Future Directions for NSF Advanced Computing Infrastructure to Support U.S. Science and Engineering in 2017-2020

The National Science Foundation (NSF) asked the National Academies of Sciences, Engineering, and Medicine to provide a framework for future decision-making about NSF's advanced computing strategy and programs. Advanced computing refers here to the advanced technical capabilities, including computer systems, software, and expert staff, that support a wide range of science and engineering research and that are of a large enough scale and cost that they are typically shared among multiple researchers, institutions, and applications. Advanced computing encompasses support for data-driven research as well as modeling and simulation. The recommendations of the Committee on Future Directions for NSF Advanced Computing Infrastructure to Support U.S. Science in 2017-2020 are aimed at achieving four broad goals: (1) positioning the United States for continued leadership in science and engineering, (2) ensuring that resources meet community needs, (3) aiding the scientific community in keeping up with the revolution in computing, and (4) sustaining the infrastructure for advanced computing.

POSITION THE UNITED STATES FOR CONTINUED LEADERSHIP IN SCIENCE AND ENGINEERING

Large-scale simulation and the accumulation and analysis of massive amounts of data are revolutionizing many areas of science and engineering research. Increased advanced computing capability has historically enabled new science, and many fields today rely on high-throughput computing for discovery. Modeling and simulation, the historical focus of high-performance computing, is a well-established peer of theory and experiment. Data-driven research, a complementary "fourth paradigm" for scientific discovery, needs data-intensive computing capabilities and resources. To support this research, NSF is a major provider of the advanced computing used for U.S. basic science, for not only its own grantees but also in support of research sponsored by other agencies, such as the National Institutes of Health and the Department of Energy.

Meeting future needs will require systems that support a wide range of advanced computing capabilities, including large-scale parallel systems and data-intensive systems. Approaches that combine large-scale computing and data resources in "converged" systems can play a role; more specialized systems may also be needed to meet some requirements. Commercial

Read, purchase, or download a free PDF of this report at http://www.nap.edu

The National Academies of
SCIENCES · ENGINEERING · MEDICINE

cloud computing offers certain advantages and can play a role in NSF's advanced computing strategy. However, NSF computing centers already exploit economies of scale and load sharing, and commercial cloud providers do not currently support very large, tightly coupled parallel applications, especially for high-end simulation workloads. For other applications, especially data-centric workloads and communities that share data sets, cloud computing is positioned today to play a growing role.

RECOMMENDATION 1: NSF should sustain and seek to grow its investments in advanced computing—to include hardware and services, software and algorithms, and expertise—to ensure that the nation's researchers can continue to work at frontiers of science and engineering.

> **RECOMMENDATION 1.1:** NSF should ensure that adequate advanced computing resources are focused on systems and services that support scientific research. In the future, these requirements will be captured in its roadmaps.

> **RECOMMENDATION 1.2:** Within today's limited budget envelope, this will mean, first and foremost, ensuring that a predominant share of advanced computing investments be focused on production capabilities and that this focus not be diluted by undertaking too many experimental or research activities as part of NSF's advanced computing program.

> **RECOMMENDATION 1.3:** NSF should explore partnerships, both strategic and financial, with federal agencies that also provide advanced computing capabilities as well as federal agencies that rely on NSF facilities to provide computing support for their grantees.

RECOMMENDATION 2: As it supports the full range of science requirements for advanced computing in the 2017-2020 timeframe, NSF should pay particular attention to providing support for the revolution in data-driven science along with simulation. It should ensure that it can provide unique capabilities to support large-scale simulations and/or data analytics that would otherwise be unavailable to researchers and continue to monitor the cost-effectiveness of commercial cloud services.

> **RECOMMENDATION 2.1:** NSF should integrate support for the revolution in data-driven science into NSF's strategy for advanced computing by (a) requiring most future systems and services and all those that are intended to be general purpose to be more data-capable in both hardware and software and (b) expanding the portfolio of facilities and services optimized for data-intensive as well as numerically-inten-

sive computing, and (c) carefully evaluating inclusion of facilities and services optimized for data-intensive computing in its portfolio of advanced computing services.

> **RECOMMENDATION 2.2:** NSF should (a) provide one or more systems for applications that require a single, large, tightly coupled parallel computer and (b) broaden the accessibility and utility of these large-scale platforms by allocating high-throughput as well as high-performance workflows to them.

> **RECOMMENDATION 2.3:** NSF should (a) eliminate barriers to cost-effective academic use of the commercial cloud and (b) carefully evaluate the full cost and other attributes (e.g., productivity and match to science workflows) of all services and infrastructure models to determine whether such services can supply resources that meet the science needs of segments of the community in the most effective ways.

