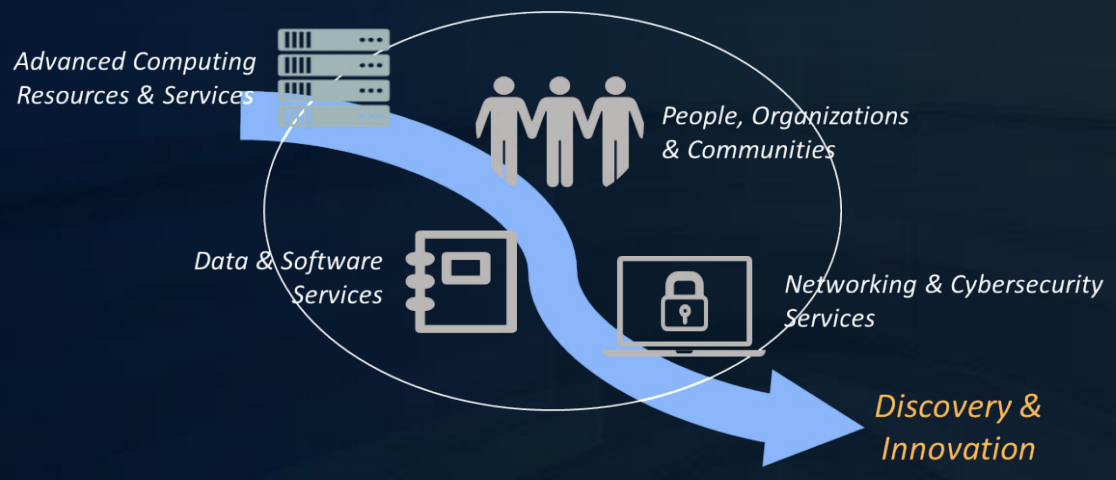
Transforming Science Through Cyberinfrastructure

NSF's Blueprint for a National Cyberinfrastructure Ecosystem for Science and Engineering in the 21st Century



Vision for a CI ecosystem, and Blueprint for Computational Resources and Services

Office of Advanced Cyberinfrastructure
Directorate for Computer & Information Science & Engineering
National Science Foundation

April 2019

Transforming Science Through Cyberinfrastructure

NSF's Blueprint for a National Cyberinfrastructure Ecosystem for Science and Engineering in the 21st Century

Executive Summary

Twenty-first century science and engineering (S&E) research is being transformed by the increasing availability and scales of computation and data. The national cyberinfrastructure (CI) ecosystem has thus become a key catalyst for discovery and innovation and now plays a critical role in ensuring US leadership in S&E, economic competitiveness and national security, consistent with NSF's mission. The vision and blueprint presented in this document have been developed by the NSF Office of Advanced Cyberinfrastructure (OAC) on behalf of NSF based on a synthesis of multiple community inputs through advisory bodies, requests for information (RFIs), workshops and conferences, and national initiatives.

A new vision. NSF envisions ***an agile, integrated, robust, trustworthy and sustainable CI ecosystem that drives new thinking and transformative discoveries in all areas of S&E research and education.*** This vision embodies the following overarching principles:

- View cyberinfrastructure more holistically as a spectrum of computational, data, software, networking, and security resources, tool and services, and computational and data skills and expertise that can be seamlessly integrated and used, and collectively enable new, transformative discoveries across S&E.
- Recognize and support the translational research continuum, from *(i)* catalyzing core cyberinfrastructure innovations essential to address disruptive changes in applications and technologies, through *(ii)* fostering the development of community tools and frameworks, to *(iii)* enabling the deployment and operation of sustainable production-quality cyberinfrastructure services.
- Develop a strategy that balances innovations with stability and continuity. For example, in the context of the computation ecosystem, develop a longer-term strategy for providing continuity in production computational capacity while ensuring that there are opportunities to explore innovations and to transition these innovations to production when appropriate.
- Work closely with the diverse S&E communities to tightly couple the cycles of discovery and innovation essential to addressing new challenges and opportunities in an era of disruptive technologies and changing science needs.
- Achieve new levels of usability by easing the pathways for discovering, accessing, understanding and utilizing powerful CI capabilities and services, and enhancing scientists' productivity and science impact.

From vision to action. New funding approaches and opportunities in the near- and mid-term by the NSF Office of Advanced Cyberinfrastructure (OAC) will target:

- Addressing needs holistically across multiple research disciplines, domains, and NSF priority areas such as the NSF Big Ideas and the NSF major facilities, and in so doing,

encouraging an expansion of the range of research communities that benefit from CI resources and services.

- Fostering CI innovation through continued investment in foundational and translational CI research and education activities that lead to deployable, scalable, and sustainable cyberinfrastructure services capable to transforming science.
- Achieving robustness, accessibility and responsiveness of the CI ecosystem through OAC investments that build integratively on previous and ongoing developments within individual CI areas and within individual CI-enabled research disciplines, and through maximal leveraging of new technologies as they emerge.
- Continuing investments in education, training and developing a broad and diverse CI community, promoting coordination and exchange between the CI and research communities; and facilitating dissemination of best practices for design, development, and operation of CI resources and capabilities.

A living document. This document consists of two parts. The first part (Sections 2 & 3) presents an evolving vision for a CI ecosystem, describing the drivers and summarizing the current NSF CI landscape. The second part (Sections 4 & 5) sets forth a blueprint for investment in the computational resources and services, and outlines an implementation plan, consisting of extensions and enhancements to current investments and new programs and opportunities in 2019 and beyond based on two strategies:

- Deploy a balanced computational ecosystem that can effectively support a broad and diverse set of requirements, users and usage modes, including Leadership Class Systems, Capacity Systems, Federated Resources, Prototypes and Testbeds, in concert with continued investments in Campus CI investments and new emphasis on inclusion of emerging cloud resources and services.
- Achieve maximal impact from the array of computational capabilities and expertise, individually and collectively, through strategic investments in crosscutting coordination, resource allocation, user services and support, and performance measurement capabilities, as well as in CI workforce development.

NSF expects that this blueprint is the first of several similar blueprints focused on different elements of the CI ecosystem. NSF will continue to work with the community to evolve and implement this blueprint for computational resources/services, as well as to develop complementary blueprints for other CI elements, with the overarching goal of realizing an integrated CI ecosystem that transforms science.

Table of Contents

Executive Summary 1
1 Introduction 4
1.1 NSF, OAC, and The National Cyberinfrastructure Ecosystem 4
1.2 The Goal and Focus of this Document 5
2 A Time of Disruptive Changes in Applications and Technology 6
2.1 Disruptive Application Pulls ..... **Error! Bookmark not defined.**
2.2 Disruptive Technology Pushes 7
2.3 Responding to Disruptive Changes 8
3 A Vision for NSF's Cyberinfrastructure Ecosystem 8
3.1 Building on Community Inputs 9
3.2 NSF's Current CI Landscape 11
3.2.1 Advanced CI Resources and Services 11
3.2.2 Software/Data CI and Services 16
3.2.3 Networking 17
3.2.4 Cybersecurity 17
3.2.5 Learning and Workforce Development (LWD) 18
3.3 Core CI Research 19
3.4 NSF Large Facilities and CI 19
3.5 NSF's Big Ideas and CI 21
3.6 Towards an Integrated CI Ecosystem 21
4 A Blueprint for a National Computational Ecosystem for Science and Engineering in the 21st Century 23
4.1 Elements of a balanced computational ecosystem 23
4.2 Putting the Plan into Action: Programs, Projects and Opportunities 25
4.2.1 Extending and Enhancing Current Investments 25
4.2.2 The Road Ahead – Programs, Projects and Opportunities in 2019 and Beyond 26
5 Ongoing Strategic Planning and Community Engagement 30
6 Conclusion 30

1 Introduction

1.1 NSF, OAC, and The National Cyberinfrastructure Ecosystem

Science and engineering (S&E) are being transformed in the 21st century by the increasing availability and scales of computation and data. As a result, the national cyberinfrastructure (CI) ecosystem is playing a central role in all areas of S&E research and education and has become a key catalyst for discoveries and innovation. It is also critical to ensuring US leadership in S&E, economic competitiveness and national security.

The overarching mission of the National Science Foundation (NSF) is *"to promote the progress of science; to advance the national health, prosperity, and welfare; and to secure the national defense; and for other purposes."* NSF envisions a nation that capitalizes on new concepts in S&E and provides global leadership in advancing research and education. In support of this mission, the NSF has effectively supported the broad availability and innovative use of CI, primarily through the investments of its Office of Advanced Cyberinfrastructure (OAC) and its antecedents (e.g., the Division of Advanced Cyber Infrastructure (ACI) and the Office of Cyberinfrastructure (OCI) before that).

OAC is responsible for the conceptualization, design and implementation of the advanced cyberinfrastructure (CI) ecosystem to support NSF's mission to advance all areas of science and engineering (S&E) research and education. These investments have spanned discipline-specific instruments and facilities; computational systems of varying capabilities and architectures optimized for different applications; virtual organizations for allocating resources and interfacing with users; the network backbone that connects and provides access to these resources; cybersecurity mechanism, tool and expertise; software and data services; as well as cross-cutting research and education programs. Over the past decade OAC (and its predecessors: ACI and OCI) has developed a complementary, comprehensive, and balanced portfolio of CI investments and has funded and coordinated exploration, development and provisioning of advanced CI resources, facilities, and services. OAC's investments blend translational computer and computational research and research cyberinfrastructure and integrate innovations from the private sector.

Overarching Characteristics of OAC Investments

- ***Science-driven***: Promotes science excellence, enabling fundamentally new advances in all areas of science and engineering.
- ***Strategic***: Intentional and aligned with science and CI priorities.
- ***Innovative***: Emphasizes innovation across all aspects of CI, including conceptualization, design, deployment, operation and use; considers both human and technical aspects of CI.
- ***Integrative/Interoperable:*** Ensures that individual CI elements can be seamlessly composed and can collectively drive science.
- ***Collaborative***: Fosters partnerships and community development; actively engages CI experts, specialists and scientists working in concert with domain scientists/CI users.
- ***Leveraged***: Builds on foundational and translational innovations; leverages existing, recognized capabilities.
- ***Measurable/Data-driven***: Encourages measurement of progress and sharing of results through management plans and metrics.
- ***Sustainable***: Provides benefits beyond the participants and the lifetime of the award.

