PROJECT SUMMARY: An EPSCoR Desktop to Teragrid EcoSystem:
Integrating Campus-based with National Cyberinfrastructure
Submitted by: South Carolina and Tennessee EPSCoR Jurisdictions

Research and Education Goals
This proposal is submitted by the South Carolina and Tennessee EPSCoR jurisdictions. Our partnership is based on existing and planned future collaborations in advanced materials and systems biology research, knowledge domains where computational science is driving new approaches and insights. We propose to build **cyberinfrastructure (CI) linked, community specific knowledge environments that embody the desktop to TeraGrid ecosystem.** In this ecosystem**,** the campus-based cyberinfrastructure at a regional research institution is an essential bridge for connecting faculty investigators to national resources such as the TeraGrid. Since NSF program guidelines for 2009 require use of the TeraGrid for all large computational resource requests, it is especially important to develop models that successfully bridge the current gap between desktop and TeraGrid.

A notable gap exists in the inadequate interfacing of TeraGrid and non-TeraGrid national computational resources and data. For users, these circumstances hinder what could be a natural way to seamlessly utilize local, campus layer, and national resources. The campus based cyberinfrastructure layer is the bridge required to address that gap. The NSF TeraGrid has attempted to deliver such an integrated structure to the user community across a set of resources designated for national access and use. However, in spite of impressive results the size of TeraGrid user community falls short of engaging the spectrum of potential users, a number that grows every day. By way of example, faculty at small undergraduate schools such as Francis Marion University and The Citadel are asking Clemson how they can use and invest in regional high performance computing facilities. This trend requires scaling national models down to regional size.

Cyberinfrastructure is more than just connecting people via advanced networks and sophisticated applications; CI is about engaging participants in the generation of knowledge. It is about creating opportunities for participants to share expertise, tools, and facilities in powerful ways that have the potential to significantly advance discovery. An essential enabler of this vision is an ability to coordinate the use of public sector information technology across scales, from desktop to TeraGrid. **This proposal will create and implement a model for that coordination, building a seamless application development and deployment, technical support, training, and personal communications bridge from desktop to TeraGrid.**

EPSCoR jurisdictions have faced many barriers (*eg*: geographic and economic) to developing CI [1]. For many years, both South Carolina and Tennessee lagged far behind in research and education network infrastructure; however, this situation has changed dramatically in recent years. Both jurisdictions have achieved significant investments in advanced networks and computational resources:

- Private sector investment in the C-Light optical network, connecting South Carolina to Internet2 and National Lambda Rail (NLR) ;
- State of South Carolina investment in the South Carolina Light Rail (SCLR) optical network, connecting C-Light to the state research universities (Clemson, USC, MUSC);
- State of South Carolina funding for the Clemson CyberInstitute and Endowed Chair in CI ($1.5M)
- Clemson University investment in a regional computational center for research in scientific computing with a strategic objective of being at 10% size of a national center (eg: achieve 100TF system, 10% of 1 petaflop system). This center includes research scientist support staff, the 45TF Palmetto cluster (ranked #62 in Top500, November 2008) ($2.2M); expertise acquired by hiring

several nationally known experts (Bottum, Pepin, Goasguen, Gemmill); membership in Open Science Grid and TeraGrid.

- Investment of $8.4M by Vanderbilt University (VU) to establish the Academic Center for Computational Research and Education (ACCRE), a 1400+ core cluster for mid-scale computing,
- VU role as a US data repository and principal analysis site (a "Tier 2" site) for the Compact Muon Soleniod-Heavy Ion experiment to be conducted on the Large Hadron Collider at CERN; VU added a dedicated 10Gbps circuit to the Southern Crossroads (SoX) in Atlanta representing a 15-fold increase in VU's external bandwidth for research purposes, and enabling order-of-magnitude increases in throughput for Internet2, National LambdaRail (NLR), and ESNet traffic.
- Federal funding (DoE, NSF) for UTK/ORNL to advance from essentially no network connectivity to ORNL to that facility being the best connected in the country while vastly improving the Tennessee R&E network infrastucture;
- Federal funding (DoE,NSF) for a TeraGrid Track 2, soon to be petascale facility, at ORNL.
- Federal funding (DoE, FCC) for South Carolina to extend SCLR to extension offices and rural health facilities throughout the state

These investments, "golden spikes in cyberinfrastructure", now provide the foundation to support cross-jurisdiction collaborations in systems biology and advanced materials. *"The conduct of science and engineering is changing and evolving. This is due, in large part, to the expansion of networked cyberinfrastructure and to new techniques and technologies that enable observations of unprecedented quality, detail and scope."* (page 2, National Science Foundation, Strategic Plan FY 2006-2011).

