

OneNSF INVESTMENTS CYBERINFRASTRUCTURE FRAMEWORK FOR 21ST CENTURY SCIENCE AND ENGINEERING (CIF21)

OVERVIEW

NSF has a long history of investments in cyberinfrastructure. Since the 1960s, NSF has sought to develop the foundational cyberinfrastructure technologies needed to maintain U.S. prominence in science and engineering (S&E). The shaping of CIF21, a new approach to managing and investing in cyberinfrastructure development and deployment, began in 2007 when NSF formed the Advisory Committee for Cyberinfrastructure (ACCI). Through its six task forces, the ACCI developed recommendations for future directions in the areas of bridging campuses, data and visualization, grand challenges, high-performance computing, learning and workforce development, and software. The recommendations, which are contained in task force final reports formally accepted in 2011, serve as a basis for CIF21 activities.

Goals

CIF21 will promote new research opportunities and partnerships across all S&E fields by providing integrated and scalable cyberinfrastructure that will leverage existing and new components. This includes identifying common methods and approaches in software, data, networking, cybersecurity, and advanced computing infrastructure, and leveraging the substantial existing cyberinfrastructure activities and efforts underway across NSF. CIF21 will ensure long-term sustainability for cyberinfrastructure by creating communities of users and providers and will also promote the development of a scientific workforce that has appropriate skills and training to develop, deploy, and use cyberinfrastructure. This will result in a greater number of scientific communities that can effectively use cyberinfrastructure to conduct research and education.

A new coordinated NSF-wide strategy for Advanced Computing Infrastructure (ACI) was developed in early FY 2012 as a key component of the CIF21 framework. This strategy seeks to position and support the entire spectrum of NSF-funded communities at the cutting edge of advanced computing technologies, hardware, and software. It also aims to move away from an approach centered on procurement and promote a more complementary, comprehensive, and balanced portfolio of advanced computing infrastructure and programs for research and education to support multidisciplinary computational and data-enabled science and engineering that supports the entire scientific, engineering, and education community. This approach is consistent with the recent PCAST review of NITRD (*PCAST, December 2010, "Designing a Digital Future: Federally Funded Research and Development in Networking and Information Technology."*)

Approach

CIF21 is governed by a steering committee, consisting of NSF assistant directors, which provides strategic direction for CIF21 activities. In FY 2012 this committee established a CIF21 Strategy and Leadership Group (SLG), with representation from each directorate and office; OCI provides both technical and administrative support. The SLG provides guidance for a portfolio of programs, including cross-directorate (common) and single directorate (unique) programs. It has four major areas of responsibility: 1) Development of solicitation guidance for common CIF21 programs; 2) coordination of common and unique CIF21 activities; 3) convening and charging of writing groups for common CIF21 programs; and 4) programmatic planning and budgeting for CIF21. The SLG has chartered initial writing groups in software development, data services, and computational and data-enabled science and

engineering (CDS&E). These groups will coordinate existing and new NSF programs, perform gap analyses, and develop new programs to fill the gaps.

INVESTMENT FRAMEWORK

FY 2012

FY 2012 is the first year of CIF21. Investments in FY 2012 begin to build the CIF21 portfolio by focusing on building new communities of developers and users of digital environments and beginning to make awards that develop core data and computational technologies to bridge scientific communities. This will position NSF to specify requirements for data systems and computational requirements and to further develop and integrate robust software resources in FY 2013.

Community Building

- The EarthCube program, involving GEO and OCI, is a multi-year initiative to support earth-system science. Early-Concept Grants for Exploratory Research (EAGER) and supplement awards are being used to develop concepts. A community-building event will be held mid-year and additional awards will be made in summer/fall 2012. OCI is providing \$2.0 million and GEO is providing \$4.0 million in support for FY 2012.
- The existing Research Coordination Networks (RCN) solicitation, supported by BIO, EHR, GEO, MPS, SBE, OPP, OISE, and OCI, has been expanded to involve two additional directorates, CISE and ENG.

