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

Overview

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

Future science, engineering, and education endeavors will be transformed by a comprehensive and scalable cyberinfrastructure that bridges diverse scientific communities and brings theoretical, computational, experimental, and observational approaches together. 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. CIF21 is a portfolio of activities to provide integrated cyber resources that will enable multidisciplinary research opportunities in all areas of science and engineering. It will leverage ongoing cyberinfrastructure investments across NSF by coordinating and deploying common approaches and components to manage data, provide computational support, and develop new multidisciplinary research communities.

Total Funding for CIF21

(Dollars in Millions)

FY 2012 Actual	FY 2012 Enacted/ Annualized FY 2013 CR	FY 2014 Request
$91.23	**$78.00**	**$155.47**

Goals

The goal of CIF21 is to accelerate the deployment and use of advanced cyberinfrastructure facilities and capabilities to support all areas of science, engineering, and education. CIF21 has three primary components: 1) establishment of a national data infrastructure; 2) development of new computational and data-intensive capabilities; and 3) community building and workforce development.

Approach

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

Organizational. 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 Directorate for Computer and Information Science and Engineering (CISE), in particular the Division of Advanced Cyberinfrastructure (ACI) (formerly the Office of Cyberinfrastructure), provides leadership for CIF21 activities, including developing coordinated CIF21 programs and

solicitations and identifying common approaches for a scalable, comprehensive cyberinfrastructure.

- The CIF21 Steering Committee of assistant directors and office heads provides oversight and advice on strategic directions and programs for CIF21.
- The CIF21 Cyberinfrastructure Leadership Group (CLG) coordinates and manages CIF21 programs across NSF, including developing solicitation guidance for common CIF21 programs, coordinating common CIF21 activities, developing and maintaining an investment roadmap, and providing planning and budgeting for CIF21.
- 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 approach 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,[3] software,[4] and workforce development for CIF21. These reports, along with on-going focused workshops and events, help to further define and prioritize programs and activities within the CIF21 framework.

CIF21 uses a combination of solicitations, Dear Colleague Letters (DCL), programs, 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. Workshops, conferences, and focused IdeaLabs will be used to reach out to new communities of researchers and educators.

Investment Framework

CIF21 Funding by Directorate

(Dollars in Millions)

Directorate/Office	FY 2012 Actual	FY 2012 Enacted/ Annualized FY 2013 CR	FY 2014 Request
Biological Sciences	$2.00	$2.00	$6.50
Computer and Information Science and Engineering	47.94	35.00	90.67
Engineering	3.70	5.00	12.00
Geosciences	4.49	8.00	16.50
Mathematical and Physical Sciences	27.60	11.50	22.30
Social, Behavioral, and Economic Sciences	5.50	5.50	7.50
International and Integrative Activities	-	11.00	-
Total	**$91.23**	**$78.00**	**$155.47**

Totals may not add due to rounding.

[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] Advanced Computing Infrastructure: Vision and Strategic Plan: www.nsf.gov/publications/pub_summ.jsp?ods_key=nsf12051
[4] Software for Science and Engineering; www.nsf.gov/od/oci/taskforces/index.jsp

FY 2012 – FY 2013

Multiple CIF21 solicitations and programs involving Biological Sciences (BIO), Computer and Information Science and Engineering (CISE), Engineering (ENG), Geological Sciences (GEO), Mathematical and Physical Sciences (MPS), and Social, Behavioral, and Economic Sciences (SBE) were issued in FY 2012. These solicitations focused on data-enabled science, software development, community research networks, new computational infrastructure, and access and connections to facilities.

In FY 2013, investments in CIF21 expanded both the scope and activities across the Foundation to support science, engineering, and education, with continued consideration of the ACCI Task Force reports. The National Big Data Research and Development Initiative remained a CIF21 centerpiece, focusing on the research and development of new capabilities for data-intensive and data-enabled science to create actionable information that leads to timely and more informed decisions. The NSF Big Data program invested in four key areas: innovation and new foundational research, cyberinfrastructure, community building, and education and workforce development. The science and engineering research and education scope of the Big Data solicitation expanded to include topics aligned with big data and new capabilities dedicated to creating large-scale, next-generation data resources and relevant analytic techniques to advance fundamental research in scientific fields supported by the SBE and Education and Human Resources (EHR) Directorates.

A unique virtual program in CDS&E was established by ACI, ENG, and MPS. It is rooted in the individual divisions and is also coordinated across directorates, responding to both disciplinary and cross-disciplinary needs.

CIF21 program activities fostered broadening of the cyberinfrastructure development and research communities in FY 2013. In particular, software programs and data activities expanded to include participation by new domains and included joint international software and data collaborations with the European Union, Australia, and China. Through the ACI-supported XSEDE (eXtreme Science and Engineering Discovery and Education) environment, new CIF21 research communities have access to advanced computational infrastructure, including two new significant computational resources that became operational in FY 2013: the University of Illinois at Urbana-Champaign Blue Waters project, and the Texas Advanced Computational Center's new Stampede project.

