

Project Summary: High Performance Computing Center Sustainability

National NSF resources, including the Track 1 and Track 2 systems, offer formidable computing capabilities to key national researchers who work on extraordinarily complex problems. However, the vast majority of researchers in Science, Technology, Engineering and Mathematics (STEM) disciplines continue to rely on departmental, campus, or regional/state research computing resources. They use these “local” computing resources (1) to fulfill their science and engineering computational requirements, and/or (2) to prepare their codes for eventual migration to national facilities, and/or (3) to educate the critically needed students that are required by the knowledge economy if we are to excel from a scientific perspective, a competitive perspective, and a national security perspective. Different institutions take a broad variety of approaches to research computing. Some universities consider research computing a strategic investment and attempt to provide sustained support for significant research computing resources, including sizeable parallel clusters. These are typically housed in formally recognized centers. More commonly universities view research computing as a tactical need, and only provide intermittent funding for research computing. All of these research computing centers are challenged to understand how best to organize, manage, fund, and utilize their hardware and staff. These centers play a critical role in educating the next generation of scientists and engineers – the very cadre that will need to extensively utilize high performance research computing resources to be globally competitive intellectuals.

The objective of this workshop is to provide a forum for an open discussion among center directors, campus Chief Information Officers and Research Officers on the topic of the sustainability of research computing centers. The discussion will yield a shared understanding of organizational models, funding models, management models and training models that result in *sustainable research computing centers*. An ancillary, but equally important, outcome will be to provide a degree of transparency across the extant centers that will provide evidentiary justification for centers that are struggling to become established and are increasingly called on to justify the significant expenses, and the resulting ROI, that naturally occur as centers become established.

Intellectual Merit: Most students are lucky to discover High Performance Computing (HPC) late in their undergraduate careers, or even while studying as graduate students. Many of the best students are never exposed to HPC at all – a distinct shortcoming in our current educational processes. One of the key reasons for this paucity of HPC opportunities is the lack of ubiquitous, well-organized, well-funded advanced research computing centers. This lack of training makes it difficult for US students to obtain the requisite level of expertise needed for entry into the modern science workforce, or entry into a STEM field that requires the extensive use of research computing. This workshop will directly impact the intellectual capabilities of our higher education institutions by initiating a sustained dialog into the increasingly critical requirement for sustained investments and pervasive use of advanced research computing, which can only be realized through thoughtful organizational change at the highest levels of the academy.

Broader Impacts: Faculty and staff who attend the workshop will return to their home institutions better prepared to elucidate and champion the need for established research computing centers, and they will have the necessary data to explain how and why such centers must be established and can be sustained. Further, this workshop will prepare higher education institutions located in economically disadvantaged areas of the country with models for successful research computing centers that, if created and sustained, can markedly impact local economies. Additionally, by developing and sharing institutionally-siloed knowledge across diverse centers, this workshop will facilitate the establishment and implementation of similar centers elsewhere and will strengthen and enrich broader learning communities. Finally, by promoting sustained research computing centers, this workshop will help to ensure early exposure to advanced computational concepts for all science and engineering students.

“NSF Workshop on High Performance Computing Center Sustainability”

1. Workshop Objective and Goals

High Performance Computing (HPC) continues to become an increasingly critical resource for an expanding spectrum of research disciplines. Both the National Science Foundation (NSF) and the Department of Energy (DoE) have created and support a powerful set of national HPC centers that provide key national users with access to state-of-the-art computing capabilities. These centers include both the NSF Track 1 and Track 2 facilities that are on-line, or will be coming on-line soon, as well as the DoE HPC centers, including the DoE Leadership Class Facilities. Applications of petascale computational science to model phenomena that are difficult or impossible to measure, to accelerate the time to discovery, and to integrate advancements are critical for the competitiveness of U.S. industry.

While national resources provide formidable computing capabilities to key researchers that work on extraordinarily complex problems, the vast majority of scientific and engineering *researchers* continue to rely on departmental, campus, or regional/state research computing resources. They use these computing resources to 1) fulfill their science and engineering computational requirements, and/or to 2) prepare their codes for eventual migration to the national facilities, and/or to 3) educate the critically needed students that are required if we are to “weather the storm” from both a competitive perspective and a national security perspective.

