CYBERINFRASTRUCTURE FRAMEWORK FOR 21ST CENTURY SCIENCE, ENGINEERING, AND EDUCATION (CIF21)

Overview

The Cyberinfrastructure Framework for 21st Century Science, Engineering, and Education (CIF21) investment accelerates and transforms the processes and outcomes of scientific discovery and innovation by providing advanced cyberinfrastructure that enables new functional capabilities in computational- and data-enabled science and engineering (CDS&E) across all disciplines.

Science, engineering, and education will be transformed by a comprehensive and scalable cyberinfrastructure that bridges diverse scientific communities and brings together theoretical, computational, experimental, and observational approaches. Large volumes of research data are being generated by scientific instruments, observing systems, surveys, mobile and embedded systems, as well as by publications, experiments, simulations, evaluations, and analyses. In addition, scientists, using thousands of distributed scientific instruments, such as gene sequencers and sensors, are generating many more small data archives – creating the long-tail of science – that today generate heterogeneous data sets at an unprecedented rate. The scientific drivers for cyberinfrastructure can be seen in the complex research problems that require advanced computational models, methods, and algorithms, including innovative, robust, and sustainable software that turn raw data into knowledge and action. CIF21 is a portfolio of activities that leverage ongoing cyberinfrastructure investments across NSF by coordinating and deploying common approaches and components to manage data and provide computational support to all areas of science and engineering while also developing new multidisciplinary research communities.

Total Funding for CIF21

(Dollars in Millions)		
FY 2013 Actual	FY 2014 Estimate	FY 2015 Request
$109.13	$145.41	$124.75

Goals

The goal of CIF21 is for science and engineering research and education communities to gain new insights and investigate ever broader and more complex research questions through the use of advanced computing systems, innovative computational approaches, data systems and repositories, major research instruments, visualization systems, and other analytic tools and techniques that together comprise cyberinfrastructure. CIF21's cross-community and multi-pronged approach has been directing these investments along a path toward a comprehensive, integrated, sustainable, and secure cyberinfrastructure that accelerates research and education through new functional capabilities driven by recent transformations in computational- and data-intensive science and engineering. CIF21 has three specific goals: 1) support foundational data and computational research, including use-inspired demonstrations; 2) enable world-class data, computation, and digital capabilities and services; and 3) ensure long-term sustainability and future growth for advanced cyberinfrastructure through building research communities with the right skills and breadth to be able to contribute to the many new, cross-cutting research domains that are now emerging.

Approach

The vision of CIF21 is to catalyze new thinking, paradigms, and practices in science and engineering by fostering a pervasive cyberinfrastructure that enables research at unprecedented scales, complexity, resolution, and accuracy. This cyberinfrastructure aims to integrate and coordinate computation, data, and experiments in novel ways, nationally and internationally.

Organizational Structure. The CIF21 organizational structure employs four interrelated groups to ensure that CIF21 continues to build upon NSF's history of providing leadership in the design, development, and use of the cyberinfrastructure required to transform science, engineering, and education in the 21st century.

- The Division of Advanced Cyberinfrastructure (ACI) in the Directorate for Computer and Information Science and Engineering (CISE) provides leadership for CIF21 activities, including developing coordinated CIF21 programs and solicitations and identifying common approaches for a scalable, comprehensive cyberinfrastructure. The other CISE divisions focus on foundational science to advance critical techniques and technologies that address challenges in data management, data analytics and scientific discovery processes that transcend disciplines.
- The CIF21 Steering Committee of assistant directors and office heads provides oversight and advice on strategic directions and programs for CIF21.
- The Cyberinfrastructure Coordination and Leadership Group (CLG) coordinates and manages CIF21 programs across NSF, including developing solicitation guidance for common CIF21 programs, coordinating common CIF21 activities, and developing and maintaining an investment roadmap.
- The NSF Advisory Committee on Cyberinfrastructure (ACCI) reviews cyberinfrastructure activities and programs across all of NSF, and provides advice and strategic feedback on NSF plans and existing efforts.

Scope. To guide the development of CIF21, the ACCI produced a set of six reports and recommendations for cyberinfrastructure.[1] These reports and recommendations have been critical in identifying new approaches and capabilities required to advance data,[2] computing infrastructure, software,[3] and workforce development for CIF21. In 2012, the Administration's Big Data Research and Development Initiative[4] became a centerpiece for CIF21, focusing research and development on new capabilities for data-intensive and data-enabled science. The Advanced Computing Infrastructure Strategic Plan,[5] published in 2012, focused on NSF leadership in creating and deploying a comprehensive portfolio of advanced computing infrastructure to accelerate the pace of discovery. Along with ongoing focused workshops and events, these reports and initiatives help to define and prioritize programs and activities within the CIF21 framework.

