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.

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 develop components of cyberinfrastructure, to use these components to break new ground in science and engineering research, and to investigate how cyberinfrastructure should be integrated into the research and education enterprise. 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.

Cyberinfrastructure Funding
(Dollars in Millions)

	FY 2006	FY 2007	FY 2008	Change over FY 2007	
	Actual	Request	Request	Amount	Percent
Biological Sciences	$84.00	$90.50	$95.50	$5.00	5.5%
Computer and Information Science and Engineering	64.37	68.00	87.00	19.00	27.9%
Engineering	52.00	54.00	58.00	4.00	7.4%
Geosciences	71.35	75.00	75.00	-	-
Mathematical and Physical Sciences	58.64	63.56	64.56	1.00	1.6%
Social, Behavioral and Economic Sciences	20.54	20.54	20.54	-	-
Office of Cyberinfrastructure	127.14	182.42	200.00	17.58	9.6%
Office of International Science and Engineering	1.00	1.05	0.75	-0.30	-28.6%
Office of Polar Programs	26.24	26.24	26.24	-	-
Subtotal, Research and Related Activities	505.28	581.31	627.59	46.28	8.0%
Education and Human Resources	15.22	16.00	16.50	0.50	3.1%
Total, Cyberinfrastructure Funding	$520.50	$597.31	$644.09	$46.78	7.8%

Totals may not add due to rounding.

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 2008 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, from climate modeling, to design and production of materials with specialized properties at the atomic level, to simulation of black hole collisions and gravity wave sources, to exploring water systems, to the dynamics of monetary flows across national boundaries, will benefit from investments in cyberinfrastructure. Together with the growing availability and capability of cyberinfrastructure tools, this emerging cyberinfrastructure is revealing new knowledge and fundamental insights. 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 missions. Some highlights of NSF's FY 2008 investments, led by the designated activity, follow:

- Investments in a Plant Science Cyberinfrastructure Collaborative (PSCIC) will create a new type of organization – a cyberinfrastructure collaborative for plant science – that 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)

- As part of the Science of Science Policy initiative, there will be a focus on (a) new techniques of data extraction leading to new science and engineering indicators and (b) the development of virtual collaboratories where social and behavioral scientists will work with scientists in specific domains in seeking to understand how to evaluate investments made in those disciplines or problem areas. In addition, the Innovation and Organizational Change program will focus attention on the effects of innovative cyberinfrastructure on companies and scientific laboratories. (SBE)

- Support will be provided to initiate 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 at the Southern California Earthquake Center utilizing the TeraGrid to construct physics-based, realistic models of earthquake wave propagation in southern California and to begin the creation of a system of models that simulate the production, transport and deposition of sediments. Both projects will have considerable societal impacts. Finally, the Request supports a variety of projects that will develop or maintain community databases. (GEO)

- The National Center for Earth Surface Dynamics (NCED), an NSF Science and Technology Center, will examine the processes that contribute to land loss in the Mississippi Delta region. The project aims to leverage the vast array of petroleum seismic images and detailed knowledge of how deltaic systems function to examine how human engineering of the Mississippi River is contributing to the loss of wetlands. Such land loss is believed to have contributed significantly to damage caused by hurricane Katrina by removing barrier islands that normally would have helped dissipate some energy and resulted in a smaller storm surge. Such knowledge could be used in the future to plan geologically realistic restoration schemes for the delta and will be of great value in similar regions, both in the U.S. and abroad. (GEO)

- 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 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)

- Support will be provided for the continuing design and development of the Global Environment for Networking Innovations (GENI). GENI cyberinfrastructure is being developed by the computing community to support fundamental research in networking and distributed systems essential to fully inform redesign of the internet to incorporate security, robustness, and openness into technological innovation. GENI will bridge the gap between physical and virtual worlds by including mobile, wireless, and sensor networks and will enable control and management of other critical infrastructures. (CISE)

- Continuing investments will be made in the National STEM Digital Library (NSDL) to support a national resource of high-quality internet-based STEM educational content and to evolve NSDL as a component of cyberinfrastructure-enabled learning cooperatives for discovery-based learning. (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 (2007-2008) 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)

- Cyberinfrastructure support will also 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. 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. Improved communications and digitized data infrastructure to enhance exploitation of astrophysical data collected from South Pole Station during IPY will be implemented. In addition, support is provided for the engineering, operations and maintenance, and security of cyberinfrastructure systems. (OPP)

- 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)