To satisfy these requirements, many universities have been focusing on identifying economies of scale, creating second and third tier centers that provide HPC resources to their researcher communities in the most cost effective and sustainable ways possible. However, the recent economic downturn is creating challenges to sustaining these centers. Second and third tier centers are faced with major challenges in funding, organizational structure, and long-term sustainability. Although the first and second tier centers funded by the NSF and those serving academic partners through the DoE have been somewhat insulated from budget pressures, unit, institutional and regional centers may face budget challenges in the coming years as the NSF transitions from the TeraGrid to a new model of funding, creating even more competition for funding. The identification of suitable center sustainability models is more important than ever, as the survival of second and third tier centers is crucial to national efforts to advance science and engineering discovery.

Different institutions take a broad variety of approaches to research computing. Some universities and university systems consider research computing a strategic investment and have attempted to provide sustained support for significant research computers, including sizeable parallel clusters, which are typically housed in formally recognized centers. Other universities view research computing as a tactical need, and may provide only intermittent funding for research computing for smaller, informal centers. In either case, these research computing centers are struggling to understand how best to organize, manage, fund, and utilize their hardware and staff.

Industry standard computing solutions provide a low cost of entry into HPC hardware, but there are significant hidden costs, including:

- Building renovations, including space, power and cooling
- Administrative staff to install, maintain and support resources and research users
- Additional infrastructure requirements such as disk storage, backup, networks and visualization
- Consulting staff that can support optimization and scaling for research codes, as well as assisting researchers in discovering and exploiting national resources and funding opportunities

The research computing ecosystem must be sustained, and a national discussion on sustainable centers is needed.

1.1 Objective

The objective of this workshop is to provide a forum for an open discussion among center directors, campus Chief Information Officers and Research Officers on the topic of research computing centers. The discussion will yield a shared understanding of organizational models, funding models, and management models that result in *sustainable research computing centers*. An ancillary, but equally important outcome, will be to provide a degree of transparency across the extant centers that will provide evidentiary justification for centers that are struggling to become established and are increasingly called on to justify the significant expenses, and the resulting ROI, that naturally occur as centers become established.

1.2 Goals

The goals of the workshop are to:

1. Discuss and document:
 - The spectrum of current research computing center models
 - A preliminary conceptual framework for organizing a national research computing ecosystem
 - New and innovative research computing centers business models
 - Models and activities that have failed, the reasons that the failures occurred, and the “lessons learned”
 - Possible regional and national collaboration and cooperation opportunities
 - Vendor relationships, including how to form useful vendor partnerships
 - How to facilitate users who must “cross the divide” and transition to the national infrastructure
2. Create a report that will provide institutions and states with:
 - The key organizational precepts and insights from the workshop
 - Methods and “blueprints” for forming unit, institutional, and regional centers
 - Possible example governance and policy documents that can be used to form new centers and to guide existing centers

2. Meeting Topics

Survival of second and third tier centers is crucial to national efforts to advance science and engineering discoveries. These centers constitute a crucial niche in the cyberinfrastructure ecosystem, providing training and education to undergraduate and graduate students and assistance to faculty prototyping new applications and using cyberinfrastructure resources for the first time. They often also play a significant role in providing access to expensive shared instruments and a knowledgeable conduit to resources and opportunities available at the national scale.

The following six topics, particularly as applied to the needs and goals of second and third tier high performance computing centers, will be the focus of the workshop:

1. Organizational Models & Staffing

Currently a number of such models exist. Centers are in place as separate entities subsidized by a consortium of individual universities. They may also exist as a part of a larger Information Technology operation on a campus, as a division within an institution’s research administration structure, as a research center associated with one or several colleges within a university, or in various hybrid forms. Leaders representing each area will present an overview of these organizational models and the advantages and disadvantages of each.

2. **Funding Models**

As central subsidies for centers decline, various fee-for-service models are being put into place. The mix of services and fee structures range across a number of categories from maintenance and management of computing resources to consultation with major research projects, to a package of fees for services. Centers are also increasingly competing for extramural funds for both research and industrial contracts. We will discuss examples of each of these funding models and the markets or situations in which they appear to be most successful.