New research communities were also established to address multidisciplinary research challenges that are emerging as a result of new data and cyberinfrastructure capabilities. These include the EarthCube communities in the geosciences, and a new community effort to develop integrated data management infrastructure in the mathematical and physical sciences, which will create communities around grand challenge problems. The Building Community and Capacity solicitation supports the development of new research communities for SBE and EHR.

FY 2014 Request

In FY 2014, CIF21 will make the following investments:

- The National Data Infrastructure program will be expanded in scope to address issues associated with open access and will invest in one or two pilots to address management and use of multidisciplinary data.
- The Big Data solicitation will be expanded to include additional themes critical to the missions of NSF directorates; workshops will be convened to develop R&D pathways for the next five to ten years, including exploring partnerships and participation with more agencies and industry.
- CDS&E program activities will be expanded to include new efforts and approaches for simulation and modeling; new prototypes in specific domains will be developed with an emphasis upon collaboration across disciplines. Based on the response to FY 2013's consolidated solicitation, additional changes will be made to the program with a specific focus to scale CDS&E efforts.

- New data conceptualization and data pilot awards will be made in collaboration with additional directorates and offices through the Data Infrastructure Building Blocks (DIBBs) program and data coordination pilots in development in BIO and GEO. The data program will broaden the base and use of advanced computational services and capabilities across disciplines that have not used advanced cyberinfrastructure resources before; and it will explore new approaches to data-intensive computational resources, including a mix of clouds, data centers, and distributed computing systems.
- Conceptualization awards, along with early pilots, prototypes, and best practices approaches, will be made to several communities to develop data and software; to resolve governance issues; and develop requirements to support multidisciplinary communities.
- Additional community-building awards will focus on the development of new research communities, including next generation data resources and access.
- The CIF21 track within the NSF Research Traineeships (NRT) will address the need to educate and support the next generation of researchers able to address fundamental challenges in: 1) core techniques and technologies for advancing big data science and engineering; 2) analyzing and dealing with challenging CDS&E problems; and 3) researching, providing, and using cyberinfrastructure that makes cutting-edge CDS&E research possible in any and all disciplines.
- EarthCube will expand its support for the development of community-guided cyberinfrastructure to integrate data into a framework that will expedite the delivery of geoscience knowledge to the science and engineering enterprise.

FY 2015 – Beyond

To further accelerate scientific discovery and innovation in FY 2015 and beyond, it is essential to develop a national data infrastructure ecosystem that provides new capabilities and functionalities to support a broad spectrum of science, engineering, and education users and communities. These activities include provision of data access and data exchange; development of at-scale multi-disciplinary data pilots and prototypes; and continued development of new algorithms, tools, and software, as well as expanding both the scope and depth of new research communities and the future CDS&E workforce.

- National Data Infrastructure efforts will be expanded to include more participation and collaboration across all NSF directorates, with an emphasis on coordination and common approaches to leverage investments and ensure interoperability. The efforts will also include development of new partnerships with campuses, other federal agencies, and international partners.
- Based on portfolio analyses, cyberinfrastructure programs and activities will be restructured to more fully support long-term data and computational needs for research and education, including open access, curation, preservation, and development of expertise to meet data-intensive science across all disciplines.
- Foundational research efforts will be expanded to include additional analysis and discovery tools, software and computational capabilities, including new domains and research communities. One focus will be on the transition of data analysis tools into practice, especially those that have applicability across multiple domains and are based on analyses of the Big Data and National Data Infrastructure portfolios and the results achieved to date. Another focus will be on new computational approaches for science that bridge across multiple domains, especially supporting long-tail science and users that have not been able to effectively leverage CDS&E resources. This will also include broadening and expanding the base and diversity of users (e.g., faculty, students, and the public) to take advantage of new capabilities in data and computation.
- Additional prototype and proof-of-concept approaches for CDS&E will be developed; involvement from other federal agencies, such as the Department of Energy, the National Institutes of Health, the National Aeronautics and Space Administration, and the U.S. Geological Survey will be pursued. Based on continued community input from all scientific domains and cyberinfrastructure communities, solicitations and programs will be revised and updated to support and develop new research communities, including effective use of new technologies, and the rapidly developing

national data infrastructure.

- The CIF21 track in NRT will be expanded to support the education of the next generation of researchers able to address fundamental challenges in advancing big data science and engineering.
- EarthCube will continue its support for the development of community-guided cyberinfrastructure to integrate data into a framework that will expedite the delivery of geoscience knowledge to the science and engineering enterprise.

Evaluation Framework

NSF will deploy a variety of tools to evaluate the scientific and educational impact and progress of the various CIF21 programs. The CIF21 Steering Committee and the CIF21 Leadership Group will consider a matrix of assessment methods and measures, including incorporating input and guidance from the NSF Advisory Committee on Cyberinfrastructure. In the short-term, these groups will conduct portfolio analyses and identify metrics. In the long-term, NSF will engage an external organization to conduct an assessment of CIF21 research, infrastructure, and education investments and outcomes.