CYBERINFRASTRUCTURE

Science and engineering research and education have become increasingly compute- and data-intensive, as a result of the proliferation of digital tools and pervasive networks through which scientific resources, methodologies, data and outcomes are collected, generated, shared, or analyzed. The comprehensive infrastructure needed to capitalize on dramatic advances in information technology has been termed cyberinfrastructure. Cyberinfrastructure integrates hardware for computing, data and networks, digitally-enabled sensors, observatories, and experimental facilities, using an interoperable suite of software and services and tools. Investments in interdisciplinary teams and cyberinfrastructure professionals with expertise in algorithm and software development, system operations, and applications development are also essential to exploit the full power of cyberinfrastructure to create, disseminate, and preserve scientific data, information, and knowledge. The enormous growth in the availability and utility of cyberinfrastructure capabilities, both technology- and human-based, is increasing scholarly research productivity, 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.

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 origins of man, are revolutionizing our understanding of the major kingdoms of life, and are 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. In the future, U.S. leadership in science and engineering will increasingly depend upon our ability to tap into this growing reservoir of digitally-encoded scientific knowledge.

Cyberinfrastructure Funding

(Dollars in Millions)

	FY 2005 Actuals	FY 2006 Current Plan	FY 2007 Request	Change over FY 2006 Amount	Change over FY 2006 Percent
Biological Sciences	$77.00	$84.00	$90.50	$6.50	7.7%
Computer and Information Science and Engineering	45.32	63.00	68.00	5.00	7.9%
Engineering	52.00	52.00	54.00	2.00	3.8%
Geosciences	71.35	71.35	75.00	3.65	5.1%
Mathematical and Physical Sciences	56.52	59.30	63.56	4.26	7.2%
Social, Behavioral and Economic Sciences	20.39	20.54	20.54	-	-
Office of Cyberinfrastructure	123.28	127.12	182.42	55.30	43.5%
Office of International Science and Engineering	0.22	1.00	1.05	0.05	5.0%
Office of Polar Programs	25.38	26.24	26.24	-	-
Subtotal, Research and Related Activities	471.47	504.55	581.31	76.76	15.2%
Education and Human Resources	20.27	15.02	15.52	0.50	3.3%
Total, Cyberinfrastructure Funding	$491.74	$519.57	$596.83	$77.26	14.9%

Increases in the capability and affordability of cyberinfrastructure are permitting the creation of powerful research and education tools and services that enable discovery, learning, and innovation across a range of science and engineering disciplines:

- Environmental scientists and engineers are drawing upon cyberinfrastructure to investigate the complexity of our environment, from the molecular to the planetary scale. This multidisciplinary work requires the collection and analysis of large volumes of data, it requires experiments with computer models that in many cases depend upon the world's most advanced supercomputers, and it relies upon the collaboration of scientists and engineers from a wide range of disciplines.
- Earthquake engineers are accessing shake tables, reaction wall facilities, geotechnical centrifuges, tsunami wave tanks, and mobile field equipment that are integrated through a common cyberinfrastructure framework, allowing them to perform tele-observation and tele-operation of experiments; to publish to and make use of curated data repositories; to access computational resources and open-source analytical tools; and to use collaborative tools for experiment planning, execution, analysis, and publication.
- Plant biologists are using cyberinfrastructure tools developed to extract implicit genome information to reveal the structure and function of plant genes at levels from the molecular to the organismal. The new knowledge and insights gained from plant genomics will lead to new discoveries and conceptual advances in our understanding of the biology of plants, as well as to the broader impact of this new knowledge in applications relating to agriculture, natural resources, the environment, health, and plant-based industries.
- Computer scientists and engineers are conducting research on next-generation systems architectures that will enable future generations of cyberinfrastructure. Research advances will enable the development of cyberinfrastructure systems that, for example, monitor and collect information on such diverse subjects as plankton colonies, endangered species, soil and air contaminants, medical patients, and buildings, bridges, and other man-made structures. Across a wide range of applications, cyberinfrastructure systems promise to reveal previously unobservable scientific phenomena.

