This report was submitted to the UC Information Technology Guidance Committee in December 2006 for its consideration. The proposals contained herein are not necessarily endorsed by the ITGC.

Focus Area on High Performance Research Computing
Report for the ITGC meeting on December 11, 2006

A. Background/Process Details

1. Charge

The High Performance Research Computing Work Group (co-chaired by Jim Davis, Associate Vice Chancellor of Information Technology, and Charles Rowley, Associate Vice Chancellor of Computing and Communications) will engage principal investigators, institute directors, and research computing experts to develop strategies that leverage UC's investment in and development of high performance computing services to the broadest possible user base throughout UC.

2. Membership

Chairs: Jim Davis, UCLA; and Chuck Rowley, UCR

Consultant(s): David Walker, UCOP

ITGC Liaison: Kristine Hafner, UCOP

Members:

- Henry Brady, UCB
- John Rundle, UCD
- Frank Wessel, UCI
- Michael Witherell, UCSB
- Peter Yellowlees, UCD

ITLC UC Research Computing Group:

Co-chairs:

- **Bill Labate**, Associate Director-Academic Technology Services, UCLA
- **Frank Wessel**, Manager, Research Computing Support and Research Physicist, Physics & Astronomy, UCI

Members:

- Ed Boring, Computer Coordinator-Earth Sciences, UCSC
- Ann Dobson, Associate Director-Central Computer Services, UCB
- Bob Grant, Director-Technology, UCR
- David Greenbaurn, Director-Strategic Technology Planning, UCB
- David Hutches, Director-Engineering Computing, UCSD
- Russ Harvey, Director-Computing Infrastructure & Computing, UCR
- Gary Jung, Scientific Cluster Support Project Manager-Information Technology, LBL
- Elise Meyer, Interim Director-Information Technology, UCSB

- Harry Mangalam, Research Computing Specialist Programmer-Network & Academic Computing Services, UCSD
- Moma Mellor, Director-Data Center, Client Services, UCD
- Brad Smith, Director-Core Technologies, UCSC
- Heidi Schmidt, Director- Library, UCSF
- David Walker, Director-Advanced Technologies, UCOP
- Tammy Welcome, Information Technology, LBL
- Dave Zavatson, Lead Programmer- Data Center & Client Services, UCD

3. Work Plan

The High Performance Research Computing Steering Group works collaboratively with the ITLC UC Research Computing Group, the ITGC Advanced Network Services Work Group, the ITGC Stewardship of Digital Assets Work Group, the Vice Chancellors for Research, and *ad hoc* faculty groups with interest in creating research cyberinfrastructure for the University of California. Two system-wide meetings regarding cyberinfrastructure have already been held in October, 2005 and May, 2006, and a third is planned for early in 2007.

4. Proposals

- Recommendation to Create UC Grid

- Recommendation to Create a Cyberinfrastructure for all UC Research and Scholarship

- Recommendation to Encourage Sharing of HPRC Resources
- Recommendation for Storage Resources within UC Grid
- Recommendation to Communicate the University's Position on Research Cyberinfrastructure

For all proposals, funding strategies should be adopted that:

- Provide stable funding for research cyber infrastructure to maintain UC's research competitiveness
    - Basic research tools and services
    - Advanced functionality
- Subsidize resource sharing
- Leverage opportunistic funding (e.g., grants) when possible
- Promote (at the state and federal level) UC research competitiveness relative to cyberinfrastructure commitments and readiness

B. Analysis

There have been opportunities to hear and assimilate opinions and comments from a wide spectrum of UC academic researchers, VCRs and CIOs, and industrial practitioners and partners, to track national trends with cyber infrastructure initiatives and to receive feedback that UC is falling behind with respect to competitiveness in cyberinfrastructure-supported and IT-enabled research. The ITGC Research oversight group will begin tackling broad-based considerations for UC. From these discussions, we can begin to pose at least four primary premises that describe areas of potential for UC leadership and national competitiveness:

Premise 1. There is a set of state/national/global problems challenging society that UC should be playing a leadership role in helping to solve. The challenge problems include energy development, use and management, homeland security, environmental impact and

management, infectious disease management, food and water supply, health management and earthquake preparedness.

