

**National Workforce Development:
Preparing the Next Generation of Practitioners and Educators
Who Will Enable Scientific Discovery Through
Effective and Sustained Use of Cyberinfrastructure**

A White Paper for the Office of Cyberinfrastructure at NSF
From the leadership of TeraGrid and Open Science Grid
(June 2009)

Vision and Goal

The potential exists for a broad and diverse population of students, motivated and engaged in different aspects of computationally intensive research, to form the cornerstone for the advancement of scientific discovery for generations to come. In order for this to be realized, we must establish a coordinated and sustained effort that will build such a pool of students and prepare them to become productive members of our national cyberinfrastructure workforce, including support staff, software developers, educators and researchers in compute and data intensive disciplines. To be clear, when discussing computationally intensive research, the required cyberinfrastructure is inclusive of a broad range of technologies including, processing, data, visualization, instruments and sensors, and networking, as well as collaborative and learning environments. The focus of this white paper is on the engagement and preparation of students to utilize these emerging technologies to advance scientific discovery.

It is important that all activities maintain a persistent emphasis on pursuing strategies that engage students and educators whose diversity reflects the diversity of the nation's population and the inherently multifaceted nature of cyberinfrastructure.

Motivation

Today computing is an increasingly essential and integral tool in all scientific research and scholarship. There are numerous reports identifying the critical need and pathways for growing a larger and more diverse community of scientists, technologists, engineers and mathematicians (STEM). The reports include *NSF's Cyberinfrastructure Vision for 21st Century Discovery* (2007), *Computational Science: Ensuring America's Competitiveness* (2005 PITAC Report), and *Fostering Learning in the Networked World: The Cyberlearning Opportunity and Challenge; A 21st Century Agenda for the National Science Foundation* (2008); and *Strategies and Policies to Support and Advance Education in e-Science* (2009). Historically, the vast majority of the members of our national cyberinfrastructure workforce emerged through their own individual motivation and persistence. As a nation we lack any structured programs to proactively develop this critical component of the national research and education human infrastructure.

Challenge

As major contributors to our national cyberinfrastructure, the Open Science Grid (OSG) and TeraGrid (TG) provide services and resources through which scientific discovery, in many fields, is accomplished. A key responsibility of these two national cyberinfrastructure providers (NCPs) includes the effective preparation of the next generation of IT/technical personnel, educators and researchers in the ever evolving area of research computing. Preparation is required on several levels, and involves a diverse population of students and educators. The future STEM workforce must be prepared to harness cyberinfrastructure at all levels for scholarship and discovery. At the

same time, the future workforce must also be positioned to continue innovation and evolution of cyberinfrastructure to support the growing and evolving needs of research and education communities. Students need guidance and mentoring to get involved in research computing and be motivated to learn how to be part of interdisciplinary teams that may consist of computer scientists, mathematicians and domain scientists. Domain experts must learn how to employ complex software tools and to harness powerful resources and services; the technical professionals must learn how to deliver dependable capabilities to demanding domain experts. This must all be done in the context of collaborating with the global cyberinfrastructure workforce—research and education is a globally connected endeavor and international collaboration and leverage must be an integral component in the development of our national workforce.

Scope

While the challenges for workforce development are very broad, the intended scope of this white paper is focused on engaging and preparing significantly larger numbers of U.S. high school, undergraduate and graduate students to become persistent members of the community of well prepared STEM practitioners to advance computationally intensive research and education computing. The scope of this effort will provide the critically necessary foundation for launching complementary efforts to further expand the nation's workforce.

Objectives

To achieve the goals above, specific objectives have been identified. These include:

- Recruit and engage a diverse community of students and educators
- Develop intensive and wide-ranging education and training content
- Establish a network of knowledge transfer workshops
- Develop a mentored internships program
- Provide opportunities for professional recognition of students.

Students need to be recruited from among U.S. institutions to address a continuum of student engagement and preparation. The objectives and activities, as well as the educators and mentors engaged, need to be adapted to each class across high school, undergraduate and graduate students. Students who engage in the activities (described below) will emerge highly motivated and well prepared to contribute to advancing the impact of computing on science discovery in all fields.

The following are key components of an integrated and coordinated set of activities for engaging and preparing students. Numerous projects have been funded that address one or a few of these elements, but there currently exists no comprehensive, integrated program that provides all of these components in a coherent and coordinated fashion to prepare students to contribute to cyberinfrastructure-enabled research.

