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

$143,060,000
+$14,100,000 / 10.9%

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 and supporting the use of advanced cyberinfrastructure that enables new functional capabilities in computational and data-enabled science and engineering across all disciplines.

Science, engineering, and education continue to be transformed by increasingly 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. New and emerging data sources are also becoming available to the scientific community, through efforts such as digitizing collections and enhancing access to records and documents. In addition, scientists, using thousands of distributed scientific instruments, such as gene sequencers, sensors, and imaging devices, are generating many more small data archives and heterogeneous data sets at an unprecedented rate, creating the long tail of science, which is yet another data cyberinfrastructure challenge.

Complex scientific research problems, such as those under other NSF priority investments (e.g. Understanding the Brain (UtB) or Innovations at the Nexus of Food, Energy, and Water Systems (INFEWS)) 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 2014 Actual	FY 2015 Estimate	FY 2016 Request
$156.75	$128.96	$143.06

Goals

The overarching goals of CIF21 are:

- Support foundational research, and discovery and innovation; develop a deep mutual relationship between science and engineering users and developers of cyberinfrastructure to simultaneously advance new research practices and open transformative opportunities across all science and engineering fields;
- Enable data, computational, and digital capabilities and services by providing an integrated and scalable cyberinfrastructure that leverages existing and new components across all areas of science and engineering supported by NSF; and
- Ensure long-term sustainability and future growth for advanced cyberinfrastructure and CDS&E through community development, learning and workforce development, and transformation of practice.

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 five 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 NSF Advisory Committee for Cyberinfrastructure (ACCI) reviews cyberinfrastructure activities and programs across all of NSF, interacts with other NSF directorate advisory committees, and provides advice and strategic feedback on NSF plans and existing efforts.
- The Cyberinfrastructure Coordination and Leadership Group (CLG) coordinates and manages the broader NSF cyberinfrastructure investment portfolio, including CIF21 programs. This coordination and management includes developing solicitation guidance for common CIF21 programs, coordinating common CIF21 activities, and developing and maintaining an investment roadmap. CLG membership is designated by CIF21 Council members.
- The CIF21 Council of NSF assistant directors and office heads provides oversight and advice on strategic directions and programmatic scope for CIF21.
- The Advanced Cyberinfrastructure (ACI) division in the Computer and Information Science and Engineering (CISE) directorate provides leadership for CIF21 activities. This leadership includes developing coordinated CIF21 programs and solicitations and identifying common approaches for a scalable comprehensive cyberinfrastructure.
- As part of CIF21, the other CISE divisions and NSF directorates focus on foundational science, and engineering, as well as domain applications. This focus leverages cutting-edge cyberinfrastructure, and also advances critical techniques and technologies to address challenges in computational science and engineering, data management and analytics, and sustained software systems. For example, the Social, Behavioral, and Economic Sciences (SBE) directorate has a working group focused on development of user-friendly, large-scale, next-generation data resources and relevant analytical techniques to advance fundamental SBE research. A Geosciences (GEO) directorate working group oversees EarthCube, a community-driven data and knowledge environment for the geosciences. A long-standing working group in the Mathematical and Physical Sciences (MPS) directorate, with regular participation from other directorates including Engineering (ENG) and CISE, coordinates the Computational and Data-enabled Science and Engineering (CDS&E) program.

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 Big Data Research and Development Initiative[4] was highlighted in 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.

---

[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

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. The ubiquity of cyberinfrastructure requires partnerships and joint collaborations with industry, other federal agencies and international groups. Principal Investigator (PI) meetings, conferences, and workshops are used to reach out to new communities of researchers and educators.

Investment Framework

CIF21 Funding by Directorate

(Dollars in Millions)

Directorate	FY 2014 Actual	FY 2015 Estimate	FY 2016 Request
Biological Sciences	$6.16	$3.75	$8.39
Computer and Information Science and Engineering	85.00	84.21	84.21
Education and Human Resources	-	2.50	2.84
Engineering	12.00	10.00	10.00
Geosciences	15.25	11.00	14.21
Mathematical and Physical Sciences	32.34	11.50	16.15
Social, Behavioral, and Economic Sciences	6.00	6.00	7.26
Total	**$156.75**	**$128.96**	**$143.06**

Totals may not add due to rounding.