NSF's CI investments are a key component of the national CI ecosystem that uniquely and strategically address the broad and diverse needs of its stakeholders across the entire scientific, engineering, and education community. These investments complement investments by other U.S. agencies and institutions including Department of Energy (DOE), Department of Defense (DOD), Department of Homeland Security (DHS), National Aeronautics and Space Administration (NASA), National Institutes of Health (NIH), National Institute of Standards and Technology (NIST), and National Oceanic and Atmospheric Administration (NOAA), and are in partnership with academia and the private sector. For example, DOE is focused on achieving exascale computing performance, NASA has focused on modular computing technology for energy efficiency, and NIH is exploring a shared interoperable cloud computing environment for high-throughput biomedical research[1]. Collectively these investments represent strategic national assets that are critical to US leadership in S&E, economic competitiveness and national security.

Additionally, recent international activities to accelerate investments in very large-scale computational resources , particularly in Europe and Asia, are providing urgency and importance for an investment strategy to maintain our nation's global leadership role in S&E. NSF continues to cooperate formally and informally with other federal agencies to ensure that the U.S. remains a world leader in the research, deployment, and use of HPC for all areas of S&E research conducted in our Nation.

1.2 The Goal and Focus of this Document

This document has two goals.

- The first part (Sections 2 and 3) presents NSF's vision for a national CI ecosystem that integrates computational, data, software, networking, and security resources, tool and services, and computational and data skills and expertise towards collectively enabling new, transformative discoveries across S&E. It also presents a snapshot (as of first quarter, calendar 2019) of the NSF-funded CI ecosystem, supported through OAC[2], highlighting its current and ongoing investments.
- The second part (Sections 4 and 5) focuses on the *computational* aspects of the CI ecosystem and presents NSF's blueprint for a computational ecosystem to support science and engineering research and education in the 21st century. It also outlines a plan to implement this blueprint, highlighting extensions and enhancements to current investments as well as new programs and opportunities in 2019 and beyond.

The growing role of CI across S&E research and education and the increasing demands for CI resources, services and expertise warrants significant increases in CI investments nationally. However, this document is driven by current budget realities and assumes as a baseline, a continuation of the overall level of support (relative to NSF appropriations) that OAC (and ACI and OCI before it) has received over the past decade.

[1] The Networking & Information Technology Research and Development (NITRD), Program, https://www.nitrd.gov/.

[2] The focus on OAC-funded activities excludes complementary projects supported by other NSF research directorates (for example, the CI resources and services at NCAR).

The vision and blueprint described in this document is informed by the community through advisory bodies, requests for information (RFIs), workshops and conferences, and national initiatives. Specifically, this document is influenced by and is, in part, a response to recent reports from a 2016 National Academy of Sciences, Engineering, and Medicine (NASEM)[3] study and 2018 community workshop[4] (discussed in more detail in Section 3.1).

This blueprint is the first of several such blueprints to be focused on different elements of the CI ecosystem, such as software and data services, networking, cybersecurity, and learning and workforce development, etc., that NSF expects to develop through continued engagements with the community.

2 A Time of Disruptive Changes in Applications and Technology

NSF's CI investments aim to support the full range of computational- and data-intensive research across all of S&E. Recent years have witnessed dramatic changes in the application and technology landscapes leading to disruptive changes in both the scale and nature of CI requirements with profound implications for OAC's strategic priorities and investments.

2.1 Disruptive Application Pulls

Recent years have witnessed dramatic changes in the number and nature of applications using NSF-funded resources, as well as the growth of new classes of applications that, for example, leverage data analytics and machine learning. These changes are evident in recent workload analyses of the NSF-funded Blue Waters[5] and Innovative HPC[6] resources. These changes are driven in part by the increasing scales and pervasiveness of compute and data and are resulting in new demands on the scale and type (e.g., extreme-scale, data-centric) and location (e.g., close to the data or close to the users) of the resources as well as in new usage modes (e.g., on-demand, elastic). Several factors are driving these changes, including:

1. *High fidelity/resolution simulations:* The desire and ability to model phenomena more holistically across multiple scales and physics are resulting in high-resolution, dynamic, coupled simulation workflows capable of running at extreme scales.
2. *Increasing availability and scales of experimental and observational data:* Increasing availability and maturity of large scale experimental and observational facilities are resulting in unprecedented volumes of data and in new and growing requirements for data management, transport, processing and storage, as well as the integration of this data into application workflows.
3. *Online data processing and actuation:* Increasing levels and resolution of instrumentation and the overall proliferation of digital data sources are leading to new opportunities for online

[3] "The National Academies Report on Future Directions for NSF Advanced Computing Infrastructure to Support U.S. Science and Engineering in 2017-2020," https://www.nap.edu/catalog/21886/future-directions-for-nsf-advanced-computing-infrastructure-to-support-us-science-and-engineering-in-2017-2020.

[4] "Workshop on Future Cyberinfrastructure: Rethinking NSF's Computational Ecosystem for 21st Century Science and Engineering," https://uiowa.edu/nsfcyberinfrastructure/.

[5] "Workload Analysis of Blue Water," 2017, https://arxiv.org/abs/1703.00924.

[6] "A Workload Analysis of NSF's Innovative HPC Resources Using XDMoD," 2018, https://arxiv.org/abs/1801.04306.

and near-real-time monitoring, data processing and actuation, and for new classes data-driven applications.

4. *Growing "long tail" applications:* Applications comprising the growing "long tail" of computational- and data-enabled S&E are increasing in complexity and scales and are quickly dominating overall computational workloads. These applications are also exploring novel ways for accessing resources. For example, a recent census of XSEDE users shows that active gateway users dominate[6].
5. *Novel data-centric/data-driven applications:* Novel data-centric application classes involving data analytics and machine learning are complementing more traditional HPC workloads and are resulting in new CI requirements. There is also an increasing use of novel and non-traditional software stacks and tools such as digital notebooks (e.g., JupyterHub), data analytics/machine-learning stacks containers, object stores, etc., which are similarly are resulting in new demands on the CI.
6. *Heightened emphasis on robust results:* A renewed emphasis on robust results across computational and data-enabled S&E is moving issues such as transparency, traceability, uncertainty, reproducibility, security, etc., as central concerns for all aspects of CI. Associated with the discussions that surround robust and reliable science is the set of policy concerns under the rubric of Open Science, which similarly seek to minimize barriers to access to research results and the supporting data and software while increasing transparency, consistent with law, national security, and economic competitiveness.[7]

2.2 Disruptive Technology Pushes

Recent years have also witnessed dramatic changes in technologies and resources resulting in significant disruptions in computational and data processing capabilities and scales. These include:

1. *Unprecedented technological advances:* Unprecedented and potentially disruptive technological advances coupled with the exploration of technologies and paradigms beyond the Moore's law era are resulting in increasing processing speeds, new classes of processors with increased parallelism and application-specific hardware accelerators, novel storage technologies and deeper storage hierarchies, faster communication fabrics both on-machine and between machines, software defined (programmable) architectures and systems, and extreme-scale systems with millions of cores.
2. *High-bandwidth/low-latency networks:* Deployment of high-bandwidth/low-latency campus, regional, national and international networks providing global connectivity and access, and enabling end-to-end data-driven workflows that were not previous possible.
3. *Growing contributions of campus CI:* Distributed (federated) and campus-based CI has grown in capacity and capability and has become a significant provider of resources and a key enabler of all S&E.
4. *Increasing edge, in-network capabilities:* Increasing capabilities (hardware and software) at the network edge as well as within the network are enabling sophisticated, real-time data

[7] See Public Access to Results of NSF-funded Research; https://www.nsf.gov/news/special_reports/public_access/; NSF 18-060, Dear Colleague Letter: Advancing Long-term Reuse of Scientific Data, https://nsf.gov/pubs/2018/nsf18060/nsf18060.jsp

processing close to the data sources and/or the users and in-transit and enabling new S&E use cases.

5. *Proliferation of commercial cloud services:* Cloud services (along with technologies such as containers) are playing an increasingly important and complementary role in supporting S&E applications and workflows. The increasing availability and affordability of these services are lowering barriers to access, supporting new models of operation and allocation (e.g., on-demand, elastic), and providing new and unique capabilities (e.g., data-analytics accelerators) to the research community. They are also providing opportunities for hybrid and multi-cloud environments.
6. *Growing concerns about energy, failures:* As devices sizes shrink and system scales grow, energy efficiency and fault tolerance are becoming first-class design concerns for all aspects of CI.

2.3 Responding to Disruptive Changes

It is essential that the NSF continue to evolve its priorities and programs in response to the disruptive changes in the S&E applications and technology landscapes outlined above, driven as always by science needs and community inputs. In the following sections we outline our vision of an integrated CI ecosystem that can address these challenges and opportunities as well as our strategy to achieving this vision.

3 A Vision for NSF's Cyberinfrastructure Ecosystem

NSF envisions ***an agile, integrated, robust, trustworthy and sustainable CI ecosystem that drives new thinking and transformative discoveries in all areas of S&E research and education.*** Such an CI ecosystem builds on foundational and translational innovations in academia and industry, is responsive to the evolving science needs, and seamlessly integrates different CI dimensions towards addressing these needs. Our strategy for achieving this vision centers on the following overarching principles:

- View cyberinfrastructure more holistically as a spectrum of computational, data, software, networking, and security resources, tool and services, and computational and data skills and expertise that can be seamlessly integrated and used, and collectively enable new, transformative discoveries across S&E.
- Recognize and support the translational research continuum, from *(i)* catalyzing core cyberinfrastructure innovations essential to address disruptive changes in applications and technologies, through *(ii)* fostering the development of community tools and frameworks, to *(iii)* enabling the deployment and operation of sustainable production-quality cyberinfrastructure services.
- Develop a strategy that is based on a separation of concerns that balances innovations with stability and continuity. For example, in the context of the computation ecosystem, develop a longer-term strategy for providing continuity in production computational capacity while ensuring that there are opportunities to explore innovations and to transition these innovations to production when appropriate.

- Work closely with the diverse S&E communities to tightly couple the cycles of discovery and innovation essential to addressing new challenges and opportunities in an era of disruptive technologies and changing science needs.
- Achieve new levels of usability by easing the pathways for discovering, accessing, understanding and utilizing powerful CI capabilities and services, enhancing scientists' productivity and science impact.

