

❑ Hire a consultant to evaluate the need and optimal distribution of minor data centers in the rest of colleges and universities in the State.
❑ Initiate the process of including the Data Centers construction in the State building plan.