CYBERINFRASTRUCTURE

Science and engineering have undergone a revolution in which the traditional approach of observation, experimentation, theory, and analysis has been dramatically enhanced by use of advanced computing and communication technology with information in digital form. Digital interfaces found in modern laboratory equipment can be networked to make it possible for scientists to directly participate in experiments from across the country or around the globe. Wired, wireless, and optical networking and the growth of autonomous systems enable researchers to deploy elaborate webs of sensors in domains as diverse as environmental science and astronomy. Advances in the analysis and visualization of digital data permit researchers to analyze large, complex collections of data. Information technology has made it possible for groups of collaborating researchers to overcome distance and work together more effectively as they tackle the hard problems of modern science and engineering.

In parallel with the emergence of digital technologies that increase the capabilities of researchers, the questions at the forefront of scientific and engineering research have become increasingly complex. Researchers wish to unravel the way multiple processes, interacting over multiple space and time scales, produce the rich variety of phenomena seen in complex systems. Such complex systems permeate the natural and engineered world, from the intricate workings of a cell to the emergence of structure in the early universe and the workings of the internet. Advances in understanding in the digital realm are providing the tools that help researchers tackle the new research challenges in the physical, biological, and social sciences.

Cyberinfrastructure Funding
(Dollars in Millions)

	FY 2007 Actual	FY 2008 Estimate	FY 2009 Request
Biological Sciences	$90.50	$97.13	$99.78
Computer and Information Science and Engineering	71.00	87.00	87.00
Engineering	54.00	56.00	60.00
Geosciences	75.00	75.00	80.00
Mathematical and Physical Sciences	61.21	64.56	71.06
Social, Behavioral and Economic Sciences	20.60	20.54	20.54
Office of Cyberinfrastructure	182.39	185.33	220.08
Office of International Science and Engineering	1.30	0.75	0.75
Office of Polar Programs	43.70	26.24	26.24
Subtotal, Research and Related Activities	$599.70	$612.55	$665.45
Education and Human Resources	17.74	16.25	16.50
Total, Cyberinfrastructure Funding	$617.44	$628.80	$681.95

Totals may not add due to rounding.

The term ***cyberinfrastructure*** was coined to encompass many of the systems used for working with digital information that have the potential to fuel advances in research, education, industry, and society. NSF has supported pioneering efforts by researchers:

- to understand the scientific foundations and develop components of cyberinfrastructure, mainly through investments made by the Directorates for Computer and Information Sciences, Mathematical and Physical Sciences, and Engineering;

- to use these components to break new ground in science and engineering research through investments made by all NSF directorates and the Office of Cyberinfrastructure, and
- to investigate how cyberinfrastructure should be integrated into the research and education enterprise, mainly through investments made by the Office of Cyberinfrastructure and the Direcorate for Computer and Information Science and Engineering.

The success of these and related programs has demonstrated the power of cyberinfrastructure and caused many research communities to express an urgent need both for greater access to and for new types of cyberinfrastructure.

Extensive input from the research and education communities led NSF to develop the document, *A Cyberinfrastructure Vision for the 21st Century,* which defines NSF's leadership roles in an integrated system of high performance computation services, services for managing massive and heterogeneous data/information, sensing and observation across multiple scales of time and space, multimode visualization and interaction, and distributed team collaboration. It also describes learning and workforce issues associated with applying cyberinfrastructure to learning as well as the learning required to use cyberinfrastructure. Achieving the vision requires linking three complementary activities: 1) research and development of tools, concepts, and technologies; 2) provisioning of leading-edge cyberinfrastructure systems; and 3) the application of cyberinfrastructure to advance our understanding of the world around us, respond to emergencies, and provide more authentic and motivational STEM learning opportunities for students, teachers, professionals, and the general public. Investments in FY 2009 are designed to capitalize on the results of the pioneering early forays into cyberinfrastructure and to advance research and education through the implementation of the strategies laid out in this vision.

Grand challenges in many fundamental research areas will benefit from investments in cyberinfrastructure. For example, analyses of DNA sequence data are providing remarkable insights into the origin of life, revolutionizing our understanding of the major kingdoms of life, and revealing stunning and previously unknown complexity in microbial communities. Sky surveys are changing our understanding of the earliest conditions of the universe and providing comprehensive views of phenomena ranging from black holes to supernovae. Researchers are monitoring socio-economic dynamics over space and time to advance our understanding of individual and group behavior and their relationship to social, economic, and political structures. Using combinatorial methods, scientists and engineers are generating libraries of new materials and compounds for health and engineering, and environmental scientists and engineers are acquiring and analyzing streaming data from massive sensor networks to understand the dynamics of complex ecosystems.