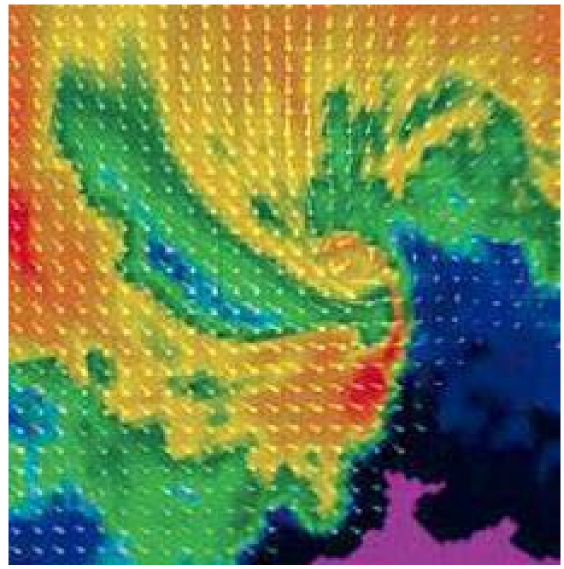
- Physical scientists and mathematicians create and use cyberinfrastructure, contributing to the development of high-performance computing, high-speed networks, data mining, software, and algorithms. Research at the frontiers of mathematical and physical sciences spans multiple length and time scales, and encompasses elementary particles; nuclear, atomic, and molecular properties, dynamics, and reactions; discovery of new materials and states of matter; complex multiscale systems and emergent behavior; supernovae and structures in the universe; and gravity waves and spacetime. Achieving a fundamental understanding of these diverse phenomena advances research both for itself and at the interfaces with the life sciences and the environment, contributes to the intellectual foundations of future cyberinfrastructure, and stimulates national competitiveness. (MPS)

- The innovative creation and use of cyberinfrastructure enables the prediction and discovery of new materials and new states of matter. It also enables the understanding of related phenomena. Software and algorithm development continue to be key to opening new approaches to reliably predict the properties of materials and phenomena, including those spanning many scales of length and time. The judicious and innovative use of cyberinfrastructure ensures community access to the best software for a wide range of research activities from the interpretation of experimental data to the visualization of simulation results for steering and interpretation to the creation of "virtual materials." Advances in the performance of information technology hardware bring previously computationally intractable or impractical methodologies or analyses within reach to tackle problems at the frontiers. (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

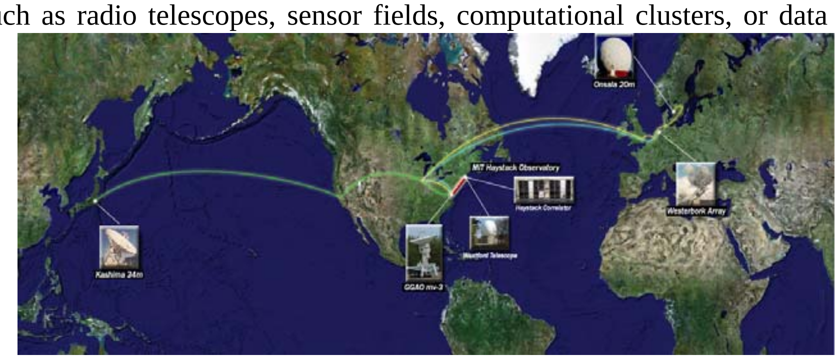
► **Reliable Tornado Prediction:** Current radar systems often miss tornadoes or confuse them with other short-lived weather events. Weather radar cannot accurately detect low-altitude activities such as tornadoes or distinguish between tornadoes and transient low-shear regions within other storms. So tornadoes must be confirmed visually. By the time a tornado has been identified, it is often too late to protect the people and structures in its path.



A simulation using the tornado detection algorithm. The swirling flow in the right–center of the image shows the tornado. *Credit: CASA ERC, University of Massachusetts-Amherst.*

Researchers at the Center for Collaborative Adaptive Sensing of the Atmosphere (CASA), an Engineering Research Center headquartered at the University of Massachusetts, Amherst, are focused on this issue. They have begun installing a network of low-power Doppler weather radar and have developed the first algorithms for dynamically detecting tornadoes close to the ground. The algorithms integrate wind field data from various radars and reconstruct the 3D wind field to define a tornado "signature". The algorithms identify tornadoes automatically and then provide information to optimize radar scanning to effectively track storms as they evolve. This allows weather forecasters to better predict the areas that will be affected by tornadoes, saving lives and property. (ENG)