Stimulate Engagement

Students and mentors who reflect the diversity of the nation's population must be recruited. The goal is to address the challenge of successfully including women, minorities, and people with disabilities. Achieving an initial enthusiasm and then sustained interest of such populations is challenging. Particular attention must be paid to methods that include already engaged experienced role models, relevant problem solving and socially relevant introductory environments.

The effort must include engaging students from rural and urban high schools, 2- and 4-year colleges and universities, minority serving institutions, and EPSCoR institutions. This will also include recruiting students from fields currently under-represented in computationally-enriched research (e.g., humanities, arts, and social sciences).

Students must be invited to meet with researchers, educators and technologists through conferences (e.g., SCxy, TGxy) involving the NCPs, workshops, and visits to computing centers and research labs.

Knowledge Transfer

Students must be trained and supported in their use of computationally and data-intensive computing resources and services offered both locally and by the NCPs. Students must learn how to: 1) develop and run computationally intensive applications both local to their own institutions and remotely on services offered by the NCPs; 2) improve code performance for emerging advanced computing systems; 3) manage large data collections; 4) utilize scientific visualization tools; 5) take advantage of other aspects relevant to computationally intensive, cloud computing and other emerging computer based paradigms for conducting scientific discovery; and 6) develop, deploy, operate and support the range of resources, software, and services that constitute the cyberinfrastructure

To address diverse learning styles and just-in-time learning, training opportunities for students must be made available through live events, synchronous distributed events, and on-demand, on-line, self-paced asynchronous learning capabilities. Content tailored specifically for the goals of the knowledge transfer must be adapted and developed based on the practices and technologies available both locally and through the NCPs.

Mentored Internships

Activities must include internships at NCP sites (as well as at international counterparts), with researchers in computationally intensive domains, and with expert developers to provide a breadth of opportunities attuned to the diverse needs and interests of the students as well as emerging workforce needs.

Students need to be provided with a range of support to nurture their involvement. The objective is to help students overcome personal and technical challenges, and sustain their continued involvement in computational STEM fields.

Activities must include the use of on-line collaborative learning environments with ongoing access to experts, teachers, mentors, and peer groups for sharing knowledge, challenges, and successes.

The students' mentors must be selected from among peers, more senior students, staff and researchers from the NCPs, and leading-edge petascale research projects from among academia, industry and government agencies both domestically and with international partners.

Recognition for Accomplishments

There must be opportunities for recognizing the unique talents and skills of the students that are involved. Students need to be empowered to demonstrate their knowledge, skills and talents through competitions, publications, presentations, and leadership roles in various computational projects and activities. Students must be encouraged (and expected!) to present their experiences and findings at relevant conferences and at professional society

meetings where they can interact with researchers and students in their domain of study. These opportunities will enhance the opportunities for the students to pursue advanced studies and professional careers in areas with strong or developing computational needs.

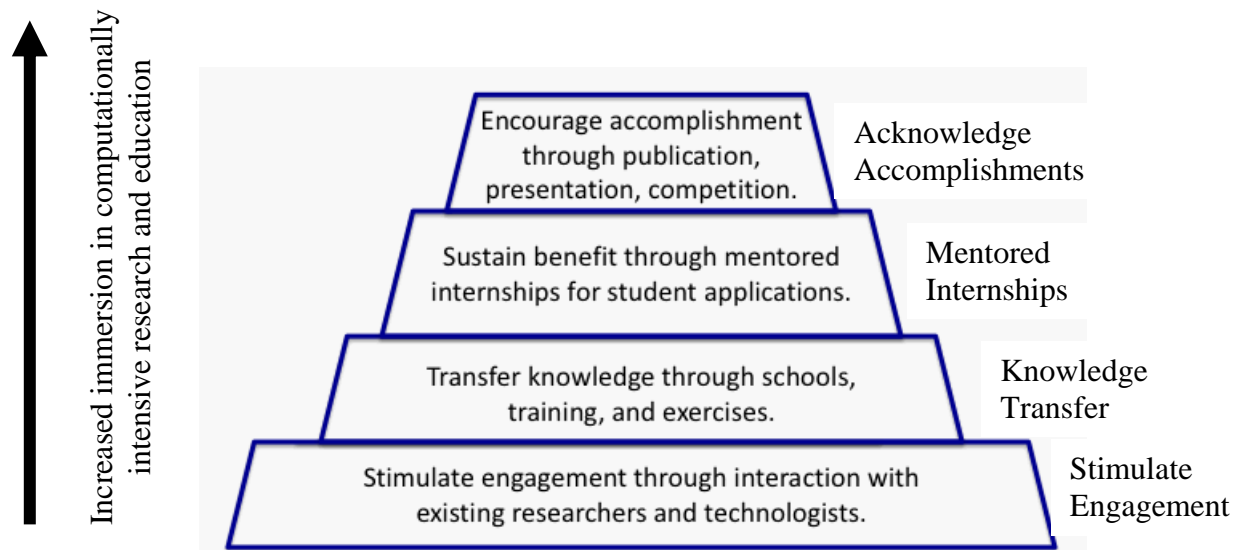
Increased Immersion of Students in Program Activities