3.1 Building on Community Inputs

NSF's vision for a national CI ecosystem and the blueprint for the computational ecosystem presented in this document is informed by the community through advisory bodies, requests for information (RFIs), workshops and conferences, and national initiatives, as illustrated in Figure 1. Key among these that have informed this blueprint include:

1. **The NSF-funded National Academies' study, "*The National Academies Report on Future Directions for NSF Advanced Computing Infrastructure to Support U.S. Science and Engineering in 2017-2020*."[3]** This study examined anticipated priorities and associated tradeoffs for advanced computing and provides a framework for future decision-making about NSF's advanced computing strategy and programs. It offers recommendations aimed at achieving four broad goals: (1) position the U.S. for continued leadership in science and engineering, (2) ensure that resources meet community needs, (3) aid the scientific community in keeping up with the revolution in computing, and (4) sustain the infrastructure for advanced computing.

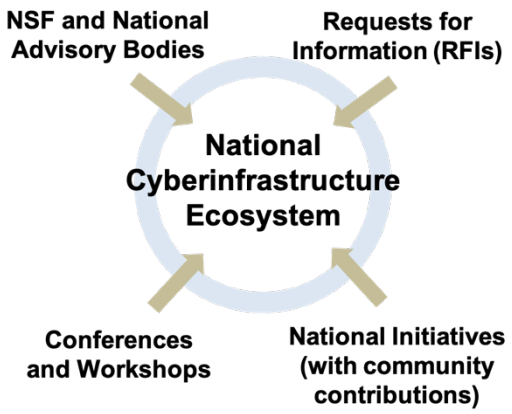


Figure 1: Gathering community input to inform NSF's CI strategic planning.

2. **Responses to the *NSF Request for Information Request for Information (RFI) on Future Needs for Advanced Cyberinfrastructure to Support Science and Engineering Research (NSF CI 2030).*[8]** This RFI, initiated on behalf of the Advisory Council on Cyberinfrastructure (ACCI), explored the advanced cyberinfrastructure the needs of the science and engineering community over the next decade. NSF received over 130 responses comprising over 300 individual contributors from across the research community – from research institutions, science-related organizations, industry, as well as many of NSF's major multi-user research facilities. The responses spanned an extraordinary range of research fields and were uniformly thoughtful and urgent in highlighting the many technical, process, and practice innovations needed to address present and anticipated trends in the scientific discovery

[8] "NSF Request for Information Request for Information (RFI) on Future Needs for Advanced Cyberinfrastructure to Support Science and Engineering Research (NSF CI 2030)," https://www.nsf.gov/cise/oac/ci2030/.

enterprise. The results have been discussed widely across the NSF research directorates and the ACCI report is available online[9].

3. **The 2018 NSF-funded *Workshop on Enabling Computer and Information Science and Engineering Research and Education in the Cloud.*[10]** The purpose of this workshop was to explore the idea of an academic cloud that provides the set of services and capabilities that serve the unique needs, workloads, and users of the CISE community. The workshop report highlighted the cloud computing needs of the CISE research and education community, with access to cloud computing resources being identified as an "equalizer" for a variety of institutions.
4. **The 2018 NSF-funded *Workshop on Future Cyberinfrastructure: Rethinking NSF's Computational Ecosystem for 21st Century Science and Engineering*.[4]** This workshop found that continued evolution of the NSF CI ecosystem must be strongly influenced by and evolve in response to dramatic, rapidly emerging increases in the number and nature of S&E research applications, increasing demands for CI capacity, as well as in the landscape of technologies, resources, and delivery mechanisms available to the S&E research community.

This vision also builds NSF's investments as part of NSF Cyberinfrastructure Framework for Twenty-First Century Science and Engineering Program (CIF21, 2012-2017)[11], a five-year initiative that saw NSF significantly expand its cross-cutting CI investments to include data and software infrastructure as well as learning and workforce development across the foundation.

Additionally, the blueprint aligns with NSF's "Big Ideas,"[12] (bold, long-term, convergent research and process ideas that are particularly well-suited for future investment at the frontiers of S&E), as well as NSF's co-leadership (in partnership with DOE and DOD) of the National Strategic Computing Initiative (NSCI).[13]

NSF's overall CI strategy and program portfolio also receives guidance and input from the NSF-wide ACCI; NSF cross-directorate Assistant Directors (AD) Council that includes ADs and Office Heads from the various NSF research directorates and offices; and the NSF-wide working group for NSCI.

[9] "CI2030: Future Advanced Cyberinfrastructure: A report of the NSF Advisory Committee for Cyberinfrastructure," https://www.nsf.gov/cise/oac/ci2030/ACCI_CI2030Report_Approved_Pub.pdf

[10] "Workshop on Enabling Computer and Information Science and Engineering Research and Education in the Cloud," https://dl.acm.org/citation.cfm?id=3233928.

[11] ''NSF Cyberinfrastructure Framework for Twenty-First Century Science and Engineering (CIF21)," https://www.nsf.gov/news/news_summ.jsp?cntn_id=122112.

[12] "NSF 10 Big Ideas," https://www.nsf.gov/news/special_reports/big_ideas/.

[13] "National Strategic Computing Initiative (NSCI)," https://www.nsf.gov/cise/nsci/.

3.2 NSF's Current CI Landscape

NSF supports the development and deployment activities along with expert services necessary for realizing the research cyberinfrastructure that is critical to the advancement of all areas of science and engineering research and education. It also supports foundational and translational research and education activities in all area of cyberinfrastructure that lead to deployable, scalable, and sustainable cyberinfrastructure services capable to transforming science. NSF's CI investments, supported through OAC programs, span advanced CI resources and services, software and data services, networking and security, as well as cross-cutting research and education programs. Key components of the NSF supported CI ecosystem are illustrated in Figure 2.

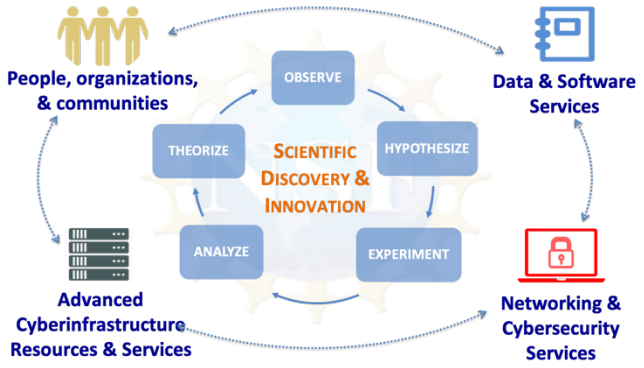


Figure 2: The NSF funded CI Ecosystem -- NSF has embraced an expansive view of CI motivated by research priorities and scientific process.

3.2.1 Advanced CI Resources and Services

NSF's current advanced CI investments span heterogeneous resources and services that, in aggregate, comprise three synergistic components. These resources and services complement each other as well as discipline-specific investments by other NSF directorates, mission-specific investments by other agencies, and cumulatively extensive but individually smaller investments by academic institutions at the regional and campus levels. This broad array of resources across all levels of investments by these different agencies and institutions supports the heterogeneous and evolving requirements of the S&E research community. NSF's three broad areas of HPC investment are:

- **Leadership-Class Computing:** The Leadership-Class Computing program aims to provide unique services and resources to advance the largest and most computationally-intensive S&E research frontiers not otherwise possible;
- **Innovative HPC/Advanced Systems and Services Investments:** The Innovative HPC program aims to provide a technically diverse advanced computing portfolio. It reflects growing and changing use of computation and data in both the research and education processes and is capable of supporting hundreds to thousands of investigators conducting cutting-edge S&E research; and
- **Support/CI Services:** These services aim to add value by coordinating the HPC resources, providing advanced assistance to the user community, supporting aggregation and federation capabilities, and broadening participation.

NSF's current advanced CI investments are summarized in Figure 3.

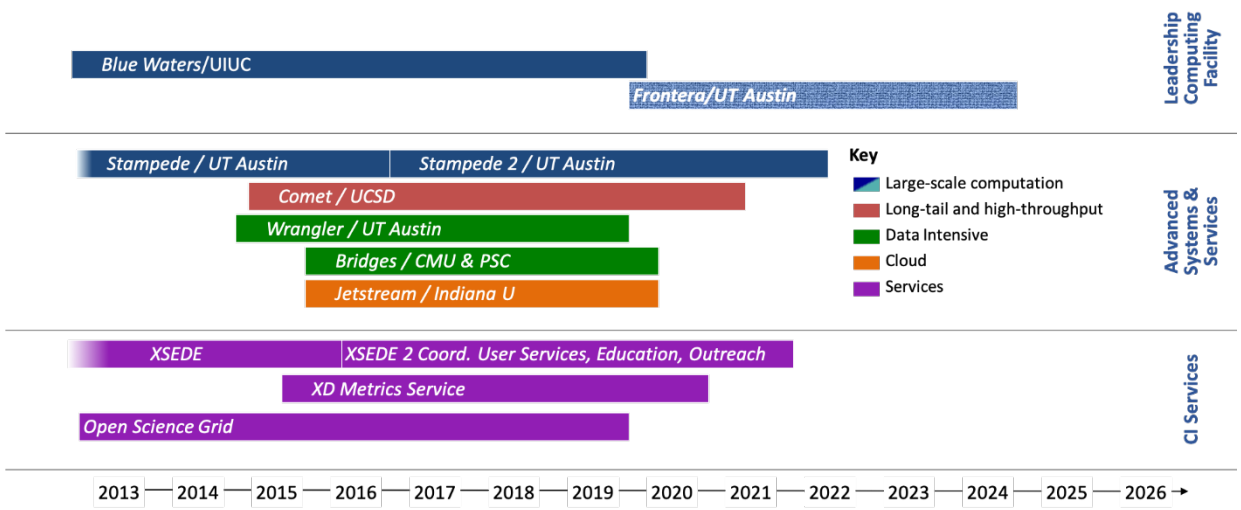


Figure 3: An overview of the NSF's current advanced CI resources and services.

3.2.1.1 Leadership Class Computing Systems

Leadership class systems have historically represented a key component of NSF's computational portfolio. NSF's current leadership-class resource is Blue Waters, which will be decommissioned at the end of 2019 and replaced by the recently-funded Frontera system that is expected to be available to early users in Spring 2019. Blue Waters was deployed at the National Center for Supercomputing Applications (NCSA) at the University of Illinois at Urbana-Champaign (UIUC) and was one of the most powerful supercomputers in the world at the time of its deployment and one of the fastest supercomputers ever deployed on a university campus. Since becoming operational in FY 2013, Blue Waters has allowed researchers to tackle much larger and more complex research challenges than ever before possible across and within disciplines as diverse as biology, astronomy, engineering, materials science, and the geosciences.