Integration

- The Software Infrastructure for Sustained Innovation (SI²) program involves BIO, ENG, MPS, and OCI. Its goal is to support the further development of existing scientific software to create tools that can be useful to broader communities. A new component of SI² is initiation of an Institutes track to establish center-scale awards of approximately \$5.0 million per year, each, that bring together tools, data, and computational resources to advance grand challenge science goals. A novel, multi-stage process is being developed that will be science driven, risk-managed, and promote cross-disciplinary collaboration.
- The SLG is forging partnerships with other cyberinfrastructure efforts at the Foundation, including those involving construction of large facilities and the cyberinfrastructure track of the Cyber-Enabled Materials, Manufacturing, and Smart Systems (CEMSS) OneNSF investment.

Data and Computational Technologies

- The BigData Senior Steering Group (BDSSG) of the Networking and Information Technology Research and Development (NITRD) program, will launch a Core Techniques and Technologies for Advancing Science and Engineering (BIGDATA) solicitation involving CISE, MPS, and OCI at NSF, as well as other agencies, including several institutes at NIH. The program aims to develop new approaches and tools to address the challenges of managing, analyzing, visualizing, and extracting useful knowledge from large, diverse, distributed, and heterogeneous data sets. This includes the development of data analytics, algorithms, and statistical and mathematical methods.
- A new program in computational and data-enabled science and engineering in Mathematical and Statistical Sciences (CDS&E-MSS) is established jointly with OCI and the Division of Mathematical Sciences (DMS). Workforce development and community needs workshops are planned in order to determine unmet needs in the mathematics and statistical science communities with respect to computational resources.
- Multiple efforts have been initiated in the area of data cyberinfrastructure. A joint effort between MPS and OCI addresses long term data storage, management, and access issues. SBE and OCI have

a joint program in the area of metadata for long-standing large scale social science surveys (META-SSS) that seeks to enhance data access and usability, and to facilitate interdisciplinary analyses. OCI, MPS, GEO, and the European Union (EU) are coordinating on a call for international data cyberinfrastructure proposals. The new OCI DataNet solicitation addresses the data cyberinfrastructure needs of multidisciplinary data-intensive researchers.

FY 2013 Request

Community Building

- The EarthCube program will develop additional community efforts and prototype systems for integrating scientific communities with respect to data, modeling, and analysis. New areas of common interest and joint solicitations will be developed with BIO, CISE, ENG, MPS, and OPP to create components of EarthCube. OCI will provide \$5.0 million and GEO will provide \$12.0 million to support EarthCube in FY 2013.
- The RCN will continue to build digital communities with new scientific communities, including OPP. Some of the evolving prototypes may transition to operational RCNs. International collaboration on data curation, standards, and access will be explored.

Integration

- Under the SI² program, it is expected that the first software institute will be established. This center will serve as a resource for software development to scientific communities. GEO and some of the large facility construction projects are expected to participate. Likely collaborators include the large telescopes, the High Energy Physics Center, the National Ecological Observatory Network, and the Ocean Observatories Initiative. A joint effort on exascale software development with the Department of Energy is being explored.
- NSF's large facilities are in many cases developers and users of significant cyberinfrastructure resources. A well attended breakout session at the spring 2011 Annual Large Facilities Operations workshop highlighted the concerns related to data and cybersecurity. A CIF21 WG consisting of program directors of large facilities is being considered. This will focus on common cyberinfrastructure needs of the major facilities.
- Efforts will continue to support cyberinfrastructure development for materials research in collaboration with the CEMMSS investment.

Data and Computational Technologies

- The Core Techniques and Technologies for Advancing Big Data Science & Engineering (BIGDATA) program will be expanded to involve BIO, ENG, and SBE at NSF, as well as additional NITRD agencies. The initiative will expand to include education and workforce development activities, as well as competitions and prizes in areas relevant to big data challenges.
- OCI data cyberinfrastructure programs will include additional solicitations on data interoperability, storage, policy, and integration with increased international coordination.
- Large facility data requirements will be considered in the data storage, management, and access efforts. Joint solicitations to include BIO, ENG, and GEO, in addition to MPS and OCI, are planned.