CIF21 uses a combination of solicitations, Dear Colleague Letters (DCLs), and focused workshops to fund the research, development, and deployment of cyberinfrastructure and related applications. Partnerships with industry are emphasized, especially as disruptive technologies change how a technology or approach should be used to support science. The ubiquity of cyberinfrastructure requires partnerships and joint collaborations with other federal agencies and international groups. Principal Investigator (PI) meetings, conferences, and Ideas Labs will be used to reach out to new communities of researchers and educators.

---

[1] NSF Advisory Committee for Cyberinfrastructure: www.nsf.gov/od/oci/taskforces/index.jsp
[2] A Vision and Strategy for Data in Science, Engineering and Education: www.nsf.gov/od/oci/cif21/DataVision2012.pdf
[3] Software for Science and Engineering; www.nsf.gov/od/oci/taskforces/index.jsp
[4] Administration Big Data Initiative: www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_press_release_final_2.pdf
[5] Cyberinfrastructure for 21st Century Science and Engineering: Advanced Computing Infrastructure Vision and Strategic Plan

Investment Framework

CIF21 Funding by Directorate

(Dollars in Millions)

Directorate/Office	FY 2013 Actual	FY 2014 Estimate	FY 2015 Request
Biological Sciences	$3.75	$6.16	$3.75
Computer and Information Science and Engineering	57.03	85.00	80.00
Education and Human Resources	-	-	2.50
Engineering	7.00	12.00	10.00
Geosciences	10.25	15.25	11.00
Mathematical and Physical Sciences	25.60	21.00	11.50
Social, Behavioral, and Economic Sciences	4.50	6.00	6.00
International and Integrative Activities	1.00	-	-
Total	**$109.13**	**$145.41**	**$124.75**

Totals may not add due to rounding.

FY 2013 – FY 2014

CIF21 began by leveraging existing software, data, and advanced computing programs in ACI. In FY 2012 – FY 2013, NSF made 24 Big Data "Small" awards and eight Big Data "Mid-Scale" awards for foundational techniques and technologies. Based on community response to the first Big Data solicitation, the second solicitation in FY 2014 was expanded to ensure that transformative big data science and engineering projects were submitted across all research areas. It is expected that approximately 20 awards will be made. In FY 2013, the Data Infrastructure Building Blocks (DIBBs) program made eight awards. Based on the response to the DIBBS solicitation, it was expanded in FY 2014 to include participation by all NSF directorates. It is expected that up to four pilot projects and two implementation awards will be funded in FY 2014.

In the EarthCube program, a collaboration between ACI and the Directorate for Geosciences (GEO), approximately 15 awards were made in FY 2013, including awards for Governance (i.e., the processes, structure and organizational elements that determine how stakeholders have a say and make decisions), Building Blocks (i.e., demonstrations of software, middleware, techniques to serve and deliver data, etc.), and EarthCube Research Coordination Networks (RCNs), which foster new collaborations that lead to improved scientific outcomes. In FY 2014, a new solicitation will encourage communities of practice and the leveraging of existing resources. Several workshops and meetings will be held to increase participation and integration of new domains and research groups. New awards will be made to expand the scale and scope of efforts, as well as to increase the level and capabilities for sharing data and resources across the Geosciences research communities. Additionally, EarthCube workshops were convened to develop new multidisciplinary research data communities.

In FY 2013, the Directorate for Mathematical and Physical Sciences (MPS) made 45 CDS&E awards and the Directorate for Engineering (ENG) made nine awards. In addition to the program description, a crosscutting web page informed the PI community about this program, as well as its relationship with complementary programs such as Physics at the Information Frontier in the Division of Physics (PHY), Computational Mathematics and Statistics in the Division of Mathematical Sciences (DMS), and Computational Chemistry in the Division of Chemistry (CHE). Coordination of these programs will continue to be managed through the CIF21 CDS&E Working Group. In FY 2014, it is expected that there will be a unified program description across CISE, ENG, and MPS for CDS&E, which will invest in new

efforts and approaches for simulation and modeling as well as prototypes for collaborating across domains.

In FY 2013, approximately 85 awards were made in the Software Infrastructure for Sustained Innovation (SI^2) program. It is expected that 20 Scientific Software Elements (SSE) awards will be made in 2014. A PI meeting was held in 2013 and one is planned for 2014.