Despite the success of Blue Waters system, the computer is reaching its natural end-of-life and will complete its operational cycle in December 2019. NSF's next leadership class system will be the recently-funded Frontera system. Frontera will be deployed by the Texas Advanced Computing Center (TACC) at University of Texas at Austin (UT-Austin) and is expected to be one of largest conventional CPU systems in the world and among the most powerful supercomputers ever deployed on a US academic campus. The system is expected to begin accepting early S&E research users in April 2019 and to be fully operational by July 2019. Additional details are presented in Section 4.2.2.1.

Science and engineering research and education activities enabled: Blue Waters is enabling investigators across the country to conduct innovative research not otherwise possible due to demanding computation and data-processing capabilities[14]. Over its lifetime, the Blue Waters project has enabled more than 700 project teams, the majority of them through the highly

[14] "2017 Blue Waters Annual Report Book," https://bluewaters.ncsa.illinois.edu/liferay-content/document-library/BW_AR_2017.pdf

competitive NSF Petascale Computing Resource Allocations (PRAC)[15] program. The research topics that the PRAC program supports include: complex biological behavior in fluctuating environments; electronic properties of strongly correlated systems; properties of hydrogen and hydrogen-helium mixtures in astrophysically-relevant conditions; electronic and magnetic structures of transition metal compounds; molecular dynamics responsible for the properties of liquid water; and propagation of seismic energy through a detailed structural model of Southern California together with prediction of ground motion and modeling of the response of buildings and other structures. Other allocations address testing hypotheses about the role of cloud processes and ocean mesoscale eddy mixing; formation of the first galaxies; turbulent stellar hydrodynamics; binary black hole and neutron star systems as sources of gamma ray bursts; and other intense radiation phenomena, contagion, and particle physics. To date, there have been more than 200 education, outreach, and training projects engaging over 3,700 individuals at over 160 institutions, including 41 institutions in Established Program to Stimulate Competitive Research (EPSCoR) jurisdictions and 14 Minority-Serving Institutions.

3.2.1.2 Innovative HPC Investments

NSF's Innovative HPC program funds acquisition of nationally-available HPC resources and services that, in aggregate, are technically diverse and reflect changing and growing use of computationally- and data-intensive computation in both the research and education processes. At the same time, they are intended to enable discoveries at a computational scale beyond the reach of an individual or regional academic institution. NSF's current computational ecosystem consists of five resources that are part of this investment, i.e., *Wrangler, Comet*, *Bridges*, *Jetstream*, and *Stampede2*. These systems are briefly described in the following paragraphs.

Wrangler came online in FY 2015 at TACC, UT-Austin, and is the most powerful data analysis system in this program. This innovative system has 10 petabytes (PB) of replicated, secure, high-performance data storage; 3,000 embedded processing cores for data analysis; 120 Intel Haswell-based servers for data access and embedded analytics; and a large-scale flash storage tier for analytics, with bandwidth of one terabyte per second (TB/s) and 275 million Input/Output Operations Per Second (IOPS), which remain at the leading edge of currently deployed production systems. Support for Wrangler operations and maintenance continues through October 2019.

Comet came online in FY 2015 at the University of California, San Diego. It supports research interests and priorities requiring large, high-throughput workloads, which in turn prompt massive amounts of computation but with moderate scalability. Notably, as a resource responsive to the "long tail of science," and is particularly well-suited for the large-scale computational needs of research community portals as well as distributed workflows. Comet's heterogeneous configuration supports not only complex simulations but also advanced analytics and visualization of outputs. Due to its role in machine learning, visualization, and advanced analytics, supplemental funding was provided to increase its graphics processing unit (GPU) component in FY 2017. Comet is planned to remain operational through March 2021.

[15] "Petascale Computing Resource Allocations (PRAC)," https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=503224.

Bridges came online in FY 2016 at the Pittsburgh Supercomputing Center on the campus of Carnegie Mellon University. Bridges provides an innovative HPC and data analytics system integrating advanced memory technologies to empower new communities. It brings desktop convenience to HPC, connecting to campuses, and intuitively integrating data-intensive workflows to increase the scientific output of a large community of scientific and engineering researchers that has not traditionally used HPC resources. Bridges extends HPC's impact to EPSCoR jurisdictions and Minority-Serving Institutions, raising the level of computational awareness at four-year colleges, and promoting computational thinking in high-schools. Bridges was augmented with GPU nodes capabilities in FY 2018 and will remain operational through November 2020.

Jetstream came online in FY 2016 at Indiana University. Jetstream is a cloud-based platform that incorporates the best elements of commercial cloud computing resources with some of the best software for solving important scientific problems. Jetstream enables new modes of sharing data and computational analysis, allowing for increased scientific reproducibility and enabling scientists and engineers to make new discoveries that are important to understanding the world around us. Jetstream's system operation has been augmented to provide additional focused staff expertise to accelerate effective researcher utilizations of the programmable cyberinfrastructure/virtual machine-enabled architecture (FY 2017) and will continue operations through November 2020.

Stampede 2 came online in FY 2016 at TACC, UT-Austin as the primary national resource for thousands of academic researchers providing capabilities beyond the reach of individual campuses and regional resources, including support for multiscale modeling, simulation, and data-intensive research. Stampede2 was deployed into production operation in three phases: Knights Landing many-core nodes deployment, demonstrating increased performance at lower power rates; deployment of the highly complementary SkyLake processors, which are responsive to data-intensive computing, resulting in a peak 18-petaflop system in FY 2018; and the final deployment phase, also in FY 2018, introducing persistent memory to the previously-deployed Skylake processors to significantly enhance overall system performance. Stampede 2 will serve the high-end, open science community through November 2022.

As previously noted, Wrangler, Comet, Bridges, and Jetstream are all scheduled to ramp down operations during the FY 2020 through FY 2021 time period. During the same period, Stampede2 and Frontera, the new Phase 1 leadership-class computing system, will ramp up to full operations to ensure continued support for the science and engineering research community. Moving forward, NSF envisions that investments will continue to foster an integrated cyberinfrastructure ecosystem that addresses the growing scale and diversity of the science and engineering community, the changing nature of science and engineering research requirements, and the rapidly evolving technology and services landscape, with the overarching goal of supporting the full range of computational- and data-intensive research across all of science and engineering. This vision is presented in more detail in Section 4.

Science and engineering research and education activities enabled: NSF's Innovative HPC investments are enabling new, world-leading, and transformative advances across the breadth of science and engineering research, in the integration of research and education, and in

broadening participation in science and engineering by underrepresented groups[16]. These investments are also stimulating new collaborations across public and private sectors to advance national security and economic competitiveness. These advances are enabled by providing researchers and educators with usable access to world-leading computational resources, expertise, and services beyond those typically available on most campuses, including the interfaces, consulting support, and training necessary to facilitate their use.

3.2.1.3 Support Services

Extreme Digital (XD): Crosscutting coordination, allocation, measurement, and user support services are primarily provided by the Extreme Digital (XD) program. The XD program adds value to the Innovative HPC investments by coordinating the HPC resources and services mentioned above, providing advanced assistance to the user community, and broadening participation. The XD program's shared services model for coherently and efficiently delivering to researchers both access and expertise to diverse, dynamic, and distributed resources is a cornerstone of our HPC ecosystem. Enabling the connection between individual campuses and national resources is an essential aspect.

XD shared services consist of several interrelated parts: allocation of resources to computational and data research projects; advanced user assistance; training, education, and outreach; architecture and operation of an integrated digital services infrastructure; metrics services; and overall coordination.

Two awards are currently active within the XD program: XD Metrics Service (XMS)[17] and the eXtreme Science and Engineering Discovery Environment (XSEDE)[18]. The smaller XMS award was made in FY 2015 for five years to the University at Buffalo – The State University of New York (SUNY-Buffalo). This award provides metrics services allowing measurement of key operational data for both resources and services. All other services are provided by XSEDE. The XSEDE award to the University of Illinois at Urbana-Champaign (UIUC) was renewed in September 2016, continuing the prior XSEDE award for another five-year period.

Open Science Grid (OSG): OSG[19] is jointly funded by NSF (by OAC and the Directorate of Mathematical and Physical Sciences (MPS)) and the Department of Energy and facilitates access to distributed high throughput computing for research in the U.S. It provides common service and support for resource providers and scientific institutions using a distributed fabric of high throughput computational services. OSG provides software and services to users and resource providers alike to enable the opportunistic usage and sharing of resources. It is primarily used as a high-throughput grid where scientific problems are solved by breaking them down into a very large number of individual jobs that can run independently. The resources accessible through the OSG are contributed by the community, organized by the OSG, and governed by the OSG consortium. OSG consists of computing and storage elements at over 100 individual sites spanning the United States. These sites, primarily at universities and national labs, range in size

[16] "XSEDE Science Stories," https://www.xsede.org/news/science-stories.
[17] XD Metrics Service (XMS), https://open.xdmod.org/7.5/index.html.
[18] eXtreme Science and Engineering Discovery Environment (XSEDE), https://www.xsede.org/.
[19] Open Science Grid (OSG), https://opensciencegrid.org/.

from a few hundred to tens of thousands of CPU cores. During 2018, OSG provided more than 1.2 billion CPU hours to researchers across a wide variety of projects. OSG services are funded through 2019.

3.2.2 Software/Data CI and Services

NSF's data and software CI and services are primarily supported through OAC's programs that have provided long-term investments in catalyzing new thinking, paradigms, and practices in developing and using data and software services to understand natural, human, and engineered systems. The current flagship software/data program is the Cyberinfrastructure for Sustained Scientific Innovation (CSSI) umbrella program that aims to create an ecosystem of CI services that scales from individuals or small groups of researchers/innovators to large communities.

The CSSI program encompasses the long-running Data Infrastructure Building Blocks (DIBBs) and Software Infrastructure for Sustained Innovation (SI^2) programs. The Data Infrastructure Building Blocks (DIBBs) program has encouraged development of robust and shared data-centric CI capabilities to accelerate interdisciplinary and collaborative research in areas of inquiry stimulated by data. The Software Infrastructure for Sustained Innovation (SI^2) program recognized that software permeates all aspects and layers of CI (from application codes and frameworks, programming systems, libraries, and system software, to middleware, operating systems, networking, and the low-level drivers), and aimed to catalyze new software thinking, paradigms, and practices in science and engineering.

