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

$100,070,000
-$32,350,000 / -24.43%

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, and mobile and embedded systems, as well as by publications, experiments, simulations, surveys and 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 being pursued through other NSF-wide investments such as Understanding the Brain (UtB) and 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 constitutes 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.

Although the CIF21 NSF-wide investment is sunsetting at the end of FY 2017, NSF will continue to support developing a pervasive cyberinfrastructure that enables research at unprecedented scales, complexity, resolution, and accuracy during the sunsetting period. In particular, some CIF21 investments will transition to a set ot focused activities in support of the Administration's National Strategic Computing Initiative (NSCI) as well as NSF's new Data for Scientific Discovery and Action (D4SDA) investment area.

Total Funding for CIF21

(Dollars in Millions)

FY 2015 Actual	FY 2016 Estimate	FY 2017 Request
$157.04	**$132.42**	**$100.07**

Goal

The goal of CIF21 is to enable the 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, software, 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 directs 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 objectives:

- To support foundational data and computational research and development, including use-inspired demonstrations;
- To enable world-class data, computational, and digital capabilities and services; and
- To 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 and emerging cross-cutting research domains.

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.

Programmatic:

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 reach out to new communities of researchers and educators.

Organizational:

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 NSF's directorate advisory committees, and provides advice and strategic feedback on NSF's existing efforts and plans.
- The Cyberinfrastructure Coordination and Leadership Group (CLG) coordinates and manages a set of activities within 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. Directorate CLG membership is designated by the corresponding member of the CIF21 Council (described below).
- 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, engaging with the CIF21 Council in setting NSF-wide strategic directions. This leadership includes developing coordinated CIF21 programs and solicitations, and identifying common approaches and an overall cyberinfrastructure framework.
- As part of CIF21, the other CISE divisions and NSF directorates have working groups that focus on foundational science and engineering, as well as on 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 in 2011.[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's leadership in creating and deploying a comprehensive portfolio of advanced computing infrastructure to accelerate the pace of discovery in all fields. Along with ongoing focused workshops and events, these reports and initiatives help to define and prioritize programs and activities within the CIF21 framework.

Investment Framework

CIF21 Funding by Directorate
(Dollars in Millions)

Directorate/Office	FY 2015 Actual	FY 2016 Estimate	FY 2017 Request
Biological Sciences	$3.75	$8.39	$8.39
Computer and Information Science and Engineering	88.34	84.21	50.00
Education and Human Resources	2.50	2.50	2.50
Engineering	10.00	8.00	4.00
Geosciences	10.99	11.00	12.21
Mathematical and Physical Sciences	34.89	11.50	16.15
Social, Behavioral, and Economic Sciences	6.57	6.82	6.82
Total	**$157.04**	**$132.42**	**$100.07**

Totals may not add due to rounding.

FY 2015 – FY 2016
Following the increased emphasis on domain-specific data-intensive applications, the Critical Techniques and Technologies for Advancing Foundations and Applications of Big Data Science & Engineering (BIGDATA) solicitation was expanded in FY 2015 to address issues of scope and scale. Reproducibility of results is an issue across the sciences, and in FY 2015 NSF held a *Director's Symposium on Robust and Reliable Science: The Path Forward*. In spring 2015, NSF announced the Big Data Regional Innovation Hubs (BD Hubs) program, which aims to augment ongoing activities and ignite new Big Data public-private partnerships through the establishment of four regional BD Hubs. In FY 2015, ENG and the Air Force Office of Scientific Research issued a DCL to support innovative research in the area of Dynamic Data Systems (DDS). This DCL resulted in EArly-concept Grants for Exploratory Research (EAGER) with the potential to transform 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.

[1] NSF Advisory Committee for Cyberinfrastructure www.nsf.gov/cise/aci/taskforces/index.jsp
[2] A Vision and Strategy for Data in Science, Engineering and Education: www.nsf.gov/cise/aci/cif21/DataVision2012.pdf
[3] Software for Science and Engineering; www.nsf.gov/cise/aci/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

In FY 2016, foundational research efforts via the BIGDATA solicitation are being broadened to address not only scalability and exploration of new data science capabilities, but also new approaches at the interface between data science and long-term data management and data interoperability for data life cycles that serve multiple stakeholders. In addition, the solicitation will support efforts to employ novel data science methodologies to significantly advance the field's knowledge base on STEM learning and learning environments, broadening participation in STEM, and increasing retention for students traditionally underserved in STEM at the K-12, undergraduate, and graduate levels. The four regional BD Hubs will come online in FY 2016. Each BD Hub will establish a governance structure for its consortium of public and private members, and the BD Hubs collectively will begin developing approaches to ensure cross-hub collaboration and sustainability over the long term. In addition, NSF will begin supporting the next phase of the BD Hubs program by funding a set of BD Spokes that will build out various sectors of particular interest to each BD Hub (e.g., transportation, energy, public safety, and education) to advance Big Data innovation throughout the Nation.