In FY 2013, the DataWay effort was redefined to coordinate the development of an integrated data infrastructure, by working with existing cyberinfrastructure and with ongoing data-related programs to foster the development of community linkages, groups, and interoperable frameworks. For example, lessons learned from EarthCube were conveyed to the Materials Genome Initiative (MGI). In FY 2014, data-related programs are being classified and potential synergies identified. As driven by community needs and interest, awards may be made to support workshops and develop early pilots, prototypes, and best practices.

The Integrated Graduate Education and Research Traineeship (IGERT)-CIF21 Dear Colleague Letter (DCL) and solicitation resulted in three FY 2013 awards. In FY 2014 this effort has been expanded in the new NSF Research Traineeship (NRT) program, which encourages U.S. universities and colleges to submit proposals focused on interdisciplinary data-intensive science traineeships.

FY 2015 Request
In FY 2015, NSF will continue to invest in the suite of programs developed over the previous three years in foundational research, cyberinfrastructure and community building, and education and workforce development.

Foundational research programs include the following:
- The Big Data program will post its third solicitation with a focus on foundational research, tool development, and the needs of additional research domains and communities. Based on the Big Data portfolio, results achieved to date, and a needs analysis of data infrastructure at scale, research solicitations and investments in prototypes will continue to be further modified.
- CDS&E will issue a revised program description based on the evolving needs of the community. Additional prototype and proof-of-concept approaches for CDS&E will be developed and other federal agencies will be encouraged to participate. Based on the results of portfolio analysis, the program will be restructured to more aggressively address issues of scope and scale.

Cyberinfrastructure and community-building programs include the following:
- DIBBS will post a new solicitation and make new technical pilot project and early implementation awards. Efforts will be expanded to include more participation and collaboration with academic institutions, as well as with other agencies and international partners. Based on portfolio analysis, the program will be restructured to support more long-term data needs for research and education.
- The SI^2 program will continue issuing new awards in the Scientific Software Integration (SSI) and SSE tracks. These new awardees will work with existing Scientific Software Innovation Institutes.
- DataWay will work toward cross-disciplinary sharing and synergy among NSF data-related programs, and external community-building activities will continue.
- EarthCube will issue a solicitation focused on critical components and conceptual governance. One goal will be to further integrate other federal agencies in this activity. Workshops and other community events that broaden the base of users as well as the breadth of science will be supported. Furthermore, integration of components and coordination across projects will start.

Education and workforce development under CIF21 is accomplished principally through the NSF Research Traineeship (NRT) program.

NRT will encourage the development of new potentially transformative and scalable models of STEM graduate training to ensure graduate students develop the skills, knowledge, and competencies needed to pursue a range of careers. It will include a focus on interdisciplinary data-intensive science in FY 2015.

FY 2016 and Beyond
NSF will identify and support further development of common elements across discipline-specific cyberinfrastructure. These efforts will also support integration with international investments in cyberinfrastructure.

Foundational research efforts in Big Data will be broadened to address issues of scale and sustainability, especially for the use and long-term management of research data. This will include both the development and deployment of new pilots and prototypes.

CDS&E will increasingly address issues associated with expanding both the base of users as well as the participation of new domains and disciplines. It is expected that the availability of new tools and technologies resulting from new sensors, mobile devices, and data clouds will provide new opportunities for research communities that previously had limited access to or need for advanced computational infrastructure.

DIBBS will accelerate efforts to address issues of extensibility across multiple domains, including pilots and prototypes involving joint implementations across Major Research Equipment and Facilities Construction (MREFC) projects. These efforts will include common approaches to data access, use, storage, and curation.

EarthCube will update its solicitations based on gap analyses. Based on the results from the governance and cyberinfrastructure communities, EarthCube will develop programs that begin to bridge and support multiple research communities. This will include integrating existing pilots and prototypes with a focus on issues of scale, including development and deployment of common approaches and structures.

The SI^2 program will begin to focus on common approaches and issues across multiple institutes and software projects. This will include integration and coordination of development and deployment, and coordination with major projects and awards including Science and Technology Centers (STC), Engineering Research Centers (ERC), and MREFC projects.

NSF will continue to support CIF21-related NRT awards, as well as to develop additional Big Data education programs to address and support the participation of new research communities.

Evaluation Framework
NSF will deploy a variety of tools to evaluate the scientific and educational impact and progress of its various CIF21 programs. The CIF21 Steering Committee and the CIF21 CLG will consider a matrix of assessment methods and measures, including incorporating input and guidance from the NSF ACCI. In the short-term, these groups will review the CIF21 portfolio and identify common elements and goals across the various programs. In the long-term, NSF will engage an external organization to conduct an assessment of CIF21 research, infrastructure, and education investments and outcomes.