The CSSI program targets services that address all aspects of CI, from embedded sensor systems and instruments, to desktops and high-end data and computing systems, to major instruments and facilities. The program will continue to nurture the interdisciplinary processes required to support the entire data and software lifecycle and will successfully integrate development, deployment, and support of CI services with innovation and research. Furthermore, the program will result in the development of sustainable CI communities that transcend scientific and geographical boundaries. The program envisions vibrant partnerships among academia, government laboratories and industry, including international entities, for the development and stewardship of sustainable CI services that can enhance productivity and accelerate innovation in science and engineering. Integrated education activities will play a key role in developing and sustaining the software and data CI over time and in creating a workforce capable of fully realizing its potential to transform science and engineering.

NSF's data portfolio also includes the Big Data Regional Innovation Hubs program (BD Hubs), launched in 2015 and currently managed by OAC, that aims to nucleate regional collaborations and multi-sector projects, and foster innovation in data science. Four BD Hubs were funded as part of this program at each of the Census Regions of the country—Midwest, Northeast, South, and West. The BD Hubs serve as a venue for building and fostering local and regional data-related activity in city, county, and state governments, in local industry and non-profits, and in regional academic institutions. Collaborative activities and partnerships emerging from a regional focus contribute to building and sustaining a successful national big data innovation ecosystem. The current *Big Data Regional Innovation Hubs (BD Hubs) – Accelerating the Big Data Innovation*

Ecosystem solicitation (NSF 18-598)[20] seeks to continues the operation of a national network of BD Hubs. It builds on demonstrated strengths of the program, which has grown to include a set of BD Spokes affiliated with the BD Hubs, and is responsive to the recent developments in data science.

3.2.3 Networking

NSF's networking investments are spread across two distinct OAC programs, one domestic, i.e., Campus Cyberinfrastructure (CC*); and the second international in scope, i.e., International Research and Education Network Connections (IRNC). The CC* program began in 2012 as primarily a campus networking infrastructure and innovation program and has since expanded to support campus CI more broadly. Its investments include improvements and re-engineering at the campus level to support a range of data transfers supporting computational science and computer networks and systems research. Over the years the program has funded close to 200 awards, geographically distributed across most states. The current CC* solicitation (NSF 19-533)[21] aims to invest in coordinated campus-level networking and cyberinfrastructure improvements, innovation, integration, and engineering for science applications and distributed research projects (in addition to campus computing and learning and workforce development). Relevant program areas in the solicitation include Data-Driven Networking Infrastructure for the Campus and Researcher, Regional Connectivity for Small Institutions, and Network Integration and Applied Innovation.

The IRNC program provides high performance network capacity linking U.S. and other regions in the world in support of global research and education (R&E) collaborations, particularly those associated with international instruments and facilities. IRNC represents over 25 years of continuous NSF investment in international network in response to data driven needs of scientific advancement through global collaboration. IRNC awards support multiple 10Gbps connections between the U.S. and Asia, Europe, Africa and the Americas.

3.2.4 Cybersecurity

Cybersecurity for CI is an important area of investment at NSF through the focused OAC program "*Cybersecurity Innovation for Cyberinfrastructure (CICI)".* The CICI program and solicitation, created in 2015, focuses on the development and deployment of hardware and software technologies and techniques to protect research data and cyberinfrastructure across every stage of scientific workflows, which have grown increasingly complex as science is conducted collaboratively and with international partners. The current CICI solicitation (NSF 19-514)[22] aims to develop, deploy and integrate security solutions that benefit the scientific community by ensuring the integrity, resilience and reliability of the end-to-end scientific workflow. This solicitation seeks three categories of projects:

- Secure Scientific Cyberinfrastructure (SSC) aimed at securing the scientific workflow by encouraging novel and trustworthy architectural and design approaches, models and

[20] "Big Data Regional Innovation Hubs (BD Hubs) – Accelerating the Big Data Innovation Ecosystem," https://www.nsf.gov/pubs/2018/nsf18598/nsf18598.htm.
[21] "Campus Cyberinfrastructure (CC*)," https://www.nsf.gov/pubs/2019/nsf19533/nsf19533.htm.
[22] "Cybersecurity Innovation for Cyberinfrastructure (CICI)," https://www.nsf.gov/pubs/2019/nsf19514/nsf19514.htm.

frameworks for the creation of a holistic, integrated security environment that spans the entire scientific CI ecosystem.

- Research Data Protection (RDP) aimed at providing solutions that both ensure the provenance of research data and reduce the complexity of protecting research data sets regardless of funding source.
- Cybersecurity Center of Excellence (CCoE) seeking to provide the NSF community with a centralized resource of expertise and leadership in trustworthy cyberinfrastructure.

In addition, as part of the CI Cybersecurity portfolio, OAC is co-funding projects through the NSF-wide Secure and Trustworthy Cyberspace (SaTC) program[23] in the transition to practice (TTP) area.

3.2.5 Learning and Workforce Development (LWD)

NSF's CI Learning and workforce development (LWD) portfolio, through programs at OAC, supports research and education to prepare, nurture, and grow the national scientific workforce for creating, employing and supporting advanced cyberinfrastructures through cross-cutting programs in advanced cyberinfrastructure research; in education and training; and in research workforce development and career advancement. Projects funded as part of this portfolio support development of new knowledge in the innovative design, development, and utilization of robust research CI, sustainable and scalable models of education and training, and research workforce development or career advancement of current and future generations of one or more stakeholders. These stakeholders include the computing and computational and data-driven scientists and engineers who are researchers and developers of new capabilities, the research computing and professional staff who support productive use, sustainable maintenance, and effective governance of hardware and software systems, and the domain scientists and engineers who effectively exploit the advanced cyberinfrastructure capabilities.

While aspects of LDW are integrated in all OAC programs[24], a flagship program in this portfolio is *Training-based Workforce Development for Advanced Cyberinfrastructure (CyberTraining)*. The current Cybertraining solicitation (NSF 19-524)[25] seeks to prepare, nurture, and grow the national scientific research workforce. The goals of this solicitation are to (i) ensure broad adoption of CI tools, methods, and resources by the research community in order to catalyze major research advances and to enhance researchers' abilities to lead the development of new CI; and (ii) integrate core literacy and discipline-appropriate advanced skills in advanced CI as well as computational and data-driven science and engineering into the Nation's educational curriculum/instructional material fabric spanning undergraduate and graduate courses for advancing fundamental research. This solicitation calls for innovative, scalable training, education, and curriculum/instructional materials—targeting one or both of the solicitation goals—to address the emerging needs and unresolved bottlenecks in scientific and engineering research workforce development, from the postsecondary level to active researchers. The

[23] "Secure and Trustworthy Cyberspace (SaTC)," https://www.nsf.gov/pubs/2018/nsf18572/nsf18572.htm.

[24] For example, the current CC* solicitation (NSF 19-533[17]) explicitly targets facilitation of the use of campus computing resources use and will fund Cyber Team expertise.

[25] "Training-based Workforce Development for Advanced Cyberinfrastructure (CyberTraining)," https://www.nsf.gov/pubs/2019/nsf19524/nsf19524.htm.

funded activities, spanning targeted, multidisciplinary communities, will lead to transformative changes in the state of research workforce preparedness for advanced CI-enabled research in the short- and long-terms.

3.3 Core CI Research

Sustained investments in fundamental and translational CI research are critical to ensuring that NSF's CI ecosystem continues to effectively drive S&E research and innovation. The OAC research core program (NSF 18-567)[26] supports all aspects of advanced CI research that will significantly impact the future capabilities of advanced research CI, as well as the research career paths of computer as well as computational and data-driven scientists and engineers. In particular, OAC supports translational research and education activities in all aspects of advanced CI that lead to deployable, scalable, and sustainable systems capable of transforming science and engineering research. OAC's research investments are characterized by their translational nature, i.e., building on basic research results and spanning the design to practice stages. They are further characterized by one or more of the following key attributes: multi-disciplinary, extreme-scale, driven by science and engineering research, end-to-end, and deployable as robust research CI. OAC encourages translational research in areas including architectures and middleware for extreme-scale systems, scalable algorithms and applications, including simulation and modeling, and the advanced CI ecosystem, including tools and sociotechnical aspects.

3.4 NSF Large Facilities and CI

NSF's major multi-user research facilities (large facilities)[27] provide shared-use infrastructure, sophisticated research instruments and platforms, services and data products that are openly, reliably and pervasively accessible by a broad community of researchers and/or educators. These include large telescopes, interferometers and distributed sensor arrays and observatories, experimental facilities, etc., that serve diverse scientific disciplines from astronomy and physics to geoscience and biological science. NSF large facilities have the potential to deliver dramatic new insights in a wide range of S&E domains and are an essential part of the national science and engineering enterprise.

NSF large facilities are increasingly dependent on advanced CI. Current and planned large facilities are producing increasing volumes of data and data products resulting growing CI requirements, including computational, networking, data management and processing requirements, and associated expertise. At the same time, the user community that depends in these facilities is growing in size, diversity and sophistication, placing new requirements for example, for scalability, cybersecurity, interoperability, and data delivery. As a result, deploying and sustaining a robust, secure, performant and usable CI has become essential and is critical to the success of NSF large facilities.

CI aspects of the large facilities have been funded and managed by the relevant domain program(s) with little direct OAC (and ACI and OCI) involvement. However, as this CI grows in

[26] "Office of Advanced Cyberinfrastructure (OAC): Research Core Program," https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=505571.
[27] "NSF Large Facilities Office," https://www.nsf.gov/bfa/lfo/index.jsp.

scale and complexity, it is essential for the facilities and CI communities to collectively explore how to most effectively provide, sustain and evolve the essential components and services to meet current and future needs. Consequently, there is a growing opportunity for OAC and the CI community to partner with the facilities to address these challenges and develop facilities CI best practices.

The 2017 NSF Large Facilities Cyberinfrastructure Workshop[28] , which OAC funded, addressed this issue. The overarching objective of this workshop was to enable direct and synergistic interactions among the NSF large facilities and the cyberinfrastructure communities to jointly address the CI needs and sustainability of existing and future large facilities. A key goal was to develop a common understanding of the current and evolving requirements, architectures, and best practices; enabling technologies; operation practices and experiences; and issues and gaps. Key recommendations of the workshop include:

- Foster the creation of a facilities' CI community and establish mechanisms and resources to enable the community to interact, collaborate, and share;
- Support the creation of a curated portal and knowledge base to enable the discovery and sharing of CI-related challenges, technical solutions, innovations, best practices, personnel needs, etc., across facilities and beyond;
- Establish a center of excellence as a resource providing expertise in CI technologies and best practices related to large-scale facilities as they conceptualize, start up, and operate; and
- Establish structures and resources that bridge the facilities and that can strategically address workforce development, training, retention, career paths, and diversity, as well as the overall career paths for CI-related personnel.