Premise 2. The State/US value share in a global market is expected to dramatically turn toward vision and application of "knowledge" and "information" that can be applied in highly targeted, individual use and use-specific ways. Targeted use encompasses 'kinds' of applications like use-specific product designs, just-in-time product production, assembly and/or delivery, highly refined social and economic analyses, just-in-time community analysis, substantially shortened product life cycles and highly refined and tuned technologies and processes, and data and information integration on macro scales.

Premise 3. The State/US mind share and IP impact of knowledge and information for use at macro system levels will stem from understandings and technologies at micro levels i.e. the cell, the molecule or the individual. There will be greatly escalated amounts of data and information used to conduct the research and development.

Premise 4. A transition to global management, e.g. global supply chains and smart manufacturing sites that operate to tighter specifications, more tightly tuned socioeconomic decision making, environment impact management, will involve much greater understanding of the processes, greater automation and decision support, more expansive use of automation, data and data interpretation, and a new generation workforce that is trained and oriented toward a knowledge and information mindshare. Across these, there are significant roles for simulation, modeling, optimization, design, ubiquitous sensing, spatially wide area data aggregation and interpretation and large scale/wide area data management, visualization and analysis but in significantly evolving ways that are supported by cyber infrastructure. There are four areas that standout for cyber infrastructure:

- Data need to be collected, mined and interpreted in increasingly sophisticated ways that involve increasing numbers and/or larger coverage of sensors to achieved more refined understanding, models and information. Massive increases in the need for data storage, management, shared access, standards and security loom large for the near future.
- Modeling, simulation, parametric analysis, concurrent evaluations, data analysis etc. are needed to shorten product life-cycles, achieve next generation products and develop processes for just-in-time, targeted use products and goods.
- Operational balances among automation, decision support and human involvement will need to shift but in a coherent manner that depends concurrently on technology development, operational and regulatory trust of automation and education, human factors development and training.
- Sophisticated understanding and practical development at the level of the molecule, the cell or the individual but used at the macro/global level of a company, a region, a state, or a country require increasingly sophisticated computation resources that need to aim toward petascale capabilities and the mathematical and computer science expertise to move to new plateaus of application code.
- Addressing macro/global problems requires sharing of data, information, data and IP.

A UC-grid is a key enabling capability that takes UC toward the virtualization of resources across geographical locations. It then becomes possible to invest in, aggregate distributed resources and optimize a given geographical location for a particular type or class of service that can then be made available to a wide range researchers regardless of location. Virtualization should not be confused with centralization. Virtualized services become components of an adaptive infrastructure that is responsive to the management and needs of the individual researcher or local research group as well as large interdisciplinary teams. High performance computational, visualization and data services can be made available without any one group or facility being over burdened with providing all services with their own resources. The grid and the exposed

resources become a sum-greater-than-the-whole enabling infrastructure for the conduct of the research and grant competitiveness. A UC-grid has the potential to significantly alter the playing field for UC's research investment:

1. What is the investment and incentive strategy for computational and storage resources that can reside at one campus and be used by another.
2. What education and behavioral incentives are needed to engender trust and demonstrate capability such that faculty notions of facility ownership can change.
3. How can infrastructure investments be positioned for grant competitiveness.
4. How can UC leverage its collective stake in IP associated with application software of IT enabled research and how can it efficiently facilitate the strategic development of software to solve bigger problems.
5. How can UC foster collaboration and facility development and invest to solve large problems of state and national interest.
6. How can UC foster the education and training needed to draw benefits from an investment in cyber infrastructure resources.

Recommendation to Create UC Grid

Objective

Create UC Grid to enable sharing of the University of California's research computing resources.

Background

The demands on the UC Research Cyberinfrastructure are growing at a phenomenal rate as a result of the increase in computational power available, data being collected and generated, and the requirement for more inter-disciplinary collaboration. This growth is taxing the capability of existing networks, support staffs, and storage and backup services along with creating an ever-increasing complexity for researchers needing access to these resources.