The American Competitiveness Initiative (ACI) describes the goal of providing world-leading high-end computing capability, coupled with advanced networking, to enable scientific advancement through modeling and simulation at unprecedented scale and complexity across a broad range of scientific disciplines. NSF investments in high-performance computing for research and education, the TeraGrid infrastructure, middleware investments, and international network connections directly contribute to the goals of the ACI. The enormous growth in the availability and utility of cyberinfrastructure capabilities, both technology- and human-based, is increasing the productivity of scholarly research, accelerating the transformation of research outcomes into products and services that drive economic growth, and enhancing the effectiveness of learning across the spectrum of human endeavor.

All NSF activities participate in support for cyberinfrastructure. The Office of Cyberinfrastructure (OCI) makes investments common to a broad range of science and engineering fields, promoting economies of scale and scope, and facilitating interoperability. Other directorates and offices make complementary cyberinfrastructure investments necessary to meet their disciplinary missions. Some highlights of NSF's FY 2009 investments, led by the designated activity, follow:

- Continued investments in the Plant Science Cyberinfrastructure Collaborative (PSCIC), a new type of organization – a cyberinfrastructure collaborative for plant science – will enable new conceptual advances through integrative, computational thinking. The collaborative will utilize new computer, computational science, and cyberinfrastructure solutions to address questions in plant science. The collaborative will be community-driven, involving plant biologists, computer and information scientists, and experts from other disciplines working in integrated teams. (BIO)

- SBE's Science of Science and Innovation Policy (SciSIP) activities are developing an evidence-based platform from which policymakers and researchers may assess the impacts of the Nation's science and engineering (S&E) enterprise, improve their understanding of its dynamics, and predict outcomes. Specifically, data collection, research, and community development components of SciSIP's activities will: (a) improve and expand science metrics, datasets, and analytical tools, yielding changes in the bi-annual S&E indicators and other data collections; (b) develop usable knowledge and theories of creative processes and their transformation into social and economic outcomes; and (c) build a community of experts in this area across the federal government, industry, and universities. SciSIP supports the development of new data, models, and tools, and also facilitates transformative research on an immensely policy-relevant topic — the ecology of innovation. SBE is at the forefront internationally of the collection of data on the S&E workforce and research and development statistics. These data, in conjunction with the new theoretical models and analytical tools being developed with support from SciSIP, will inform and enhance the success of the ACI. In addition, the Innovation and Organizational Change program focuses attention on the effects of innovative cyberinfrastructure on companies and scientific laboratories. (SBE)

- Continued support will be provided for the development of a versatile, open-source, community Ocean Modeling Environment and to identify and refine best practices and describe trade-offs between alternatives for simulating a range of important ocean processes. A second thrust is to develop and assess the capability to dynamically configure the grid resolution of future ocean models. The Budget Request also supports work of the Community Surface Dynamics Modeling System (CSDMS), a national effort to develop, support, and disseminate to the geoscience research and teaching community, integrated software aimed a predicting the erosion, transport, and deposition of sediment and solutes in landscapes and their repository sedimentary basins. A key science issue to be explored by this initiative includes understanding the fluxes, reservoirs, and flow paths associated with the physical, biological, and chemical transport processes in the Critical Zone -- the skin of the earth. Both projects will have considerable societal impacts. Finally, the Request supports a variety of projects that will develop or maintain community databases. (GEO)

- The Computational Infrastructure for Geodynamics (CIG) will develop, support, and disseminate software for the geoscience community, from model developers to end-users. The software is designed to investigate problems ranging widely from mantle and core dynamics, crustal and earthquake dynamics, magma migration, and seismology. The long-term goal of CIG is to create a set of computational tools and data structures that can be commonly applied within the geodynamics community. A common set of computational tools will enable the development of models of Earth evolution that intimately couple lithosphere, convecting mantle and core, with the capability to eventually simulate the planet as a whole. (GEO)

- Continued support will be provided for software and services that facilitate complex science and engineering research and that advance ACI goals in data-intensive applications. These include innovative approaches to data management and stewardship, and middleware for distributed applications, distributed collaboration, interactive remote observation and the tele-operation of experimental facilities. Continued investment in leading-edge computational infrastructure and international network connections will support the research of U.S. investigators and their ability to collaborate internationally in projects such as the Large Hadron Collider (LHC). Investments will be made in numerical models, data analysis tools and new algorithms in strategic science and engineering research areas in order to take advantage of forthcoming petascale computing systems. (OCI)