NSF has been responding to these issues and recommendations. For example, OAC has funded the NSF Cybersecurity Center for Excellence (CCoE) [29] that has been actively assisting cyberinfrastructure projects by analyzing gaps in cybersecurity technology from a CI perspective, addressing the application of software assessment to complicated cyberinfrastructure software stacks, and fostering the transition of cybersecurity research to practice. CCoE has been working with multiple large facilities to provide expertise on achieving and maintaining effective cybersecurity programs. CCoE also organizes an annual Cybersecurity Summit that targets the NSF large facilities. More recently, OAC in partnership with the Directorate of Biological Sciences (BIO) funded a pilot Cyberinfrastructure Center of Excellence (CI CoE)[30]. The goal of this pilot project is to develop a model for such as center of excellence that can facilitate community building and sharing and apply knowledge of best practices and innovative solutions for facility CI. This pilot, in partnership with the National Ecological Observatory Network (NEON)[31] large facility, will explore how such a center would facilitate CI improvements for existing facilities and for the design of new facilities that exploit advanced CI architecture designs and leverage establish tools and solutions and address issues related to training and workforce development.

[28] "NSF Large Facilities Cyberinfrastructure Workshop," Alexandria, VA, September 2017, http://facilitiesci.org.
[29] "The NSF Cybersecurity Center for Excellence (CCoE)," https://trustedci.org/.
[30] "Cyberinfrastructure Center of Excellence (CI CoE)," http://cicoe-pilot.org.
[31] "National Ecological Observatory Network (NEON)," https://www.neonscience.org/.

3.5 NSF's Big Ideas and CI

In 2016, the NSF unveiled a set of "Big Ideas[12]" - 10 bold, long-term research and process ideas that identify areas for future investment at the frontiers of science and engineering. The Big Ideas represent unique opportunities to position our Nation at the cutting edge of global science and engineering leadership by bringing together diverse disciplinary perspectives to support convergence research. NSF is implementing the Big Ideas vision through focused NSF-wide programs and solicitations that involve all NSF directorates and offices. CI challenges and opportunities cut across and the NSF Office of Integrative Activities (OIA) and OAC are working with the other directorates and offices to provide CI leadership.

Among the NSF Big Ideas, Harnessing the Data Revolution (HDR)[32] explicitly addresses data CI. HDR is a national-scale activity to enable new modes of data-driven discovery that will allow fundamental questions to be asked and answered at the frontiers of science and engineering. The HDR vision is realized through a set of interrelated efforts seeking to establish theoretical, technical, and ethical frameworks that will be applied to tackle data-intensive problems in science and engineering, contributing to data-driven decision-making that impacts society. The HDR Institutes activity seeks to create an integrated fabric of interrelated institutes that can accelerate discovery and innovation in multiple areas of data-intensive science and engineering by harnessing diverse data sources and developing and applying new methodologies, technologies, and cyberinfrastructure for data management and analysis. The HDR Institutes will support convergence between science and engineering research communities as well as expertise in data science foundations, systems, applications, and CI. The conceptualization phase for HDR Institutes will be implemented in FY 2019 via two complementary funding opportunities. The Ideas Lab opportunity (NSF 19-543[33]) will facilitate the formation of interdisciplinary teams with critical expertise in different areas of science and engineering and data science to address important data-intensive challenges by recruiting participants from both the research (all science and engineering disciplines) and technical (computer and computational science, cyberinfrastructure, mathematics, statistics, and information science) domains. The second opportunity in FY 2019 will encourage applications from teams of researchers proposing frameworks for integrated sets of science and engineering problems and data science solutions.

3.6 Towards an Integrated CI Ecosystem

As noted above, NSF envisions *an agile, integrated, robust, trustworthy and sustainable CI ecosystem* that drives new thinking and transformative discoveries in all areas of S&E research and education, and we are strategically and tactically evolving our programs and investments towards achieving this vision of a holistic and integrated CI ecosystems, following the guiding principles outlined in the beginning of this Section and illustrated in Figure 4. This includes ensuring that our programs and investments across our portfolio can be integrated and composed so that they can be seamlessly brought together towards collectively addressing the end-to-end needs for current and future S&E research and education. Furthermore, we will

[32] "Harnessing the Data Revolution," https://www.nsf.gov/cise/harnessingdata/.

[33] "Harnessing the Data Revolution (HDR): Institutes for Data-Intensive Research in Science and Engineering - Ideas Labs (I-DIRSE-IL)," https://www.nsf.gov/pubs/2019/nsf19543/nsf19543.htm.

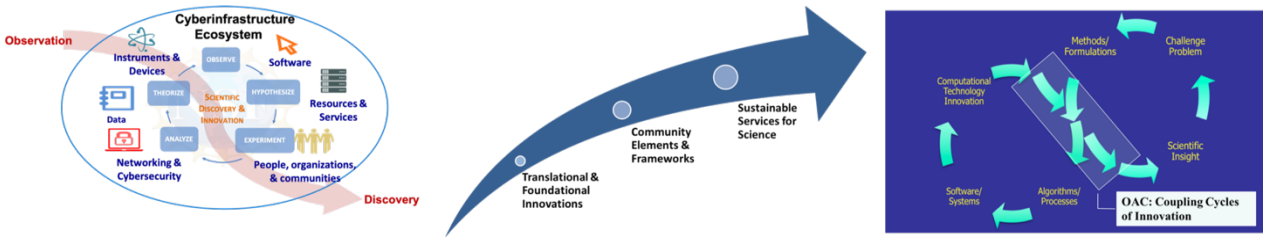


Figure 4. Achieving NSF's vision for a holistic, integrated CI ecosystem: (1) Ensuring components of the CI ecosystem can be composed to address S&E needs; (2) Supporting the translational research continuum; and (3) Coupling cycles of innovation across diverse S&E communities, ensuring usability, productivity and impact.

ensure that our programs support the translational research continuum, including catalyzing core CI innovations, the development of tools and frameworks, and the transitioning these tools and frameworks to sustainable cyberinfrastructure services and that our investments balance necessary innovations in CI with stability and continuity. And most importantly, we will work closely with the diverse S&E communities to tightly couple the cycles of discovery and innovation essential to addressing new challenges and opportunities, and to ensure usability and enhance scientists' productivity and science impact.

4 A Blueprint for a National Computational Ecosystem for Science and Engineering in the 21st Century

As noted in Section 1.2, this document has two goals. The first is to present NSF's vision for an integrated national CI ecosystem for enabling new, transformative discoveries across S&E, and the second is to outline NSF's blueprint for a computational ecosystem. This section focuses on the second goal and presents NSF's forward-looking blueprint for a computational ecosystem to support science and engineering research and education in the 21st century and outlines a plan to translate this blueprint into action and highlights extensions and enhancements to current investments as well as new programs and opportunities in 2019 and beyond. This computational ecosystem is complemented by the other components of the CI ecosystem including software and data services, networking, cybersecurity, and learning and workforce development. NSF (and OAC) expects to work with the community to develop complementary blueprints for these components in the future.

As detailed in Section 3.2, the vision and blueprint described in this document is informed by the community through advisory bodies, requests for information (RFIs), workshops and conferences, and national initiatives. Specifically, this document is influenced by and is, in part, a response to recent reports from a 2016 NASEM study and 2018 community workshops.

4.1 Elements of a balanced computational ecosystem

NSF aims to conceptualize and implement the advanced CI ecosystem that is critical to the advancement of all areas of S&E research and education. As a result, it is essential that we deploy a balanced computational ecosystem that can effectively support a broad and diverse set of requirements, users and usage modes. The key elements of the envisioned balanced computational ecosystem are illustrated in Figure 5, and include Leadership Class Systems, Capacity Systems, Federated Resources, Prototypes and Testbeds. Furthermore, such a computational ecosystem must include (and leverage) Campus CI investments as well as Cloud Services and must be composable to support end-to-end science needs. These elements evolve NSF current CI portfolio of investments as outlined in Section 3.2.1 in response to changing application/community needs and technology landscapes. Clearly, these elements are not mutually exclusive. For example, capacity or capability systems may include federated resources and leverage campus investments and/or cloud services. Also central to this ecosystem are the crosscutting coordination, allocation, measurement, and user support services. These key elements are discussed below.

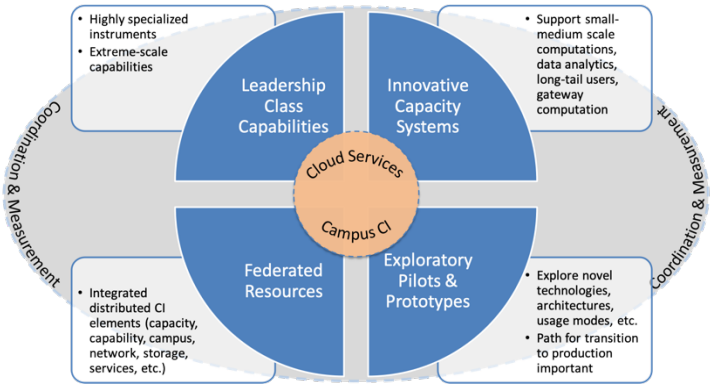


Figure 5: Elements of a balanced computational ecosystem.

Leadership Class Capabilities: Leadership class systems are highly-specialized computation and data-analytics instruments aimed at providing extreme-scale capability to support S&E research

that would not be possible otherwise. In recent decades, the development of powerful computing and data analytics capabilities have enabled ever larger and more detailed physical simulations and data-driven models, resulting in significant discoveries of new knowledge of the natural world, as well as improved economic competitiveness, health, prosperity and security of nations. As researchers in many areas of S&E continue to develop novel techniques for using computer simulation to advance their research, there has been an equally rapid expansion in the use of predictive data-driven models, derived from large experimental data sets often from disparate sources, in the scientific discovery workflow. As a result, advanced computational instruments with powerful computation and data analytics capabilities are now a fundamental tool in many areas of S&E, such as materials sciences, aerodynamics, climate and earth system dynamics, space weather modeling, seismic hazards modeling, cosmology, stellar astrophysics, biochemistry, biophysics, machine learning, economics and social science, bioinformatics, as well as many other topics within physics, engineering, biology and the geosciences.

Innovative Capacity System: Innovative capacity systems represent innovative approaches, architectures, and delivery models that maximize the computing capacity and throughput available to the NSF S&E community to support the broad range of computation and data analytics needs in a scalable and sustainable manner. These systems are motivated by current and future demand for computational and data analytics capacity in the broad and diverse S&E research community and target small to mid-scale jobs (from one to a few thousand cores per job) across broad areas of S&E, including support for "long-tail science" applications, as well as new classes of applications, such as artificial intelligence/machine learning/deep learning applications. The systems may also explore novel architecture that can leverage federated and/or distributed resources, regional and/or campus supported resources, and/or commercial cloud services.