To deal with this demand, a new set of tools and - more importantly - a new and adaptive infrastructure must be developed. To this end, the design goal of the UC Grid Project is to give researchers a unified view of, and access to, their computational resources, regardless of their physical location. From any desktop equipped with a web browser and a network connection (PDAs can also be used), a researcher can perform all the tasks associated with the submittal of serial, parallel and pre-defined commercial application computing jobs to any computational resources they have access to. The UC Grid will give a researcher the ability to view the overall status of a cluster, number and types of jobs in various queues, busy and free nodes, and the ability to move data and program files within and between clusters utilizing a secure, certificate-based access model. It will also provide access to a rich set of visualization tools that work in conjunction with UC Grid resources. The UC Grid will be built on the Open Source Globus Toolkit. The UCLA-built Grid Portal and Appliance overlay software, on which the UC Grid will be built, has already been released to the Open Source community.

The ultimate vision of UC Grid Project is the complete virtualization of resources across geographical locations, where any given location could be optimized for a particular type or class of service and then be made available to a wide range of researchers through a unified, web-based Grid interface. In this way extremely high performance computational, visualization and data services can be made available without any one group or facility over-burdened by attempting to provide all services with their own resources.

Action Items

- By the end of 2007, create a production UC Grid with participation from all UC campuses and national labs. It is anticipated that this will involve a phased implementation starting with three pilot campuses. As part of this work, the following issues should be addressed:
    - A cost model for expansion of UC Grid.
    - Identification of resources that are appropriate for broad (UC incented) sharing within UC's research community.

Recommendation to Create a Cyberinfrastructure for all UC Research and Scholarship

Objective

Create a cyberinfrastructure-enhanced infrastructure in support of all research and scholarship at UC.

Background

Cyberinfrastructure is more than just high-performance computing clusters. Its scope must include all aspects of research and scholarship that can be enhanced by information technology. For example, all researchers can benefit from the following:

- Collaboration tools and services, such as video conferencing, wikis, and blogs, in addition to electronic mail and telephones.
    - This is an opportunity to link not just data sets and computing resources, but researchers and research teams, from main 10 campuses, the 5 med centers and the national labs in ways that should greatly increase research opportunities, ideas and outcomes. In the medical area this is especially important in terms of translational research, where both UCD and UCSF have both been successful in getting 2 of the first 12 national Clinical Translational Science Center awards and the HPC report could greatly support and drive translational research (ie bench to bedside to community).
- Electronic publishing tools and services that assist in the creation and dissemination of information in electronic form.
- Substantial high quality research data sets that could be used by any UC researcher
- Digital preservation services that ensure availability of information in the future.
- Access to federal and University services related to grant application and administration.

These services are currently available only as uncoordinated services. We can expect them to become more numerous, as well as more integral to the research endeavor. In order to avoid greater distraction or UC's researchers from their research, UC should integrate these services into a common service infrastructure for all of UC.

Action Items

- By the end of 2007, and consulting extensively with the research community, identify the IT-based tools and services that can best enhance UC's research mission with phased completion of components through 2010. As part of this work:
    - Determine which of these services and tools should be included as a base-level set of services that should be funded for all UC researchers, and which might require additional funding.
    - Outline a project that integrates these services and tools into a "Research Portal" environment. The portal strategy must accommodate both services provided system-wide, as well as services provided by the campuses to their unique communities.
    - A user support model that appropriately leverages local (discipline knowledgeable), campus, and system-wide support personnel.

- Appropriate integration with other academic and administrative tools and services.

Recommendation to Encourage Sharing of HPRC Resources

Objective

Implement strategies that encourage sharing of IT resources in support of research and scholarship when it is appropriate but do not force resource sharing when it is not appropriate.