FY 2014 – FY 2015
Following the increased emphasis on domain-specific data intensive applications in FY 2014, the BIGDATA solicitation is being further expanded in FY 2015 to address issues of scope and scale. In FY 2015, NSF is holding a workshop on Big Data and replication of results. Reproducibility of results is an issue across the sciences, and the federal Networking and Information Technology Research and Development (NITRD) Big Data Senior Steering Group is spearheading development of this workshop to focus specifically on how Big Data can drive replicability and confirmation of hypotheses.

In FY 2014, the Data Infrastructure Building Blocks (DIBBs) solicitation was expanded to include participation by all the NSF directorates and the Office of International and Integrative Activities (OIIA). The seven NSF directorates and OIIA supported 16 pilot-demonstration and two early-implementation awards in FY 2014. During FY 2015, efforts in data infrastructure continue to focus on collaboration among all directorates and the formulation of a plan for developing a critical set of national-scale data archive pilots in FY 2016. DIBBs is increasing its scale and scope, emphasizing the value of sharing data beyond a specific institution to the wider science, engineering, and education communities. Investments in multi-campus and/or multi-institutional regional cyberinfrastructure sought to leverage high performance network paths among campuses, enabling integration of new data-focused services, capabilities, and resources to advance scientific discoveries, collaborations, and innovations.

In FY 2014, the Computational and Data-Enabled Science and Engineering (CDS&E) program, led by MPS, ENG, and CISE, was expanded to include new efforts and approaches for simulation and modeling, along with a specific focus on scaling. Prototypes in specific domains were developed with an emphasis upon collaboration across disciplines. The BIO and GEO directorates also participated informally in the proposal and review processes. A total of 71 CDS&E awards were made in FY 2014, half of which were co-funded between two or more divisions in the participating directorates. In FY 2015, additional prototype and proof-of-concept approaches for CDS&E will be developed. More involvement from all NSF directorates and other federal agencies will be pursued. Based on the results of continuing portfolio

analysis, the program will be structured to address emerging issues of scope and scale.

In FY 2014, EarthCube supported the planning phase for a science community-based governance structure, encouraging the participation and interaction of new geoscientists across domains, cyberinfrastructure researchers, and other agencies, as well as the private sector. The EarthCube All-Hands Meeting brought awardees and the broader EarthCube community together to make decisions on the governance framework, on mechanisms to coordinate projects and find common goals, on ways to articulate science drivers for infrastructure development, and on ways to approach the development of an integrated cyberinfrastructure architecture. In FY 2015, EarthCube will continue to support community coordination activities, such as research coordination networks, workshops, and other engagement mechanisms. These activities will broaden the base of users as well as the breadth of science conducted via EarthCube. Early efforts include development of common approaches including some level of integration and coordination across projects.

The Software Institutes for Sustained Innovation SI^2 program (led by CISE/ACI) issued a Dear Colleague Letter (DCL) jointly with the Science of Science and Innovation Policy (SciSIP) program (led by SBE) in FY 2014, calling for EArly-concept Grants for Exploratory Research (EAGER) and workshop proposals related to norms and practices for software and data citation and attribution, which could in turn incentivize sustained software in the long term. SI^2 continued to support software reuse across NSF. The program also began a community activity called "Working towards Sustainable Software for Science: Practice and Experiences (WSSSPE)," with two workshops at technical conferences, including the Supercomputing Conference 2014 (SC14), and a journal Special Issue with 19 papers. In FY 2015, SI^2 began working more closely with DIBBs to encourage proposals that have both software and data elements. SI^2 is also developing a supplement mechanism through which PIs can apply for additional funds to support creating and supporting open source software developed through general (non-SI^2) projects. These supplements are co-funded by the SI^2 program and the funders of the original award. This mechanism is replacing much of the existing software reuse activity funded as part of CIF21. Additionally, the SI^2 program will hold a PI meeting for grantees in FY 2015.