Prototypes and Testbed Systems: This category targets systems that are initially deployed as a prototype/testbed system and are aimed at supporting S&E research through delivery of novel forward-looking capabilities and services. Systems in this category may explore, for example, the deployment of new technologies, system architectures, or usage modalities at scale, with plans for developing a national S&E user community that will benefit from these capabilities. Such systems could include novel processor architectures supporting artificial intelligence applications, distributed systems leveraging edge devices, domain-specific architectures and technologies, such as but not exclusively, reconfigurable and/or software defined systems, systems designed for streaming data and/or real-time processing, etc. It is important that investments in this category clearly define the target classes of application as well as have a clear plan for bringing these classes of applications to the system. While these systems are initially expected to be prototypes, it is expected that there is a clear plan for scaling these systems and transitioning them to production.

Distributed/federated CI: Distributed/federated CI systems and services support the integration of distributed capacity and capability systems, including systems in the above 3 categories as well as shared on-premise campus CI, in-network capabilities, resources at experimental facilities and observatories, and cloud services, with the goal of collectively supporting S&E research needs. These systems/services can support, for example, emerging science workflows, exploit data localities, and integrate local and regional resources.

Leveraging Campus CI, Cloud Partnerships: It essential that the computational ecosystem effectively integrates and leverages investments by NSF and others in campus CI. Campus CI represent resources managed and operated by campuses and serving campus, regions (and sometimes national) needs. These may or may not be located at the campus, and can be an aggregation of campus, regional and/or federal-funded resources. It is also essential to leverage capabilities and capacities provided by commercial cloud services.

4.2 Putting the Plan into Action: Programs, Projects and Opportunities

NSF is aggressively moving ahead to put the blueprint for the national computational ecosystem outlined in Section 4.1 into action through programs and projects in FY 2018, as well as in FY 2019 and beyond, as outlined in the next sections. These programs and projects target investments over a 10-year timespan (contingent as always on budgets and national priorities). This action plan has included enhancing current computational systems as well as issuing solicitations for the next generation of systems and services. NSF's action plan is summarized in Figure 6 and is also described.

4.2.1 Extending and Enhancing Current Investments

As NSF implements the blueprint outlined in the preceding sections, it is important to ensure continuity in the resources and services available to the S&E research and education community. NSF has ensured this by extending and enhancing its current portfolio of computational resources and services presented in Section 3.2.1. For example, supplemental funding was provided to ***Comet*** and ***Bridges*** for enhanced GPU capabilities to support machine learning, visualization, and advanced analytics, and to ***Jetstream*** to augment its system operation with focused staff expertise to accelerate effective researcher utilizations of its programmable

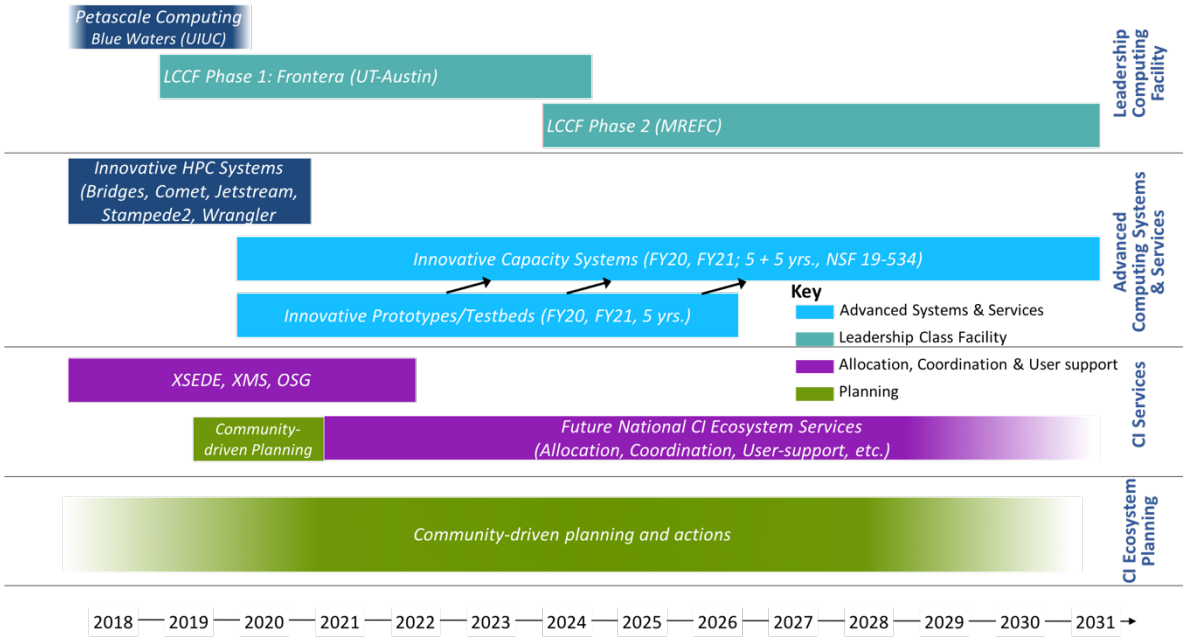


Figure 6: Putting the plan into action: Programs, projects and opportunities in 2019 and beyond.

cyberinfrastructure/virtual machine-enabled architecture. The period of operation for these systems was also extended to the FY 21-22 timeframe.

4.2.2 The Road Ahead – Programs, Projects and Opportunities in 2019 and Beyond

4.2.2.1 Future Leadership Class Facilities

In anticipation of the operational end date of Blue Waters, NSF current leadership class system, OAC issued a competitive solicitation (NSF 17-558)[34] in FY 2017 for the first phase (Phase 1) of a two-phase plan to deploy a new leadership-class computing facility. The solicitation called for a Phase 1 system that would be two to three times more powerful in application performance than Blue Waters. The operations and maintenance phase of the leadership-class HPC system is anticipated to commence in FY 2019 and last for five years under a separate award. The Phase 1 solicitation noted that, when fully deployed, the Phase 1 system would be expected to broadly extend new large-scale HPC benefits across the academic landscape, including support for previously unattainable research advances in large-scale modeling and simulation; use of robust data analytics at unprecedented scales for research; and use of HPC in dynamic workflows combining large-scale computation with big data streaming.

In addition to the acquisition of the Phase 1 system, the solicitation requested project plans for the conceptual design of a Phase 2 computing facility, which will enable a further 10 times performance improvement over the Phase 1 system. Phase 2 planning activities will be funded as supplements to this award and will be reviewed and approved according to NSF Major Research Equipment and Facilities Construction (MREFC) policies[35]. Phase 2 construction will be funded from the MREFC account. A related FY 2019 award to the Phase 1 awardee is anticipated to fund a Conceptual Design in anticipation of a future MREFC for Phase 2 of the leadership-class HPC system. The Phase 2 computing facility is expected to be deployed in 2024. A high-level overview of the ten-year plan for the NSF leadership-class facility is shown in Figure 7.

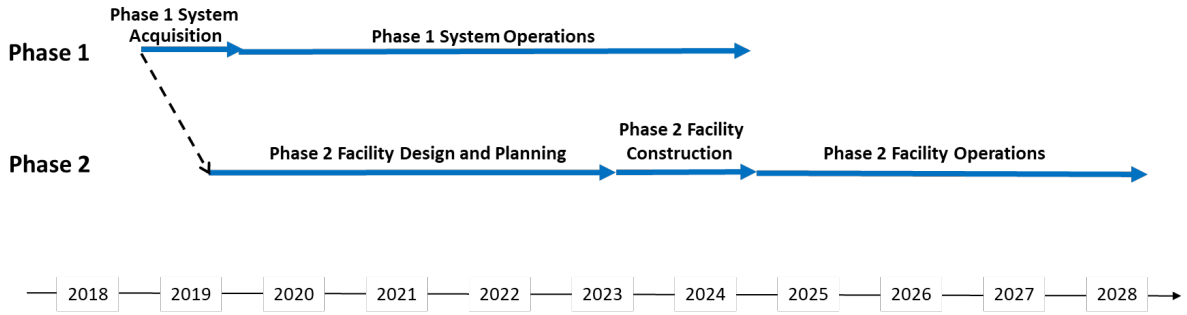


Figure 7: Overview of NSF's Ten-Year Plan for an NSF Leadership-Class Computing Facility.

In response to the Phase 1 solicitation, the National Science Board (NSB) at its July 2018 meeting authorized the Director at her discretion to make an award to TACC, UT-Austin for the acquisition of the system described in a proposal titled "Computation for the Endless Frontier," in an amount not to exceed $60,000,000 for a period of 60 months. The resolution also authorized that,

[34] "Towards a Leadership-Class Computing Facility - Phase 1," https://www.nsf.gov/pubs/2017/nsf17558/nsf17558.htm.
[35] "NSF Large Facilities Manual," (NSF 17-066), https://www.nsf.gov/pubs/2017/nsf17066/nsf17066.pdf.

pending appropriate approval associated with NSF MREFC policies, an additional amount not to exceed $8 million may be made available to TACC in the form of supplemental funding to this award to advance the design of the Phase 2 leadership-class system. Based on this authorization, the Phase 1 acquisition award for the Frontera systems was made to TACC in FY 2018[36].

4.2.2.2 Advanced Computing Systems and Services

Building on the innovative HPC investments and experiences as well as responding to community inputs, OAC released solicitation NSF 19-534[37] in FY 2019 titled *Advanced Computing Systems & Services: Adapting to the Rapid Evolution of Science and Engineering Research*. The intent of this solicitation is to request proposals from organizations willing to serve as service providers (SPs) within the NSF Innovative HPC program to provide advanced CI capabilities and/or services in production operations to support the full range of computational- and data-intensive research across all of S&E. The current solicitation is intended to complement previous investments in advanced computational infrastructure by provisioning resources, broadly defined in this solicitation to include systems and/or services, in two categories:

- Category I, Capacity Systems: production computational resources maximizing the capacity provided to support the broad range of computation and data analytics needs in S&E research; and
- Category II, Innovative Prototypes/Testbeds: innovative forward-looking capabilities deploying novel technologies, architectures, usage modes, etc., and exploring new target applications, methods, and paradigms for S&E discoveries.