- Our society now faces the mammoth challenge of synthesizing information and deriving knowledge from massive, dynamic, ambiguous and possibly conflicting digital data. In FY 2009, CISE researchers will create the foundational computational concepts, models, tools, and algorithms that will allow human-centered visualization systems of the future to more effectively create new knowledge from data. (CISE)

- Continuing investments will be made in the Directorate for Education and Human Resources to support a national resource of high-quality internet-based STEM educational content. (EHR)

- Improved understanding and design approaches will be pursued for auto-reconfigurable engineered systems enabled by cyberinfrastructure. Autonomous reconfigurability is a promising concept for ensuring appropriate operational levels during and after unexpected natural or man-made events (e.g. hurricanes, pandemics, or terrorist attacks) that could impact critical engineered systems in unforeseen ways. (ENG)

- The International Geophysical Year (IGY: 1957-1958) or International Polar Year (IPY-3) ushered in the modern era of polar research and provided the first detailed measurements of the polar ocean, atmosphere, land, and space. Emerging advances in cyberinfrastructure that occur during the IPY (2007-2009) will for the first time link remote instruments with scientists in the field and institutions around the world, allowing scientists to observe and record the pulse of our planet in real-time. (OPP)

- Continued cyberinfrastructure support will be provided for the Arctic Systems Sciences (ARCSS) Data Coordination Center that serves as a central point for deposition of data deriving from ARCSS-funded research. Continued support is also provided for Arctic modeling, distributed field sites, and autonomous flux towers. In the Antarctic, funds support data center/data repositories, 3-D bathymetric data fusion, and environmental monitoring, both marine and terrestrial. In addition, support is provided for the engineering, operations and maintenance, and security of cyberinfrastructure systems. (OPP)

- Research at the frontiers of mathematical and physical sciences spans multiple length and time scales, ranging from studies of elementary particles to studies of nuclear, atomic, and molecular properties, dynamics, and reactions to the discovery of new materials and states of matter. At the other end of the length scale are studies of supernovae, structures in the universe, and space-time. Achieving a fundamental understanding of these diverse phenomena involves the use of cyberinfrastructure. In addition, cyberinfrastructure enables the prediction and discovery of new materials and of new states of matter. (MPS)

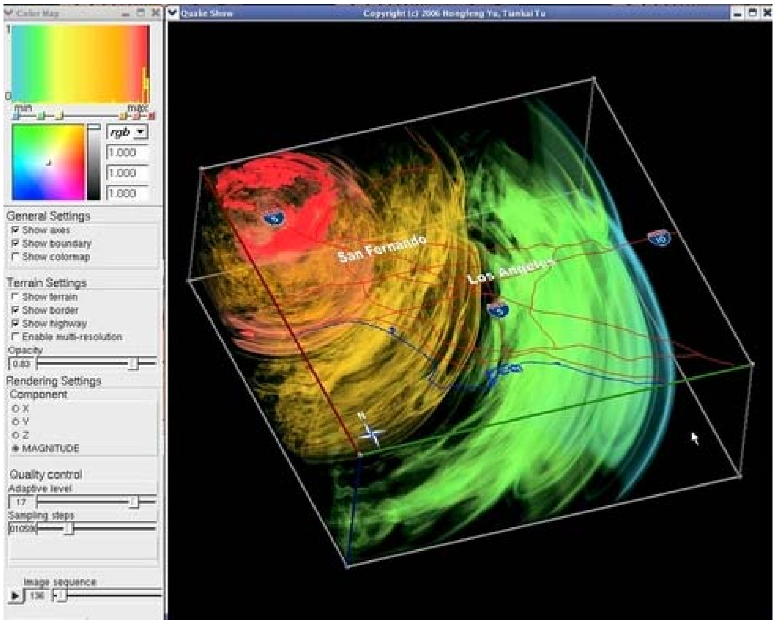
- Advances in the performance of information technology hardware bring previously computationally intractable problems within reach. Physical scientists and mathematicians contribute to the development of high-performance computing, high-speed networks, data mining, software, and algorithms. Cyberinfrastructure ensures community access to the best software for a wide range of research activities ranging from the interpretation of experimental data to the visualization of simulation results. It enables scientists and students to easily tap into the hundreds of terabytes of archived data available around the world. Such databases are fast becoming an essential component of the physical sciences research environment. Emphasis will continue on the development of global GRID network technology to enable discoveries to be made in petabyte data sets generated at unique facilities and involving international collaborations, *e.g.*, LIGO and LHC. (MPS)