3. **Vendor Relationship**

Smaller centers do not have the buying power of the major centers and thus are less likely to receive the pricing and array of options available to those larger entities. Strategies that emerge from this situation range from creating strong ties with a single vendor to encouraging long-term and better support to developing local expertise and expending staff resources on assembling heterogeneous systems. We will discuss various strategies and problems associated with vendor relationships, as well as the potential for regional and/or national cooperation that might lead to a broader set of options.

4. **Succession Planning**

Many centers have limited staff and thus face potential major problems as key leaders or key staff retire or take positions elsewhere. With a very limited pool of expertise in high performance computing, such transitions can lead to the demise of a center unless actions are taken to anticipate possible changes and to provide a succession plan that will work. At the same time, many centers are being asked to transition from one organizational model to another. Such transitions pose similar problems, as staff may resent the changes and thus may move to alternative jobs. These issues will be discussed and potential approaches to their solution will be discussed.

5. **Metrics of Success and Return on Investment**

As budgets become tighter, centers are increasingly asked to justify their return on investment. Metrics are therefore becoming an increasingly important aspect relating to the survival of HPC centers. Approaches to defining metrics of success such as return on investment, gathering and maintaining the necessary data such as resource usage and usability, depth and breadth of impact, and effective means of presenting them to key decision-makers will be discussed.

6. **Industry Relationships**

As both industry and academic centers are pressed by budget limitations, there are opportunities for joint projects with and services to industry that could become an important aspect of center activities. Examples of industry partnerships, services and service models, and the challenges of developing an industrial customer base will be addressed at the workshop.

7. **Recent Related Meetings and Talks**

- A recent workshop hosted by the EDUCAUSE Net@Edu campus Cyberinfrastructure Working Group (CCI) and the Coalition for Academic Scientific Computation (CASC) ⁱⁱ emphasizes the challenges that must be addressed to build a coherent CI that elegantly spans from the local to the national level, and the potential opportunities that can be created if such a CI ecosystem can be realized.
- A report published by Educause focuses on Business Models for Sustainable Data Centers <http://www.educause.edu/Resources/BusinessModelsforSustainableDa/160195>
Educause Net@EDU (State Networks) - 2/5/2007
[Guy T. Almes \(Texas A&M University\)](#)
[Paul Killey \(University of Michigan-Ann Arbor\)](#)

[James Pepin \(Clemson University\)](#)
[Dallas Thornton \(University of California, San Diego\)](#)

- A talk at the University of Oklahoma Supercomputing Center for Education & Research Symposium 2006 on Sustainable Funding Models for Academic High Performance and Research Computing was given by Dan Stanzione - Director ASU.

8. **Organizing Committee**

- Stanley C. Ahalt, Ph. D.
Executive Director, Ohio Supercomputer Center
ahalt@osc.edu
614-292-9524
- Amy W. Apon, Ph. D.
Director, Arkansas High Performance Computing Center, University of Arkansas
aapon@uark.edu
479-575-6794
- Henry Neeman, Ph. D.
Director, OU Supercomputing Center for Education & Research, University of Oklahoma
hneeman@ou.edu
405-325-5386
- David Lifka, Ph. D.
Director, Cornell University Center for Advanced Computing
lifka@cac.cornell.edu
607-254-8621

9. **Proposed Location, Date and Announcement**

The event will be during the first two weeks of May 2010 at Cornell University, Ithaca, New York in the Statler Hotel. An exact date will be posted on an official workshop web site when funding is confirmed.

Cornell University: <http://www.cornell.edu/>

Statler Hotel: http://www.statlerhotel.cornell.edu/statler_experience/

Announcement of the workshop will be via CASC, SURA & TeraGrid Mailing lists as well as a press release on HPC Wire and Supercomputing Online. A conference web site/wiki will provide conference details and logistics, and a place for potential participants to submit 2-5 page position papers and to provide suggestions and feedback on the agenda.

10. **Meeting Organization, Agenda, and Dissemination or Results**

A call for participation requiring a 2-5 page position papers will be widely publicized to academic research computing community.