Resources supported through awards from this solicitation will be incorporated into and allocated as part of NSF's Innovative HPC program. This program complements investments in leadership-class computing and funds a federation of nationally-available HPC resources that are technically diverse and intended to enable discoveries at a computational scale beyond the research of individual or regional academic institutions. As in previous solicitations, this solicitation recognizes the value and sustained institutional commitment required to building and retaining staff skilled in interdisciplinary computational and data science by allowing, at the discretion of NSF, a longer time horizon of eight to 10 years for a single institutional awardee. This timeline begins with a competitively awarded acquisition but allows for the possibility of a renewal acquisition award four years after the original award.

4.2.2.3 Catalyzing Campus CI Investments

NSF, through OAC programs, has consistently catalyzed investments in campus CI broadly and campus computational and data CI capabilities specifically through its programs and projects including the CC* program (introduced in Section 3.2.3)[38]. The goals of the CC* program include seeding or augmenting scientific computing access at the campus level and incentivize the

[36] "NSF awards $60 million for next-generation supercomputer," https://nsf.gov/news/news_summ.jsp?cntn_id=296431.
[37] "Advanced Computing Systems & Services: Adapting to the Rapid Evolution of Science and Engineering Research," https://www.nsf.gov/pubs/2019/nsf19534/nsf19534.htm.
[38] The OAC CC* is complementary to NSF-wide instrumentation program such as Major Research Instrumentation (MRI), https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=5260.

sharing and coordination of campus computing investments across campuses. While campus networking is a key focus on CC*, the solicitation also allows for campus computing strategies that explore the integration of cloud resources into the set of capabilities provided across scientific and engineering research campus-wide. Specifically, the current CC* solicitation (NSF 19-533)[21] recognizes that local campus computing resources have emerged as an important aggregated and shared layer of scientific computing. The *Campus Computing and the Computing Continuum* program area in the solicitation promotes coordinated approaches for scientific computing at the campus level as well as for incentivizing multi-campus and national resource sharing. Specifically, the *Campus Cluster Resource* categories aims at seeding and augmenting shared computing resources at the campus level through investments in capacity computing in campus clusters by providing funding for the acquisition of a shared, high-performance network-connected compute resource available to scientific computing users on campus and outside of campus. The program will also support campus use of commercial cloud services as well hybrid campus computing environments that integrate cloud and campus computing services, as described below.

4.2.2.4 Leveraging Cloud Services

Cloud services are playing and increasingly important role in supporting academic and research computing complementing more traditional research/academic computing environments such as specialized high-end systems and campus infrastructure. Furthermore, recent years have seen an increasing convergence between cloud systems and research/academic computing systems in terms of their architectures and technologies, as well as their service-based usage model. As a result, NSF is exploring models and mechanisms for partnering with commercial cloud service providers and for integrating cloud services as part of its CI ecosystem. Currents efforts in this direction include:

1. The current CC* solicitation (NSF 19-533) as part of its *Campus Computing and the Computing Continuum* program area is seeking to enable campuses to explore the potential use of cloud computing services and its analytics platforms, as well as combinations of both local computing resources and access to remote cloud computing, in supporting their community's scientific research computing assets and available resources. NSF is partnering with Amazon Web Service and Google Cloud Platform as part of this solicitation to provide cloud credits/resources to campuses whose scientific research requires additional and external computational and storage resources. These providers are also expected to provide support and training to those campuses.
2. The current solicitation NSF-19-510[39] titled "*Enabling Access to Cloud Computing Resources for CISE Research and Education (Cloud Access)"* seeks to fund an entity that can serve as a principal interface between the CISE research and education community and public cloud computing providers. Through this solicitation, NSF will support an entity that will have multiple responsibilities, including: 1) establishing partnerships with the various public cloud computing providers; 2) assisting NSF in allocating cloud computing resources to qualifying

[39] "Enabling Access to Cloud Computing Resources for CISE Research and Education (Cloud Access)," https://www.nsf.gov/pubs/2019/nsf19510/nsf19510.htm.

CISE-funded projects; 3) managing cloud computing accounts and resources allocated to individual CISE projects; 4) providing user training and other support to CISE researchers and educators using cloud computing in their work; and 5) providing strategic technical guidance for CISE researchers and educators interested in using public cloud computing platforms.

3. NSF continues to explore the role of Cloud services in enabling S&E research and education. For example, in FY18, OAC funded Internet2, a nonprofit computer networking consortium, to build partnerships with commercial cloud computing providers and support science applications in new and more effective uses of cloud computing capabilities. As part of the project, *Exploring Clouds for Acceleration of Science (E-CAS)*[40], Internet2 is partnering with Amazon Web Services, Google Cloud Platform and Oracle Cloud[41] to investigate the viability of commercial clouds as an option for leading-edge research computing and computational science supporting a range of academic disciplines. The E-CAS project aims to help researchers understand the potential benefit of larger-scale commercial platforms for simulation and application workflows such as those currently using NSF's High-Performance Computing (HPC), and explores how scientific workflows can innovatively leverage advancements in real-time analytics, artificial intelligence, machine learning, accelerated processing hardware, automation in deployment and scaling, and management of serverless applications in order to provide digital research platforms to a wider range of science.

4.2.2.5 Monitoring and measurement services

Deep instrumentation, monitoring, measurement, and reporting across all layers of the systems and services making up the CI ecosystem are essential to providing the situational awareness necessary for achieving desired levels of efficiency, understanding, autonomous operations, robustness, and performance. Providers, operators (those responsible for the CI), and users require a comprehensive system view into CI structure, behavior and operation. The user and her application workflow should have access to services and capabilities available at execution and should be able to adapt to changing system conditions. Similarly, operators and providers require their own forms of visibility in order to deliver production quality CI. NSF's CI investments continue to integrate such monitoring and measurement capabilities with the overarching goal of achieving the desired levels of instrumentation and insight. For example, OAC has invested in a portfolio of award activities in monitoring and measurement research and development over the years, ranging from distributed HPC usage metrics (e.g., XSEDE's XDMoD[17]) to networking-based monitoring and measurement (e.g., CAIDA[42], Route Views[43], NetSage[44] and perfSONAR[45]). One challenge forward is how to build on this base of experience, knowledge, and capabilities to

[40] "Exploring Clouds for Acceleration of Science," https://www.internet2.edu/vision-initiatives/initiatives/exploring-clouds-acceleration-science/.

[41] Note that other NSF programs have also included partnerships with Microsoft Azure and IBM, such as for example, "Critical Techniques, Technologies and Methodologies for Advancing Foundations and Applications of Big Data Sciences and Engineering (BIGDATA)," https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=504767.

[42] "Center for Applied Internet Data Analysis," http://www.caida.org/home/.

[43] "University of Oregon Route Views Project," http://www.routeviews.org/routeviews/.

[44] "NetSage," http://www.netsage.global/.

[45] "perfSONAR," https://www.perfsonar.net/.

achieve greater insights and understanding on the behavior, performance, and use of scientific CI environments.

5 Ongoing Strategic Planning and Community Engagement

NSF will continue its strategic planning and community engagement activities as it explores other elements of computational ecosystem such as, for example, the crosscutting coordination, allocation, measurement, and user support services currently provided by the Extreme Digital (XD) program. Given the importance of such services to the effective use of the CI ecosystem and recognizing the dramatic changes in the nature and requirements of applications supported by this CI ecosystem as well as the landscape of CI technologies and resources, such strategically thinking about the future nature and implementation of these services to respond to changing application and CI landscapes is essential to ensure that they continue to enable science in the 21st century. It is also similarly important to think strategically as a community about other elements of the national CI ecosystem, including software and data systems and services, networking, cybersecurity, and learning and workforce development. As previously noted, NSF expects that this blueprint is the first of several similar blueprints focused on different elements of the CI ecosystem that will be developed in partnership with the community.

Looking to the future, the road ahead for research CI promises to be exciting with many new opportunities. In the near term, NSF Big Ideas are moving into full gear with multiple new solicitations and are complemented by recent relevant national initiatives in areas such as Quantum Information Science (including a National Quantum Initiative Act signed into law)[46] and Artificial Intelligence (including an Executive Order on Maintaining American Leadership in Artificial Intelligence)[47]. Furthermore, NSF's investments, for example, as part of the Quantum Leap Big Idea[48], and the Quantum Computing & Information Science Faculty Fellows (QCIS-FF) program (NSF 19-507)[49], as well as its investments in fundamental and applied Artificial Intelligence and Machine Learning[50] research, will help define the nature and structure of the CI ecosystem over the longer term. NSF looks forward to continuing to work with the community to define the future of cyberinfrastructure research and research cyberinfrastructure, with the overarching goal of realizing and integrated CI ecosystem that transforms all of S&E research and education.

6 Conclusion

The NSF-funded CI ecosystem is playing an increasingly critical role across all of S&E research and education, enabling discoveries and driving innovation. It is an important part of the national CI ecosystem that is critical for ensuring US leadership in S&E, economic competitiveness and

[46] "National Quantum Initiative," https://www.congress.gov/115/bills/hr6227/BILLS-115hr6227enr.pdf.
[47] "Executive Order on Maintaining American Leadership in Artificial Intelligence," https://www.whitehouse.gov/presidential-actions/executive-order-maintaining-american-leadership-artificial-intelligence/.
[48] "NSF Quantum Leap Big Idea," https://www.nsf.gov/news/special_reports/big_ideas/quantum.jsp.
[49] "Quantum Computing & Information Science Faculty Fellows (QCIS-FF)," https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=505535.
[50] "Artificial Intelligence (AI) at NSF," https://nsf.gov/cise/ai.jsp.

national security. As a result, it is essential that NSF strategically rethink and evolve this CI ecosystem in response to the changing nature of needs of S&E, driven by the changing technology landscape, and informed by community inputs. This document articulated NSF's vision for a national CI ecosystem that integrates computational, data, software, networking, and security resources, tool and services, and computational and data skills and expertise towards collectively enabling new, transformative discoveries across all of S&E. It also presented a snapshot of NSF's CI ecosystem, supported through OAC, highlighting its current and ongoing investments. The document then presented NSF's blueprint for a computational ecosystem to support science and engineering research and education in the 21st century. Finally, it outlined a plan to implement this blueprint, highlighting extensions and enhancements to current investments as well as new programs and opportunities in 2019 and beyond. The vision and blueprint presented in this document have been informed by the community through advisory bodies, requests for information (RFIs), workshops and conferences, and national initiatives. NSF intends to continue to work with the community to evolve and implement the vision and blueprint presented in this document, as well as to develop complementary blueprints for other CI elements.