CHAPTER 4
VIRTUAL ORGANIZATIONS FOR DISTRIBUTED COMMUNITIES (2006-2010)

I. New Frontiers in Science and Engineering Through Networked Resources and Virtual Organizations

With access to state-of-the-art cyberinfrastructure services, many researchers and indeed entire fields of science and engineering now share access to world-class resources spanning experimental facilities and field equipment, distributed instrumentation, sensor networks and arrays, mobile research platforms, HPC systems, data collections, sophisticated analysis and visualization facilities, and advanced simulation tools. The convergence of information, grid, and networking technologies with contemporary communications now enables science and engineering communities to pursue their research and learning goals in real-time and without regard to geography. In fact, the creation of end-to-end cyberinfrastructure systems – comprehensive networked resources – by groups of individuals with common interests is permitting the establishment of Virtual Organizations (VOs) that are revolutionizing the conduct of science and engineering research and education. A VO is created by a group of individuals whose members and resources may be dispersed geographically and/or temporally, yet who function as a coherent unit through the use of end-to-end cyberinfrastructure systems. These CI systems provide shared access to centralized or distributed resources and services, often in real-time. Such virtual organizations supporting distributed communities go by numerous names: collaboratory, co-laboratory, grid community, science gateway, science portal, and others.

During the past decade, NSF funding has catalyzed the creation of VOs across a broad spectrum of science and engineering fields, creating powerful and broadly accessible pathways to accelerate the transformation of research outcomes into knowledge, products, services, and new learning opportunities. With access to enabling tools and services, self-organizing communities can create end-to-end systems to: facilitate scientific workflows; collaborate on experimental designs; share information and knowledge; remotely operate instrumentation; run numerical simulations using computing resources ranging from desktop computers to HPC systems; archive, e-publish, access, mine, analyze, and visualize data; develop new computational models; and deliver unique learning and workforce development activities.

Through VOs, researchers are exploring science and engineering phenomena in unprecedented ways. Scientists are now defining the structure of the North American lithosphere with an extraordinary level of detail through EarthScope, which integrates observational, analytical, telecommunications, and instrumentation technologies to investigate the structure and evolution of the North American continent and the physical processes controlling earthquakes and volcanic eruptions. The Integrated Primate Biomaterials and Information Resource assembles, characterizes, and distributes high-quality DNA samples of known provenance with accompanying demographic, geographic, and behavioral information to advance understanding of human origins, the biological basis of cognitive processes, evolutionary history and relationships, and social structure, and provides critical scientific information needed to facilitate conservation of biological diversity. The Time-sharing Experiments for the Social Sciences (TESS) allows researchers to run their own studies on random samples of the population that are interviewed via the Internet. By allowing social scientists to collect original data tailored to their own hypotheses, TESS increases the precision with which social science advances can be made. Through the Network for Earthquake Engineering Simulation (NEES), the coupling of high performance networks, advanced data management, distributed coordination and computational tools, and 15 experimental facilities enables engineering

The GPS station shown on the opposite page monitors transform plate movement in California as part of EarthScope. The data collected will be compared with data from other GPS stations to better understand the interplay of faults in the region.

researchers to test larger scale and more comprehensive structural and geomaterial systems to create new design methodologies and technologies for reducing losses during earthquakes and tsunamis.

reliable, accessible, usable, pervasive, persistent and interoperable, and that are able to exploit the full range of research and education tools available at any given time.

The NEES@buffalo shake table recreates movements of the forces generated by an actual earthquake. NEES allows earthquake engineers and students located at different institutions to share resources, collaborate on testing, and exploit new computational technologies.

This chapter describes the establishment of a national cyberinfrastructure framework into which the HPC environment described in Chapter 2 and the national data framework described in Chapter 3 are integrated, enabling the development, deployment, evolution, and sustainable support of end-to-end cyberinfrastructure systems that will serve as transformative agents for 21st century science and engineering discovery and learning, promoting shared use and interoperability across all fields.

II. The Next Five Years: Establishing a Flexible, Open Cyberinfrastructure Framework for Virtual Organizations

NSF's five-year goals are as follows:

- To catalyze the development, implementation and evolution of a functionally complete national cyberinfrastructure that integrates both physical and cyberinfrastructure assets and services to support VOs.
- To promote and support the establishment of world-class VOs that are secure, efficient, reliable, accessible, usable, pervasive, persistent and interoperable, and that are able to exploit the full range of research and education tools available at any given time.
- To support the development of common cyberinfrastructure resources, services, and tools enabling the effective, efficient creation and operation of end-to-end cyberinfrastructure systems for and across all science and engineering fields, nationally and internationally.

The following principles will guide the agency's FY 2006 through FY 2010 investments:

- NSF's investments in end-to-end cyberinfrastructure systems are driven by science and engineering opportunities and challenges.
- Common needs and opportunities are identified to improve the cost-effectiveness of NSF investments and to enhance interoperability.
- NSF investments promote equitable provision of and access to all types of physical, digital, and human resources to ensure the broadest participation of individuals with interest in science and engineering inquiry and learning.
- Existing projects and programs inform future investments, serving as a resource and knowledge base.
- End-to-end cyberinfrastructure systems are ap-

propriately reliable, robust, and persistent such that end users can depend on them to achieve their research and education goals.