In FY 2015, the Data Infrastructure Building Blocks (DIBBs) program continued to focus on collaboration among all directorates and the formulation of discipline-specific data science and data management pilots. DIBBs increased its scale and scope, emphasizing the value of sharing data beyond a specific institution to the wider science, engineering, and education communities through the Campus Cyberinfrastructure – Data, Networking and Innovation (CC*DNI) program solicitation. As a result of this solicitation, NSF made investments in multi-campus and/or multi-institutional regional cyberinfrastructure 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 2016, the DIBBs program is building on existing community development activities in and across NSF's directorates and offices as well as the previous three years of experiences in the program. NSF directorates are expanding the scale and scope of directorate and multi-directorate Data Science Pilots, building on pilots initiated in previous years and in coordination with the evolving directorate Data Management Plans that are being guided by the *NSF Public Access Plan* (NSF 15-52). Based on individual directorates' investment priorities, expanded scope includes expansion of data analytic algorithms and tools aligned with research priorities, as well as evolution of data management, including the associated social and technical infrastructure. It is anticipated that the program will support activities related to data reproducibility; interoperability of research data as well as specific non-research data; data sustainability plans particularly for high-value community datasets; institutional data policies and community governance; security, privacy, integrity and trustworthiness of research data; exploration of innovative economic/operating models for archiving and curation of research data; and learning and workforce development. The program will pursue increased scale toward national-level, international, and multi-agency activities.

The CDS&E program, led by MPS, ENG, and CISE, includes efforts and approaches for simulation and modeling, along with a specific focus on scaling. The BIO and GEO directorates also participate informally in the proposal and review process. In FY 2015, additional prototype and proof-of-concept approaches for CDS&E were developed. Based on the results of continuing portfolio analysis, the program was structured to address emerging issues of scope and scale. In FY 2016, CDS&E efforts are addressing issues associated with expanding both the base of researchers as well as the participation of new domains and disciplines. In addition, the availability of new tools and technologies resulting from research and infrastructure advances in computation and data are providing new opportunities for communities that have had limited access and use of research data and advanced computing infrastructure.

The Advances in Biological Infrastructure (ABI) program represents NSF's support for cyberinfrastructure for the biological sciences community. While this investment has not previously been highlighted specifically, investments in this area have always been a component of CIF21. In FY 2015, the ABI

program supported mid-scale cyberinfrastructure and community coordination activities, such as Research Coordination Networks (RCNs), workshops, and other engagement mechanisms. In FY 2016, ABI is investing in cyberinfrastructure resources that have the potential to advance or transform research in biology supported by BIO. A primary focus is crosscutting, high-priority areas, such as UtB and Genotype to Phenotype.

In FY 2015, EarthCube continued to support community coordination activities, such as RCNs, workshops, and other engagement mechanisms. These activities broadened the base of users as well as the breadth of science conducted via EarthCube. The Integrative Activities track of the EarthCube solicitation supported integration of resources and expansion of cyberinfrastructure to serve other geoscience domains. Early efforts included the development of common approaches and standards including some level of integration and coordination across projects. An EarthCube All-Hands meeting provided a venue for demonstrations of integrative technologies and plans for integrating or employing those technologies in the subsequent few years.

In FY 2016, EarthCube is continuing to support community coordination activities and the development of common approaches. Based on the results from the governance and cyberinfrastructure communities, EarthCube is developing programs that begin to bridge and support multiple geoscience communities. This includes integrating existing pilots and prototypes to address issues of scale and research support. In addition, EarthCube is focusing on at-scale issues including development and deployment of common approaches and structures. This includes development of common tools, data systems, and virtual organizations to support the emerging geoscience communities and to coordinate regional and national cyberinfrastructure facilities.

In FY 2015, the Software Infrastructure for Sustained Innovation (SI^2) program began working more closely with the DIBBs program to encourage proposals that have both software and data elements. SI^2 also developed a supplement mechanism through which NSF awardees could apply for additional funds to support creating and supporting open source software developed through general (non-SI^2) projects. These supplements were co-funded by the SI^2 program and the funders of the original award. This mechanism replaced much of the existing software reuse activity funded as part of CIF21. Additionally, the SI^2 program held a PI meeting for grantees in FY 2015. In FY 2016, the SI^2 program is expanding the scope of its award portfolio by introducing the first Software Infrastructure for Sustained Innovation (S2I2) awards. As it continues to develop software sustainability models for scientific software, it is also beginning to focus on common approaches and issues across multiple institutions and software projects, including integration and coordination of development and deployment as well as validation and verification. This includes a focus on software architecture and infrastructure for major NSF scientific and engineering priorities and investments including possible international joint funding activities.