In FY 2014, NSF transitioned from IGERT to the NSF Research Traineeship (NRT) program. This new graduate education initiative is designed to encourage the development of new, potentially transformative, and scalable models for STEM graduate training that ensure that graduate students develop the skills, knowledge, and competencies needed to pursue a range of STEM careers. In this first year of the program, NRT specified one priority theme to address fundamental challenges advancing computation- and data-enabled science and engineering. This was directly aligned with CIF21. Based on the response to NRT in FY 2014, it is anticipated that this theme will continue in FY 2015, and a new set of five-year NRT awards is planned in this space.

In FY 2014, a revised and updated version of the Building Community and Capacity for Data-Intensive Research in the Social, Behavioral, and Economic Sciences and in Education and Human Resources (BCC-SBE/EHR) solicitation was released, and a set of awards were funded with a focus on community involvement in the design of the infrastructure. In FY 2015, a new solicitation, Resource Implementations for Data Intensive Research in the Social Behavioral and Economic Sciences (RIDIR), was issued for user-friendly large-scale next-generation data resources and relevant analytic techniques to advance fundamental research in SBE areas of study. Successful proposals will construct databases and/or relevant analytic techniques and produce a finished product to enable new types of data-intensive research in multiple disciplines or fields. The databases/techniques should have significant impacts across multiple fields by enabling new types of data-intensive research that includes, but is not necessarily limited to, the SBE sciences.

Also in FY 2014, ENG, in collaboration with the Air Force Office of Scientific Research (AFOSR), issued a DCL seeking EAGER proposals with the aim of transforming the ability to understand, manage, and control the operation of complex, multi-entity natural or engineered systems through innovative approaches that consider new dimensions in Big Data, Big Computing, and a symbiotic combination of Data and Computing. NSF and AFOSR supported highly innovative projects in their early stages that sought to address unique challenges and identify fruitful directions for analytics to transform engineering and scientific practice across various relevant disciplines and scales.

FY 2016 Request

- The DIBBs program will build on existing community development activities in and across the directorates and offices as well as the past three years of experiences in the program. As part of this, NSF directorates will expand the scale and scope of directorate and multi-directorate Data Science Pilots, reflecting maturity of the pilots initiated in previous years through co-investments with CISE/ACI. Based on individual directorates' investment priorities, expanded scope could include data reproducibility; interoperability of specific research data; sustainability plans; data policy and governance; security, privacy, data integrity and trustworthiness; exploration of innovative economic/operating models for archiving and curation; and learning and workforce development. Depending on scientific and engineering priorities, increased scale toward national-level and multi-agency activity will be explored.
- Foundational research efforts via the BIGDATA solicitation will be broadened to address not only scalability and the exploration of new data science capabilities, but also the interface between data sciences, multiple stakeholders, and sustainability, especially in use and long-term management of research data. This will complement both the development and deployment of new pilots and prototypes.
- CDS&E efforts will address issues associated with expanding both the base of researchers as well as the participation of new domains and disciplines. The availability of new tools and technologies resulting from research and infrastructure advances in computation and data will provide new opportunities for communities that have had limited access and use of research data and advanced computing infrastructure.
- Based on the results from the governance and cyberinfrastructure communities, EarthCube will develop programs that begin to bridge and support multiple communities. This will include integrating existing pilots and prototypes to address issues of scale and research support. In addition, EarthCube will focus on at-scale issues including development and deployment of common approaches and structures. This will include development of common tools, data systems, and virtual organizations to support the emerging geoscience communities and to coordinate regional and national cyberinfrastructure facilities.
- The SI^2 program will continue to expand its joint activities with DIBBs and related cyberinfrastructure, such as security and networking. As it continues to develop software sustainability models for scientific software, it will also begin to focus on common approaches and issues across multiple institutions and software projects including integration and coordination of development and deployment. This will include a focus on software infrastructure for major projects and awards, including Science and Technology Centers (STCs), Engineering Research Centers (ERCs), and Major Research Equipment and Facilities Construction (MREFC) projects.
- Through the NRT program, NSF will continue to support the training of STEM graduate students and the development of transformative and scalable models for STEM graduate education in computation- and data-enabled science and engineering.
- The EHR directorate will emphasize the development of communities of researchers and data scientists with the capacity to pursue graduate education research questions in technology-rich learning environments, and address complex issues of privacy and data sharing. This will occur through the EHR Core Research (ECR) program and other programs as appropriate.