The above objectives in effect provide a pyramid (see Figure 1) of coordinated and complementary activities that allow the students to become more and more deeply involved in learning and supporting scientific discovery through the use of computationally rich resources, tools, and methods. As students rise through the levels of involvement described above and illustrated in the pyramid below, they will become more deeply immersed in computationally intensive opportunities that advance their skills, knowledge, and motivation to pursue advanced studies and careers.

It is important that students are retained in the field as they rise through the levels of the pyramid of engagement. The success of any program will be assessed by the number of students that graduate from each of the levels of the pyramid.

To fully achieve the vision and goals, these efforts must engage experts from among the NCPs, whose staff, along with their national and international collaborators, can provide students with a unique base of real-life experiences in conducting research and development that directly complements and expands upon the formal classroom learning experiences of students.

Figure 1: Student Workforce Preparation Immersion Pyramid



Outcomes

The goal is that a large portion of the students recruited continue within the field, completing the highest level of the pyramid. As an indicator of success, another significant outcome would be intentions – and commitment – on the part of a significantly large number of the students to pursue advanced studies and professional careers that contribute to the advancement of scientific discovery, either directly as researchers or educators, or as professional staff providing and supporting the cyberinfrastructure that researchers and educators will harness.

Any work in this area must also produce a set of guidelines for replicating successful efforts, as well as candid advice and recommendations for avoiding pitfalls and problems encountered

during implementation of the program. Guidelines must include metrics for both immediate and long-term success, including indicators of students' likelihood of continuing to pursue studies and careers in technologies employed by compute intensive science.

Efforts in this area must track students' continued involvement in the field as long as possible beyond their completion of their participation in activities to assess their long-term career choices. "Graduates" of any programs will create a pool of talent from which to recruit mentors for subsequent students.

External evaluation is crucial to success. It needs to be involved early, to provide both formative and summative evaluations. Students' sustained participation in the evaluation process is essential to assess impact on their attitudes, interests, and likelihood of pursuing STEM careers. Formative evaluation must provide ongoing advice for revisions and improvement, while yielding documented advice and guidance for replication and scaling-up to larger communities of students.

Foundations for Success

The members of the NCPs are uniquely poised to support this program's success in transforming the preparation of today's students to develop a community of researchers, educators and practitioners for advancing computational scientific discovery.

Social scientists can help address student engagement, mentoring and support, as well as help develop evaluation strategies and mechanisms for long-term tracking of the program's impact on students beyond the period of their involvement in the program.

An advisory group (that includes students) must be established to provide periodic advice for short-term and long-term improvement, while also exploring approaches for both sustaining and scaling activities to have even larger national impact.

Action Plan

The critical national need for achieving the goals outlined in this document are well documented; the challenges require a commitment to human capacity development, and the expectations for successful outcomes are very high in all sectors of society. While the expertise, talent and research-based proven strategies for success exist today to accomplish these goals, there is no cohesive articulated national strategy to guide leadership, expenditures and decision-making.

A critical next step must be for the Office of Cyberinfrastructure (OCI) at NSF to launch a five-year pilot program to implement the vision outlined in this white paper. Five years are needed to be able to assess the impact on and retention of students as they progress through the educational system and into advanced studies and professional careers.

The pilot effort must have both an extensive evaluation plan for assessing the impact of the program, as well as an Advisory Council formed by NSF to assess the program impact. The results of the evaluation and the advice of the Advisory Council will be critical to informing NSF in formulating strategies for launching larger scale programs for developing the next generation of computationally experienced workforce.