All programmatic Directorates and Offices support cyberinfrastructure. The Office of Cyberinfrastructure makes investments common to a broad range of science and engineering fields, promoting economies of scale and scope, and facilitating interoperability. Other programmatic Directorates and Offices make cyberinfrastructure investments necessary to meet their missions.

FY 2007 Areas of Emphasis:

- Support will be provided for the Protein Data Bank (PDB), the international repository and primary source for information about the structure of biological macromolecules, a key research resource and central component to our understanding of living systems.
- As recommended in the 2004 report of the High-End Computing Revitalization Task Force, funding for research on high performance computing architectures will leverage interagency coordination and collaboration activities.
- Acquisition of a leadership-class high performance computing system will begin in FY 2007, to support petascale simulations essential to progress in many science and engineering fields. As recommended by the High End Computing Revitalization Task Force, this acquisition will be conducted in close collaboration with other agencies with a stake in high performance computing.
- 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. All projects

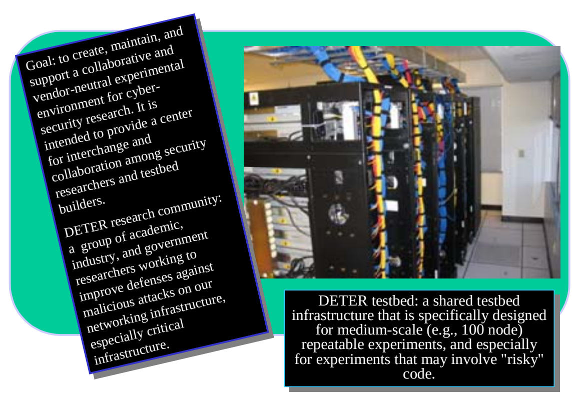
that produce data are expected to deposit their data at the Center. In cases where data is more appropriately deposited elsewhere, the Center keeps track of metadata. The Center ensures that data sets are published on-line and that they are broadly accessible to the community.

- An agency-wide programmatic activity, called CI-TEAM, will prepare thousands of scientists and engineers at the undergraduate, graduate, postdoctoral and faculty levels to use cyberinfrastructure in their research and education activities.
- Support for the National Radio Astronomy Observatory and the National Optical Astronomy Observatory will include enhanced efforts to make available long-term data archives for astronomical research and education.
- Substantial investments will be made in major social and behavioral science data collections, and will address issues such as confidentiality protections and means for securing worldwide, user-friendly access.
- 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 services to support learners at all levels.
- Projects that provide the nation's multidisciplinary computational science and engineering community with access to high performance computing resources and services will be supported, with an increasing investment in software infrastructure. High performance resources permit researchers to perform: lattice gauge calculations of properties of elementary particles; astrophysical simulations such as in the study of supernovae, accretion and jets, or galaxy dynamics; calculations of astrophysical processes using, for example, strong-field general relativity to compute gravitational wave signatures of colliding black holes; and first-principles calculations of the properties and dynamics of atomic, molecular, and nuclear systems using QED and the nucleon-nucleon force.
- Development of next-generation data management systems and tools will improve support of domain specific data types, such as sequences, pathways, and time series data.

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

► Collaborative Research: Cyber Defense Technology Experimental Research Network (DETER): Cybercrimes, including hacking, denial of service attacks, worms, viruses and other malicious attempts to corrupt or take control of networked computers, are a major threat to our nation's security. Yet it has been difficult to fully evaluate the effects of such attacks, or to test solutions to thwart these crimes, because researchers have not had access to suitable test environment.



DETER testbed: a shared testbed infrastructure that is specifically designed for medium-scale (e.g., 100 node) repeatable experiments, and especially for experiments that may involve "risky" code.

To address this challenge, NSF funds a vendor-neutral testbed called DETER: a collaborative project of the University of California at Berkeley, University of California at Davis, and University of Southern California. DETER provides academic, government and industrial scientists a safe environment to contain, model and analyze malicious attacks -- especially those that might result in catastrophic damage to public networks supporting critical infrastructure. In the isolated and controlled DETER environment, researchers can unleash “risky code” (such as worms, viruses, denial of service attacks and other targeted strikes), and then aggressively test counter-responses without fear of infecting public networks.