Innovative use of Cyberinfrastructure

Now that this essential infrastructure is in place, strategic investment in research computational scientists who can partner with domain scientists to use larger scale computational resources is needed. Writing and compiling software that can make full use of mult-core, large scale systems is a challenge. Researchers in advanced materials and systems biology may not have the expertise to solve (or solve in reasonable time) important computational problems that arise in the course of their investigations. We will address that need through a new, multi-tiered model that provides campus researchers with a seamless integration of local, regional, and national computational expertise to address their computational challenges. This multi-tiered model, illustrated in Figure 1, will provide a communication and

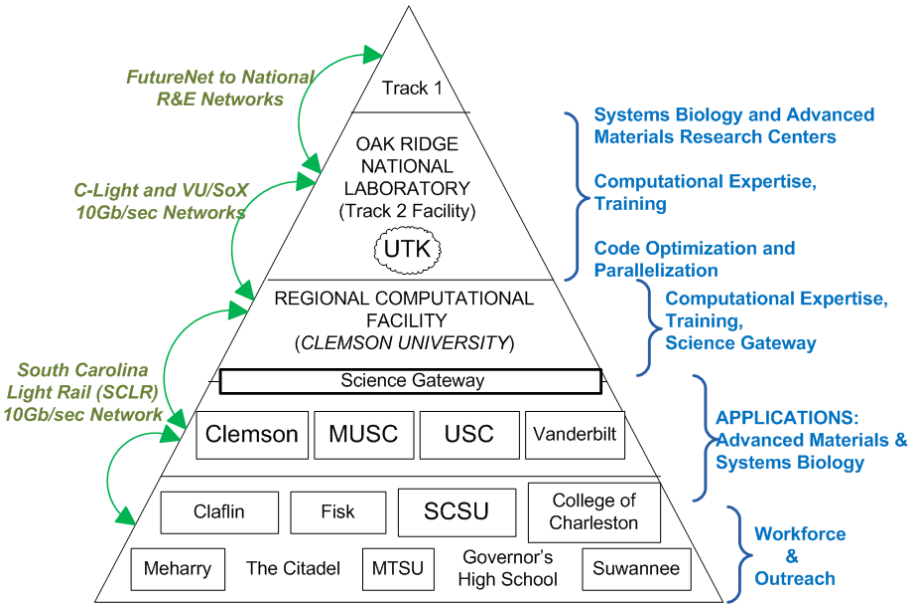

Figure 1. *The proposed desktop to TeraGrid ecosystem leveragesCyberinfrastructure investments in networks and high performance computing systems to advance approaches to Advanced Materials and Systems Biology. A campus-based approach and science gateway facilitate use of the TeraGrid.*

organizational fabric to bring undergraduate students into internships at a national laboratory, build communication and working partnerships among regional experts and support personnel, and leverage nationally funded training and support resources.

Our specific objectives are described in Section 3 and, organized using the NSF's four cyberinfrastructure pillars, include:

- **Computation and Networks:** (a) improvements to network infrastructure in the two states; (b) extension of Clemson's "condominium cluster" model to a group of EPSCoR researchers; (c) facilitated use of the TeraGrid by two EPSCoR knowledge communities; (d) significant increase in computational expertise via hiring of six additional research scientists; petascale scaling of at least one project application;
- **Data:** (a) High Throughput data transfer via dedicated optical network paths, (b) storage and high throughput i/o for data
- **Virtual Organizations:** a Science Gateway allowing researchers to harvest spare computational capacity; a "Systems Biology Workbench" for use by participants in this project; a "Computational Endstation for Advanced Materials", for use by participants in this project
- **Education and Outreach:** (a) increased number of internships at ORNL for South Carolina under-graduates; (b) new members for the Oak Ridge Affiliate University program; (c) dissemination of our results to the higher education CI community and broad public. (d) strategic advances in the consortium's diversity and awareness of CI among diverse populations by leveraging ORNL outreach programs and partnership the Minority-Serving Institution CI Empowerment Coalition.

An additional objective is:

- **Assessment/Oversight:** program evaluation under the direction of Tom Finholt and external advisory board members Ian Foster, Mark Ellisman, Rebecca Doerge, and a fourth person TBD;

Intellectual merit

We are creating a model desktop to TeraGrid ecosystem anchored by a campus-based CI layer that might be termed a "Track 3" regional research resource bridging campus investigators to TeraGrid. The need for such a model has been identified by the EDUCAUSE Campus Cyberinfrastructure Working Group and the Coalition for Academic Scientific Computing (report from workshop, July 2008). We will define and describe how campuses can successfully leverage the TeraGrid program by enabling the systems biology and advanced materials applications described in this proposal. We will achieve progress in STEM diversity by our partnership with undergraduate and Historically Black Colleges and Universities (HBCUs). We will advance a TeraGrid Track 2 site's experience with and understanding of the national middleware program while building mechanisms to access regional and TeraGrid resources seamlessly through a state Science Gateway.

Broader impacts

Solving these technical and communication challenges is essential to scaling use of the TeraGrid. Our SC/TN consortium includes partner undergraduate universities (College of Charleston, the Citadel) and HBCUs (Claflin University and South Carolina State University (SCSU)) along with engagement and training outreach opportunities for the region as well as an innovative dissemination plan. Claflin and SCSU are located in Orangeburg, South Carolina, in a county ranked in the bottom 10 counties for per capita income, and 83% minority; **these institutions' role in drawing more interest in research oriented careers and awareness of the academic path to those jobs, cannot be over stated.**

This proposal leverages a total <$$XXXX> private, state, federal investment in CI;