- NSF will continue to support development of user-friendly, large-scale, next-generation data resources and relevant analytical techniques to advance fundamental SBE research through increased investment in the RIDIR program.

FY 2017

Foundational research within the CISE-led Big Data program will lead to use-inspired approaches, techniques, and paradigms for collecting, managing, and analyzing Big Data across all areas of science and engineering as well as enhancing analytical capabilities in health, government, policy, and other areas of national priority.

NSF support for Data Infrastructure for Research and Education through the DIBBs program will build on data and computational science foundational advances guided by research and education priorities within and across directorates. Individual research community considerations of accessibility, reproducibility, efficient sustainability, policy, confidentiality and privacy in addition to a range of scalability needs will result in a mature, extensible, flexible data ecosystem tailored to the nation's research priorities.

Today's science requires the highest levels of computational science and engineering and associated expertise. CDS&E will enable researchers to address the most challenging scientific problems of our time. It will help to create computational resources and a sustainable career path for computational and data scientists so that they become a permanent part of the scientific workforce.

EarthCube will produce an integrated framework of cyberinfrastructure for the open and easy discovery and access of geoscience data, software and services, information, and computational resources. It will also facilitate the coordination of geoscience data and software facilities to better serve the science requirements of the entire research community. The academic geoscience community will gain a stable venue to coordinate future infrastructure advances as science drivers and technologies change. Additionally, EarthCube will enable new transformative geoscience research and education. This will be achieved through the improved ability to access and analyze geoscience data, using effective software, models, and analytical tools that can simulate and examine complex and interrelated Earth processes.

SI^2 will lead to an increase in shared software for use across many scientific fields. It will also increase the incentive to develop such shared software without direct NSF support, which will make the concept of software as infrastructure sustainable, and this will result in increased and improved science and engineering research. SI^2 projects will continue to develop software for use by broad communities, with specific metrics based on the size of the user communities and their science and engineering research productivity. Those metrics support the goal of creating long-term, sustainable impact.

Prior CIF21 investments NRT investments will continue to support graduate students in computational- and data-enabled science and engineering, and to lead to the development and testing of potentially transformative and scalable models for graduate education more broadly. These investments will continue to facilitate broad training of computational- and data-enabled science and engineering graduate students, including the development of technical and professional skills that will prepare them for successful careers in various settings within or outside academe.

EHR will provide leadership in the use of institutional data, technology-based learning data, and issues of privacy for learning data with NSF and the community. EHR anticipates initiatives in FY 2017 that will involve large-scale use of education and learning data for the improvement of education at scale in the undergraduate and K-12 levels, and will conduct workshops in FY 2015 and 2016 in preparation for these activities.

The projects funded through the RIDIR solicitation are expected by FY 2017 to lead toward (1) new large-scale databases, substantial expansion or revision of extant databases, and/or the merging of extant databases that will enable data-intensive SBE research (i.e., research involving data resources that are well beyond the storage requirements, computational intensiveness or complexity that is currently typical in SBE areas of research); and/or (2) analytic tool(s) that would serve to enhance database use to address significant SBE research questions. RIDIR will enhance the ability to conduct data-intensive research that will address broad, important, fundamental SBE research questions.

Evaluation Framework
NSF has deployed tools to evaluate the scientific and educational impact and progress of the various CIF21 programs. The progress of the implementation of CIF21 was monitored and reviewed quarterly as part of a performance goal in FY 2014 and FY 2015. For more information about monitoring key program investments, see the FY 2014 Annual Performance Report in the Performance chapter.

The CIF21 Council will also consider a matrix of assessment methods and measures, incorporating input and guidance from the NSF ACCI. Planned evaluation activities will address each major CIF21 goal.