The NSF Research Traineeship (NRT) program is designed to encourage the development of new, potentially transformative, and scalable models for STEM graduate training that ensure graduate students develop the skills, knowledge, and competencies needed to pursue a range of STEM careers. In FY 2015, NRT supported the priority theme Data-Enabled Science and Engineering (DESE). DESE is aligned directly with CIF21, and includes an emphasis on computational- and data-enabled science and engineering. NRT will continue to support the CIF21/DESE theme in FY 2016.

In FY 2015, a revised and updated version of the Building Community and Capacity in Data Intensive Research in Education (BCC-EHR) solicitation was issued, along with a new solicitation, Resource Implementations for Data Intensive Research in the Social, Behavioral, and Economic Sciences (RIDIR). RIDIR is designed to enable user-friendly, large-scale, next-generation data resources and relevant analytic techniques to advance fundamental research in SBE areas of study. RIDIR awards are constructing databases and/or relevant analytic techniques, and will ultimately produce finished products to enable new

types of data-intensive research in multiple disciplines or fields. The databases/techniques will have significant impacts across multiple fields by enabling new types of data-intensive research that includes, but is not necessarily limited to, the social, behavioral, and economic sciences. In FY 2016, NSF will issue a revised and updated version of the BCC-EHR solicitation and will continue to support development of user-friendly, large-scale, next-generation data resources and relevant analytical techniques to advance fundamental social, behavioral, and economic sciences research through increased investment in the RIDIR program.

FY 2017 Request

- Big Data-related activities supported under CIF21 will begin transitioning to new activities within D4SDA in FY 2017. NSF will continue to build out various sectors of particular interest to each BD Hub to advance sector innovation in each region.
- NSF support for data infrastructure for research and education through the DIBBs program will begin transitioning to the D4SDA investment area in FY 2017. This support will build upon foundational data and computational science advances guided by research and education priorities within and across directorates. This support will also be informed by NSF's evolving policies for data management, data citation, and public access. 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.
- CDS&E will continue to 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. Support for CDS&E will continue beyond the end of the CIF21 activity.
- ABI will continue to support midscale investments that advance data, software, and collaborative infrastructure in support of several priority research investments including Understanding the Brain, Genotype to Phenotype, and Plant Genome.
- 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 geosciences community will gain a stable venue to coordinate future infrastructure advances as science drivers and technologies change. Additionally, EarthCube will enable new transformative geosciences research and education, through improved ability to access and analyze geosciences data, using effective software, models, and analytical tools that can simulate and examine complex and interrelated Earth processes. Support for EarthCube will continue beyond the end of the CIF21 activity.
- 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 shared 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 sizes of the user communities and their science and engineering research productivity. These metrics support the goal of creating long-term, sustainable impact. Funding for SI^2 will begin to transition to new activities under NSCI in FY 2017, focusing on scientific software architecture and infrastructure for new approaches to computational and data-enabled discovery.
- Prior investments in the NRT CIF21/DESE priority theme will continue to support graduate students in computational- and data-enabled science and engineering, and lead to the development and testing of potentially transformative and scalable models for graduate education more broadly.
- In FY 2017, EHR will fund research addressing the following issues:

  - Employing data science methodologies to significantly advance the field's knowledge base on STEM learning and learning environments, broadening participation in STEM, and increasing retention for students traditionally underserved in STEM at the K-12, undergraduate and/or graduate levels; and
  - Advancing, adapting and developing novel data science methods and techniques that are distinctly suited to answering educational research questions.
- 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 (such as 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.

FY 2018

While CIF21 ends in FY 2017, the deep integration of cyberinfrastructure with research will continue to transform the discovery process across all disciplines. NSF will develop a subsequent, focused set of activities as part of NSCI. In response to multiple technological transitions, the above-mentioned data themes, and the intimate role that computation plays in research leadership, NSF will continue to focus efforts on advancing the Nation's computational infrastructure for research. In addition, through the new D4SDA investments, NSF will support research and research infrastructure for advancing data science, data management, and data policy.

Evaluation Framework

NSF has deployed a variety of 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 an agency performance goal in FY 2014 and FY 2015. For more information about monitoring key program investments, see the FY 2015 Annual Performance Report in the Performance chapter.