- Distributed computing, where the public is able to download software that enables them to add their personal computers to the computations associated with a particular problem allows for a novel "collaboration" with the general public. Distributed computing directly connects with the NSF goal of enhancing discovery and understanding while promoting teaching, enhancing infrastructure for research and education, and broad dissemination to enhance scientific understanding. (MPS)

Over time, NSF investments will contribute to the development of a powerful, stable, persistent, and widely accessible cyberinfrastructure to enable the work of science and engineering researchers and educators across the Nation and around the world.

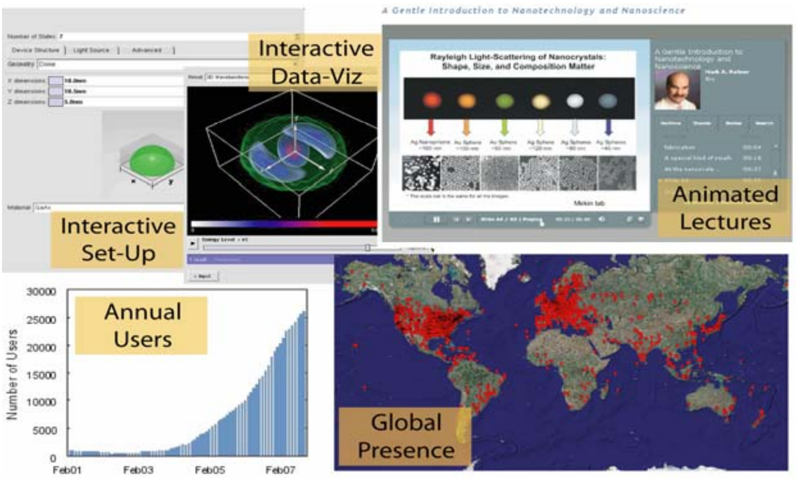
Recent Research Highlights

► **Earthquakes on Demand:** Researchers created a ground-breaking capability to use thousands of computer processors to simulate in fine detail earthquakes and other events. The process brings together diverse aspects of the computer science field to work on real scientific problems provided by earthquake scientists. The system enhances the ability of large supercomputer centers and networks to provide powerful tools that support scientists and engineers. These tools improve the ability to predict strong ground-motion during earthquakes and will eventually help zoning code developers produce more accurate building requirements. They will also make critical infrastructure, such as power-plants and waste facilities, safer. This work has wide-ranging effects in computer science, earth science, civil engineering, and computational modeling. (CISE)



The laptop screen during a real-time animation of a simulation of the 1994 Northridge earthquake, while the simulation is running on 2050 processors of the Cray XT3 at PSC. The animation shows the displacements of the ground over time. Onscreen controls allow the user to interactively change the parameters of the volume rendering algorithm while the simulation and rendering algorithms are running on the processors of a remote supercomputer. *Credit: Hongfeng Yu, Tiankai Tu, CMU.*

► **Economic Impact of Globus:** Globus was started in 1996 as a research project to identify enabling mechanisms for resource federation across enterprises. That work led to the creation of the widely used open source Globus Toolkit and to the large and aggressive "grid" community in eScience and industry. Intel, Raytheon, and Cisco have publicly described their Globus work, job ads abound for Globus experts, and IBM promotes Globus as its standard open source grid platform. The indirect impact of Globus is also significant. Virtually every major computer vendor has a "grid product," and most of the Fortune 500 companies have a "grid strategy." While some of the activity might be characterized as "just" clustering solutions, others are more ambitious resulting in significant benefits. (OCI)



nanoHUB user community. *Credit: Gerhard Klimeck, Purdue University*

► **Simulating Nanotechnology:** The NSF Network for Computational Nanotechnology (NCN), a network of six universities headquartered at Purdue, has a mission to accelerate the advancement of nanotechnology. The NCN also created a unique cyberinfrastructure, a nanoHUB, which offers online simulation on-demand, through an ordinary Web browser. Its third-generation middleware supports integrated data analysis and visualization capabilities without having to download, compile, or install software. Simulations run on powerful computers in the nation's grid infrastructure for unprecedented reach into laboratories, classrooms, and other areas for greater utilization. The user community has doubled every year, and the number of simulation users has quadrupled. It executed over 230,000 simulation jobs for more than 5,900 registered users nationwide. (OCI)