- NSF partners with relevant stakeholders, including academe, industry, other federal agencies, and other public and private sector organizations, both foreign and domestic.
- Tools and services are networked together in a flexible architecture using standard, open protocols and interfaces, designed to support the creation and operation of robust networked resources and VOs across the scientific and engineering disciplines supported by NSF.

A digital optical module is lowered into a hole over 1,450 meters deep in the ice at South Pole station. These sensors are deployed as part of the IceCube international collaboration to further understanding of the Universe by finding evidence of high energy subatomic particles.

The cyberinfrastructure framework developed will integrate widely accessible, common cyberinfrastructure resources, services, and tools, such as those described in other chapters in this document, enabling individuals, groups and communities to efficiently design, develop, deploy, and operate flexible, customizable networked resources and VOs to advance science and engineering.

In facilitating the creation and support of effective virtual organizations, NSF will focus on three essential elements: the creation of a common technological framework that promotes seamless, secure integration across a wide range of shared, geographically-distributed resources; the establishment of an operational framework built on productive and accountable partnerships developed among system architects, developers, providers, operators, and end users who span multiple communities; and the establishment of an effective assessment and evaluation plan that will inform the agency's ongoing investments in cyberinfrastructure for the foreseeable future.

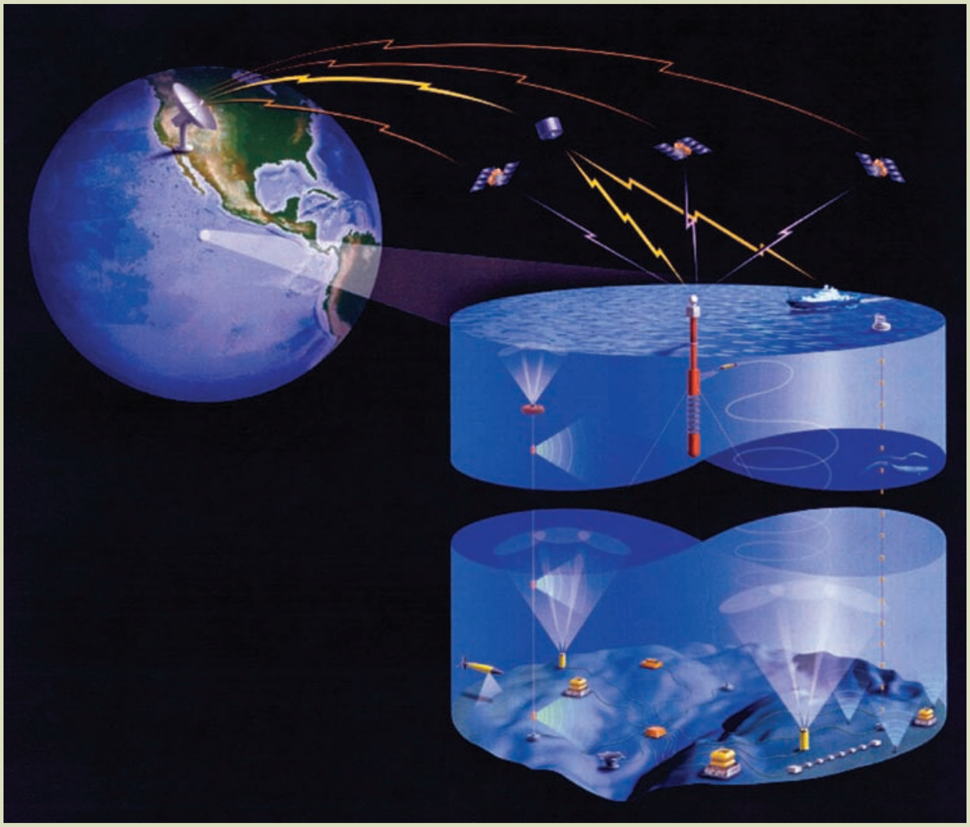
A. Open Technological Framework

To facilitate the development of an open technological framework, NSF will support cyberinfrastructure software service providers to develop, integrate, deploy, and support reliable, robust, and interoperable software. Software essential to the creation of networked resources and VOs encompasses a broad range of functionalities and services, including enabling middleware; domain-specific software and application codes; teleobservation and teleoperation tools to enable remote access to experimental facilities, instruments, and sensors; collaborative tools for experimental planning, execution, and post-analysis; workflow tools and processes; system monitoring and management; user support; web portals to simulation software and domain-specific community code repositories; and flexible user interfaces to enable discovery and learning.