Participants – 75 leaders from around the US, including the leaders of the major state, regional, and university centers – will be invited to participate in the one and a half day workshop that will consist of morning presentation and panel discussions designed to stimulate detailed discussion in the five breakout sessions held the afternoon of the first day.

These breakout session topics will include:

- Organizational Models & Staffing
- Funding Models
- Industry & Vendor Relationships
- Succession Planning
- Metrics of Success and Return on Investment

Each breakout session will have a leader selected by the committee ahead of time from among the participants, based on their position paper and their ability to lead an open and positive discussion. Each breakout session will also have a person focused on capturing the key points of discussion and identify areas of consensus or lack thereof. The breakout session leaders will then report the findings back to the full group of workshop attendees in the form of a short informal set of power point slides. Discussion will be encouraged with a goal of consensus where possible.

The second day will include additional presentations and an industry panel discussion designed to foster discussion on funding strategies, opportunities and developing industry partnerships.

The organizing committee will collect notes, presentations and comments from the workshop wiki and Twitter, and will create a plan to generate a workshop report that we will circulate to all workshop attendees, to ensure that we have stated the findings accurately and succinctly.

10.1 Workshop Agenda

Day 1:	12:00 p.m. 6:00 p.m. – 8:00 p.m.	Afternoon Hotel Check-in Informal Reception
Day 2	8:00 a.m. – 8:30 a.m. 8:30 a.m. – 9:00 a.m. 9:00 a.m. – 9:45 a.m. 9:45 – 10:30 a.m. 10:30 a.m. – 11:00 a.m. 11:00 a.m. – 11:45 a.m. 11:45 a.m. – 12:00 p.m. 12:00 p.m. – 1:15 p.m. 1:15 p.m. – 3:00 p.m. 3:00 p.m. – 3:30 p.m. 3:30 p.m. – 5:00 p.m. 5:00 p.m. – 5:30 p.m. 5:30 p.m. – 8:30 p.m. 8:30 p.m.	Continental Breakfast Welcome: Overview, Goals and Brief Introductions by Participants – Stan Ahalt “The Cornell Center for Advanced Computing Sustainability Model” – Dave Lifka “Bridging Campuses to National Cyberinfrastructure – Overview of OCI Sustainable Center Activities” ▪ NSF Program officer to be specified Break “The Penn State Sustainability Model” – Vijay Agarwala Afternoon agenda discussion and breakout planning – Stan Ahalt Lunch Breakout Sessions ▪ Organizational Models & Staffing ▪ Funding Models ▪ Industry & Vendor Relationships ▪ Succession Planning ▪ Metrics of Success and Return on Investment Break Reports from the breakout sessions Bus to Dinner MV Manhattan Dinner Cruise on Cayuga Lake Bus to Hotel(s)

Day 3	8:00 a.m. – 8:30 a.m.	Continental Breakfast
	8:30 a.m. – 8:45 a.m.	Welcome and Agenda Review
	8:45 a.m. – 9:45 a.m.	“Federal Funding Opportunities and Strategies for Tier 2 and Tier 3 Research Computing Centers” – Jim Bottum
	9:45 a.m. – 10:15 a.m.	Open Discussion on the need for Collaboration and Advocacy – Henry Neeman
	10:15 a.m. – 10:30 a.m.	Break
	10:30 a.m. – 11:30 a.m.	Panel Discussion – Industry and Vendor Relationships (Moderators: Stan Ahalt, Amy Apon, Dave Lifka, Henry Neeman)
	11:30 a.m. – 12:00 p.m.	Wrap up including identification of areas of consensus or lack thereof and report planning – Stan Ahalt, Dave Lifka
	12:00 p.m.	Box Lunch and Adjourn
	12:00 p.m. – 1:00 p.m.	Organizing committee generate report writing assignments and deadlines

10.2 Dissemination of Results

The organizing committee will generate a complete report on the findings of the workshop. We will post the report on the on the CASC website and submit it to EDUCAUSE for publication. We will also leave the workshop wiki up for a period of at least 3 years, making the participants’ position papers, presentations and notes available to the community. Dissemination of results will also be done to the same communities from which participants are recruited (see below).