Background

With few exceptions, researchers at UC currently are on their own with respect to acquiring, supporting, and housing the information technology (IT) resources they require to do their research. While this strategy has the potential of close alignment between the IT resources and specific research needs, it has a number of potential negative impacts:

- The cost of physical facilities to house IT resources (*e.g.*, space preparation, power, air conditioning) can be very high, typically as high as the computing resources themselves. In fact, as the unit price of IT resources declines, we can expect the relative cost of facilities to exceed and continue to grow with respect to the cost of IT resources. This can result in less than adequate facilities for resources that support critical research.
- System administration is a rare skill and, particularly in the area of high-performance computing, a rapidly developing one. Researchers who do not find talented system administrators can be severely disadvantaged.
- Individual researchers must acquire the maximum resource they will ever need. There is no capacity for extraordinary, short-term needs.
- When the target of research is not specifically computing, the time spent on issues related to the IT infrastructure is a distraction from useful work.

The University needs to develop strategies to encourage appropriate sharing of these critical resources:

- Data center space
- System administration
- Computing resources
- Storage resources

Action Items

- By the end of FY 2007/2008, create shared computing and storage resources within UC Grid, leveraging internal and external funding sources, that address the common needs of researchers at UC. This work should:
  - Be capable of providing a base level of computing and storage to all UC researchers (See Recommendation to Create a Cyberinfrastructure for all UC Research and Scholarship.)

    - Result in allocation mechanisms that guarantee resource delivery to funders, while making otherwise-unused resources available to the entire UC research community
    - Address data center space needs
    - Address system administration
    - Address end-user support needs
    - Be "reusable" for resources that are shared less broadly within UC.
- By the end of 2007, create short and long term plans for data center space within the University. These plans must address:
    - Projected need for shared data center space at the system-wide and campus levels.
    - Projected need for unshared space for specific research projects.
    - Criteria determining the need for physical proximity of specific resources and people (*e.g.*, the research requires physical access, or speed of light delays would slow computation significantly).
    - Funding strategies to ensure the availability of sufficient space to support UC research.

- By the end of FY 2007/2008, create a collaborative strategy for administration of HPRC system resources that leverages expertise at all campuses.

Recommendation for Storage Resources within UC Grid

Objective

Create a data storage infrastructure within UC Grid.

Background

The ability to store data is a common theme across many of the ITGC work groups

- High performance research computing often requires access to large storage resources to store raw data and processed results.
- Stewards of digital information need resources within which to store the information collections they manage.
- The Instructional Technology group has identified the need to create a collection of course-based information.
- Many of the new modes of collaboration, such as wikis, enable collaborative management of highly dynamic collections of information.

Not all storage requirements are the same. They may differ in multiple ways, including retention, availability, responsiveness, integrity, and access controls. For example, the intermediate results of a simulation may have a short retention with low requirements for availability and integrity, but the responsiveness of access must be very fast. On the other hand, a PhD thesis written in 1970 would have long retention, high requirements for availability and retention, but slow response to access requests might be acceptable. Both of these examples have a high requirement for access control over modification of the data, but low requirements for control over "read" access.

It should be noted that, in this context, a data storage service is not digital stewardship. Rather it is a service required by digital stewardship.

Action Items

- By the end of 2007, identify a framework for data storage resources within UC Grid. This work should include representatives from the research community, as well as the CDL and other stewards of digital information. The framework should address the following issues:
    - Packaged service offerings appropriate to UC's research community, such as high-availability storage, "scratch" storage, or backup storage.
    - Facilities and system administration appropriate for very large collections of storage systems
    - Cost models for the various service packages
    - Interoperability standards among data storage services, as well as between data storage services and other UCGrid resources.
    - Access control.

Recommendation to Communicate the University's Position on Research Cyberinfrastructure

Objective

Develop a concise statement of the University of California's research cyberinfrastructure strategy.

Background

The University of California is one of the foremost research institutions in the world, and cyberinfrastructure has or is taking a central role in how research is conducted in nearly all disciplines. The University does not, however, have an articulated research cyberinfrastructure strategy that can be used to educate federal and state agencies about its capabilities to use information technology to enhance research.

Such a statement, communicated appropriately through UC's government relations channels, can greatly increase competitiveness for funding, as funders will be likely to be familiar with UC's capabilities before receiving grant proposals.

Action Items

- By the end of 2007, develop a statement of cyberinfrastructure strategy and a communications plan to educate federal and state agencies about its capabilities to use information technology to enhance research.