Many of the projects listed in Appendix E have produced fundamental software and/or new integrated environments to support interdisciplinary research and education. NSF's strategy leverages this body of work, harvesting promising tools and technologies that have been developed to a research prototype stage, and further hardening, generalizing, and making them available for use by multiple individuals and/or communities. For example, many communities require access to services that build scientific workflows and rich orchestration tools, and to integrate the intensive computing and data capabilities described in Chapters 2 and 3, respectively, into collaborative and productive working environments. While work has already been done to develop and deploy components and packages of needed software through NSF and other support, existing software needs to be hardened, maintained and evolved. New software and enhancements to existing software must be developed as new uses and new user requirements continue to emerge.

Cybersecurity pervades all aspects of end-to-end cyberinfrastructure systems and Virtual Organizations, and includes human, data, software and facilities elements. Security requires coordination, the development of trust, and rule setting through community governance. NSF will require awardees developing, deploying and supporting networked resources and virtual organizations to develop and apply robust cybersecurity policies and procedures, thereby promoting a conscientious and consistent approach to cybersecurity. The agency will support development of strong authentication and authorization technologies and procedures for individuals, groups, VOs, and other role-based identities. In so doing, the agency will leverage research prototypes and new technologies produced in the Cyber Trust program, the U.S. government's flagship program in cybersecurity research and development.

Interoperable, open technology standards will be used as the primary mechanism to support the further development of interoperable, open, extensible networked resources and VOs. Ideally, standards for data representation and communication, connection of computing resources, authentication, authorization, and access should be open, internationally employed, and accepted by multiple research and education communities. Where appropriate, conventions that are employed internationally should be favored. The use of standards creates economies of scale and scope for developing and deploying common resources, tools, software, and services that enhance the use of cyberinfrastructure in multiple science and engineering communities. This approach allows maximum interoperability and sharing of best practices. A standards-based approach will ensure that access to cyberinfrastructure will be independent of operating systems; ubiquitous, and open to large and small institutions. Together, web services and service-oriented architectures are emerging as a standard framework for interoperability among various software applications on different platforms. They provide important characteristics such as standardized and well-defined interfaces and protocols, ease of development, and reuse of services and components, making them potential central facets of the cyberinfrastructure software ecosystem.



The Ocean Observatories Initiative promises to provide the ocean sciences research community sustained, long-term and adaptive measurements in the oceans via a fully operational research observatory system.

B. Operational Framework

NSF will promote the development of partnerships to facilitate the sharing and integration of distributed technological components deployed and supported at national, international, regional, local, community, and campus levels. Significant resources already exist at the academic institution level. It is important to integrate such resources into the national cyberinfrastructure fabric.

In addition, NSF will encourage partnerships with industry to develop, maintain and share robust, production-quality software tools and services and to leverage commercially available software. NSF will engage commercial software providers to identify value propositions, to identify software needs specific to the research and education community, and to facilitate technology transfer to industry, where appropriate. Efforts to ensure that NSF cyberinfrastructure investments complement those of other federal agencies will be intensified.

With the increasing globalization of science and engineering and its attendant cyberinfrastructure, NSF supports international efforts of strategic interest. NSF will endeavor to: (i) facilitate U.S. researchers' collaboration with international partners through cyberinfrastructure; (ii) identify exemplars of international collaboration and partnerships, utilizing cyberinfrastructure, that offer efficient and beneficial relationships and build on these; and (iii) encourage international collaboration in the development of cyberinfrastructure.

The cost-effective penetration of cyberinfrastructure into all aspects of research and education will require the full engagement of the broad science and engineering community. Incorporating the contributions from multiple communities and reconciling their interests is one of the major challenges ahead. Community proxies must be identified and empowered to find common interests to avoid duplication of effort and to minimize the balkanization of science and engineering.

C. Evaluation and Assessment

Cyberinfrastructure is dramatically altering the conduct of science and engineering research and education. Accordingly, studying the evolution and impact of cyberinfrastructure on the culture and conduct of research and education within and across communities of practice is essential. NSF will also support projects that study how ongoing and future cyberinfrastructure efforts might be informed by lessons learned and by the identification of promising practices. Among other things, NSF seeks to build a stronger foundation in our understanding of how individuals, teams and communities most effectively interact with cyberinfrastructure; how to design the critical governance and management structures for the new types of organizations arising; and, how to improve the allocation of Cyberinfrastructure resources and design incentives for its optimal use. These types of activities will be essential to the agency's overall success. NSF will support studies of the evolution and impact of cyberinfrastructure on the culture and conduct of research and education within and across different research and education communities. It will also encourage systemic design of virtual organizations with embedded evaluation to better address the complex interaction of technological and social issues with the user community.

HPWREN, a high performance wireless research network, expands the reach of cyberinfrastructure into remote environments in and surrounding San Diego County to support a range of science, engineering, education, and emergency response initiatives.

The rapidly evolving nature of cyberinfrastructure requires ongoing assessment of current and future user requirements. Comprehensive user assessments will be conducted and will include identification and evaluation of how the physical infrastructure, networking needs and capabilities, collaborative tools, software requirements, and data resources affect the ability of scientists and engineers to conduct transformative research and provide rich learning and workforce development environments. Other issues to be addressed include the degree to which cyberinfrastructure facilitates federated inquiry, interoperability, and the development of common standards and new social norms.