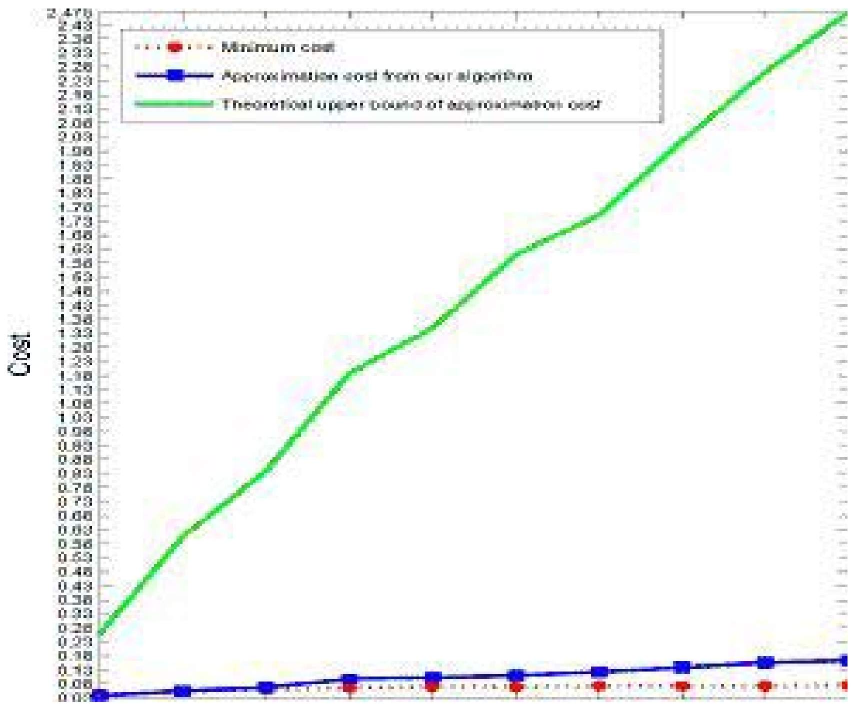
► **DRAGON (Dynamic Resource Allocation via GMPLS Optical Networks):** Scientists of the DRAGON networking research project have developed advanced network technologies that allow emerging "e-science" applications to construct customized and dedicated networks linking scientific resources worldwide. Resources such as radio telescopes, sensor fields, computational clusters, or data repositories can be integrated into a single coherent tool available to an individual scientist or science team. The networks can be established in a matter of seconds, when and where needed, held for as long as needed, and then released when no longer required.



Linking radio telescopes for Very Long Baseline Interferometry studies.

DRAGON research combines pre-commercial technology, tiny mirrors etched into integrated circuits and used to switch very high-speed photonic telecommunications channels, and Generalized Multiprotocol Label Switching (GMPLS) technology, a set of software protocols that automate the set-up process for these channels. Reconfigurable optical add/drop multiplexers provide an automated means to reconfigure optical networks dynamically and efficiently. These capabilities have been refined as a direct result of the DRAGON collaboration and the resulting switching devices have been deployed in advanced research and commercial carrier networks worldwide. GMPLS protocols, while standardized, had not seen significant adoption until they were deployed, with extensions for the global environment, in the DRAGON network. Today, DRAGON GMPLS software is deployed in the U.S. and in Europe and Japan. Leveraging this reach, researchers at MIT Haystack Observatory link radio telescopes and correlation facilities in Massachusetts, Maryland, Europe and Japan for groundbreaking Very Long Baseline Interferometry studies. Other corporations, agencies, and universities actively and tangibly support DRAGON development. The visibility and viability of these technologies in a high-profile testbed has spurred development and/or deployment well beyond the initial collaborating organizations. The NSF DRAGON project is the seed around which the rest of the global lattice is crystallizing. (OCI)

► **Minimum Cost Sensor Wireless Coverage:** A wireless sensor network monitors a given area by using small sensing devices that communicate amongst themselves using wireless communication. To adequately sense the physical world, sensors are typically deployed in numbers ranging from tens-to-thousands. Now, to ensure complete monitoring at a minimal cost, computer scientists have devised a computational method to optimize sensor placement in the area to be covered.



Computer scientists developed a new method to ensure complete coverage in a wireless sensing field at minimum cost. *Credit: Jie Wang.*

Wireless sensor networks offer new solutions for the 3-dimensional monitoring of varied settings including rainforests, oil platforms, bridges, city blocks, and high-rise buildings. To adequately monitor a sensing field and build in a safety factor in the event a sensor in the network fails, several sensors may be needed for every viewing point. Accordingly, the new method also calculates when certain sensors can "sleep" in order to save energy and increase the working lifespan of sensors in the network. (CISE)