CYBER-ENABLED DISCOVERY AND INNOVATION (CDI)

Objective: *Broaden the Nation's capability for innovation by developing a new generation of computationally based discovery concepts and tools to deal with complex, data-rich, and interacting systems.*

Fundamental research has created the foundation of concepts, techniques, and tools on which today's digital age is based. The Nation's preeminence in information technology is widely recognized as an essential element in our competitiveness, both as a leading technology and in the way it has enabled innovation across many fields, starting with science and engineering and permeating all aspects of society.

Conduct of science and engineering has been revolutionized by the infusion of computational science and simulation in the traditional experimentation-observation-analysis-theory loop and by eliminating the geographic constraints for collaboration and experimentation. Ongoing investments by NSF and others have demonstrated the power of these approaches and led to our major investments in cyberinfrastructure that enable their application. Many important challenges are being tackled with these techniques and it is clear that this needs to be accelerated.

At the same time, looking beyond today's cyberinfrastructure, it is clear that new concepts and tools are needed to address the challenges posed by a world of petascale computers, massive data flows and databases, and an economy dependent on digitally enabled activity. Future competitiveness will clearly be dependent on scientific leadership and education that goes well beyond today's applications of the earlier concepts underlying the digital age.

CDI is a research initiative that will stimulate new capabilities to deal with these and similar challenges, including directly addressing NSF's "Cyberinfrastructure Vision for the 21st Century." The insights from this activity can be translated to other non-research systems and will be an essential contribution to the Nation's renewed focus on competitiveness.

Description

Scientists and engineers are faced with research problems that often have many complex internal feedback processes that defy simple analysis, or that must be studied at scales that are much different than processes occurring in nature. Some research problems require extensive complex networked observations. These challenges often need massive datasets for simulation, have heterogeneous data sources that must be linked, or generate massive, high-dimensional datasets from experiment or observation, and will soon be beyond today's capabilities.

This initiative aims to explore radically new concepts, approaches and tools at the intersection of computational and physical or biological worlds to address such challenges. It includes five distinct themes - knowledge extraction, interacting elements, computational experimentation, virtual environments, and education for computational discovery. New means of computational discovery will augment the traditional discovery-innovation loop with novel computational concepts to aid knowledge discovery, analysis and experimentation. This will accelerate the discovery of knowledge buried in massive datasets, creation of models to understand complex phenomenon, and understanding of rare events.

Areas of Emphasis

Knowledge extraction. Finding genomic expressions of cereal DNA, discovering new, fundamental particles predicted by the standard model, and finding new planets and proto-stars – all are like 'finding a needle in a haystack.' Knowledge extraction encompasses a variety of techniques – data mining, visualization, utilization of basic concepts from computation, geometry and topology – to help scientists and engineers find what is most important in the almost infinite amounts of data from sensors, telescopes, satellites, the media, the Internet, surveys, etc. Combining underlying conceptual ideas with heterogeneous data from multiple sources, high-bandwidth communications, and tera- to petascale computational power, scientists and engineers will be able to make sense of the massive volumes of data that bear on their grand challenges and address one of the most daunting challenges of the present century. This theme will improve across current techniques, which are insufficient for these challenges.

Complex interactions. Analyzing the flow of electricity or information across the electric power grid or the Internet, describing protein folding and unfolding, and finding superposition principles for scaling from the quantum- to the nano- to the macro-scales are examples of grand challenges that require scientists and engineers to understand interacting systems. Such systems, ranging from particles to galaxies and from computer networks to societies, are at the heart of many science and engineering grand challenges, and their understanding and control are major sources of innovation. Key factors in such systems are the large number of interacting elements, non-linearity of interactions, and aggregate or emergent phenomena observed at certain scales. This theme will improve both forward (predictive) and inverse (deductive) capabilities in order to better understand nature, and be able to design, control and make decisions about complex systems.

Computational experimentation. We cannot generate a hurricane to see how it develops and progresses; use routine brain surgery to experiment on neural synapses in the brain; or rerun the "Big Bang" to see how the physics of the universe develops. Computational experimentation allows insight into complex systems by enabling the creation of a virtual description (algorithmic or computational) that can interact with elements from the real world. Simulation and other dynamic modeling techniques allow us to experiment with complex systems in ways that would be unimaginable in the real world, and to constrain our understanding of the system characteristics or underlying physical phenomena. Furthermore, it allows us to guide real world operations and experimentation in cases that have potential for unforeseen or extreme events. Research in this area will provide needed new modeling techniques ranging from mathematical formulations to multi-scale simulation techniques.

Virtual environments. Scheduling and operation of distributed facilities and sensor arrays, data extraction and analysis, international real time comparative analyses of global climate models, and injecting discovery and innovative environments in STEM learning and training - all require the use of virtual environments as important mechanisms to enhance discovery, learning, and innovation. Virtual environments enabled by cyberinfrastructure permit collaboration among diverse populations spread across geographic distances and time zones. This theme will develop new techniques for building and utilizing virtual environments, especially in the context of cyberinfrastructure.

Educating researchers and students in computational discovery is essential. Without explicit attention to this, the promise of new capability, as well as the translation of these capabilities into other segments of the economy, will not be realized. We will emphasize integrating computational discovery techniques into the basic education of all scientists and engineers as well as development of new techniques for use in all areas, not just science and engineering. Research on the role and impact of computational discovery in education, educational practice and learning in general is also important.

Special focus will be placed on the utilization of virtual environments and cyberinfrastructure in education at all levels. By enhancing human cognition/perception in dealing with complexity, computational tools provide an essential handle on workforce development for the 21st century.

Several NSF directorates and the Office of Cyberinfrastructure will participate in developing advanced applications of the new concepts and tools flowing from the basic objectives of this investment. It is fundamental work clearly appropriate for NSF. Work will be coordinated with other agencies via the NITRD Subcommittee and other relevant subparts of National Science and Technology Council. It is critical over the next five years to leverage on-going investments, such as cyberinfrastructure, and to prepare for the coming computational discovery challenges in all fields.

Funding

A five-year initiative with a first-year total of $51.98 million is needed, growing to at least $150.0 million by the third year. Growth of about $50.0 million per year for a full 5 years is suggested.

Long-term Funding for Cyber-enabled Discovery and Innovation

(Dollars in Millions)

FY 2008 Request	FY 2009	FY 2010	FY 2011	FY 2012
$51.98	$100.00	$150.00	$200.00	$250.00

It is clear that new concepts and tools are needed to address research at the frontier. With CDI, at the end of five years, we expect to have: **enhanced ability** to deal with research requiring petascale cyberinfrastructure, a strengthened **technical basis** for a new generation of computational discovery in all areas of S&E, and **significant progress** in educating computational discoverers.