Sustainability

- Efforts in programs across the portfolio of CIF21 programs will ensure development of the workforce required to sustain the Nation's cyberinfrastructure.

FY 2014 – FY 2016

In FY 2014 through FY 2016, it is expected that many of the development efforts will begin to produce practical tools for deployment and use by research communities. Relying on a spiral development approach, promising areas for additional investment will be considered by the Strategy and Leadership Group and AD Steering Committee for further development. Also, programs will begin to experiment with different funding approaches, including grand challenge prizes.

EarthCube is expected to develop a suite of tools to integrate the geoscience communities, and many are expected to become available and adapted for use by other scientific communities outside the geosciences.

The SI² program will mature by supporting a total of 2-3 institutes that will serve as a resource to software developers in academic communities, as well as a number of smaller software projects. These combined efforts will expand the network and library of accessible and useable software tools.

The Core Techniques and Technologies for Advancing Big Data Science & Engineering (BIGDATA) solicitation will include education and workforce development activities, as well as competitions and prizes in areas relevant to big data challenges.

OCI data cyberinfrastructure programs will provide increasing levels of access to federally funded scientific data for national and international scientific and engineering collaborations.

Research to understand and solve issues of cyberinfrastructure sustainability, including shared governance, succession, and management of technological trajectories will be supported. The transition to practice focus and efforts in OCI help bridge the gap from idea to deployment. Awards will ensure that tools are created for scientists and that scientists develop skills and relationships with industry to create a full cycle of innovation for cyberinfrastructure.

CIF21 Funding

(Dollars in Millions)

Directorate/Office	FY 2012 Estimate	FY 2013 Request
BIO	\$2.00	\$4.00
CISE	12.00	16.00
ENG	5.00	11.00
GEO	4.00	12.00
MPS	11.50	19.55
SBE	5.50	6.00
OCI	23.00	32.03
OISE	-	1.00
OPP	4.00	4.50
IA	11.00	-
EHR	-	-
Total, NSF	\$78.00	\$106.08

Totals may not add due to rounding.

EVALUATION FRAMEWORK

The CIF21 investment will be evaluated by the Strategy and Leadership Group (SLG) in the short term and by an independent contractor in the long term.

In the short term (FY 2012 – FY 2013) the SLG, in concert with its writing groups, will be performing assessments to inform decision making at the close of each fiscal year. The writing groups will be evaluating directorate and office efforts and solicitations that support ongoing discipline-driven cyberinfrastructure to assess how these efforts can be coordinated with existing and planned CIF21 programs. These assessments will inform decisions concerning the overall CIF21 portfolio balance as well as provide a basis for suggested changes to other directorate and office programs involving cyberinfrastructure to focus on common approaches and leverage CIF21-specific programs.

Prior to the end of FY 2013, the investment-wide assessment will be shared with both the AD steering committee as well as the external advisory committee, ACCI, in order to gain insight in strategic directions. Adjustments to existing programs and formulation of plans will result. A proposal for an independent evaluation by an outside party may be initiated before the end of FY 2013 to help further define data needs with respect to overall portfolio evaluation.

In FY 2014 or FY 2015 a summative evaluation will be carried out through an outside party. The evaluation will examine outcomes of awards to establish the degree to which the goals of scalability, integration, and sustainability have been achieved, as well as the degree to which the cyberinfrastructure has supported new scientific discoveries. Specific metrics might include the value of CI tools associated with new discoveries, the level of use of new coupled systems, the degree of coupling of existing systems, disciplinary diversity of system users, and perceptions of the level of CI support in various communities. The impact of CIF21 in computing and cyberinfrastructure programs will extend beyond the programs directly affiliated with CIF21, so it will not be appropriate to compare programs under the CIF21 framework with those outside it.