Maintaining leadership in advanced computing will be challenging. The resources available for advanced computing are inherently limited by research budgets, even as the demand for computing is growing and changing rapidly across the scientific enterprise and as the gap between supply and demand grows. If NSF is unable to increase or better leverage its resources for advanced computing, it seems inevitable that it will be unable to meet future demand for computational resources and will have to reduce the size of the very largest research projects that are supported by its advanced computing facilities.

ENSURE THAT RESOURCES MEET COMMUNITY NEEDS

Despite various ongoing efforts to collect and understand requirements from some science communities and occasional efforts to chart strategic directions, the overall planning process for advanced computing resources and programs is not systematic or uniform and is not visibly reflected in NSF's strategic planning, despite its foundation-wide importance. The creation of an ongoing and more regular and structured process would make it possible to collect requirements, roll them up, and prioritize advanced computing investments based on science and engineering priorities.

RECOMMENDATION 3: To inform decisions about capabilities planned for 2020 and beyond, NSF should collect community requirements and construct and publish roadmaps to allow NSF to set priorities better and make more strategic decisions about advanced computing.

> **RECOMMENDATION 3.1:** NSF should inform its strategy and decisions about investment trade-offs

using a requirements analysis that draws on community input, information on requirements contained in research proposals, allocation requests, and foundation-wide information gathering.

RECOMMENDATION 3.2: NSF should construct and periodically update roadmaps for advanced computing that reflect these requirements and anticipated technology trends to help NSF set priorities and make more strategic decisions about science and engineering and to enable the researchers that use advanced computing to make plans and set priorities.

RECOMMENDATION 3.3: NSF should document and publish on a regular basis the amount and types of advanced computing capabilities that are needed to respond to science and engineering research opportunities.

RECOMMENDATION 3.4: NSF should employ this requirements analysis and resulting roadmaps to explore whether there are more opportunities to use shared advanced computing facilities to support individual science programs such as Major Research Equipment and Facilities Construction projects.

The roadmaps would reflect the visions of the science communities supported by NSF, including both large users and those (in the "long-tail") with more modest needs. The goal is to develop brief documents that set forth the overall strategy and approach rather than high-resolution details. They would look roughly 5 years ahead and provide a vision that extends about 10 years ahead. The roadmaps would help inform users about future facilities, guide investment, align future procurements and services with requirements, and enable more effective partnerships within NSF and with other federal agencies.

The roadmapping and requirements process could be strengthened by developing a better understanding of the relationships among requirements, the costs of different approaches (roadmap choices), and science benefits. Such information would inform program managers about the total cost of proposed research, help focus researcher attention on effective use of these valuable shared resources, and encourage more efficient software and research techniques.

RECOMMENDATION 4: NSF should adopt approaches that allow investments in advanced computing hardware acquisition, computing services, data services, expertise, algorithms, and software to be considered in an integrated manner.

RECOMMENDATION 4.1: NSF should consider requiring that all proposals contain an estimate of the

advanced computing resources required to carry out the proposed work and creating a standardized template for collection of the information as one step of potentially many toward more efficient individual and collective use of these finite, expensive, shared resources. (This information would also inform the requirements process.)

RECOMMENDATION 4.2: NSF should inform users and program managers of the cost of advanced computing allocation requests in dollars to illuminate the total cost and value of proposed research activities.

AID THE SCIENTIFIC COMMUNITY IN KEEPING UP WITH THE REVOLUTION IN COMPUTING

Computer architectures are changing rapidly along with programming models to use the hardware, creating challenges for the science community, which depends on and has invested significantly in science codes written for yesterday's systems. The rise of data-intensive science brings with it new software and systems. Better software tools, technical expertise, and more flexible service models (ways of delivering software and computing resources) can improve the productivity of researchers both today and in the future.

RECOMMENDATION 5: NSF should support the development and maintenance of expertise, scientific software, and software tools that are needed to make efficient use of its advanced computing resources.

RECOMMENDATION 5.1: NSF should continue to develop, sustain, and leverage expertise in all programs that supply or use advanced computing to help researchers use today's advanced computing more effectively and prepare for future machine architectures.

RECOMMENDATION 5.2: NSF should explore ways to provision expertise in more effective and scalable ways to enable researchers to make their software more efficient; for instance, by making more pervasive the XSEDE (Extreme Science and Engineering Discovery Environment) practice that permits researchers to request an allocation of staff time along with computer time.

RECOMMENDATION 5.3: NSF should continue to invest in and support scientific software and update the software to support new systems and incorporate new algorithms, recognizing that this work is not primarily a research activity but rather is support of software infrastructure.

If NSF were to invest solely in production, it would miss some key technology shifts, and its facilities would become obsolete quickly. By taking a leadership role in defining future advanced computing capabilities and helping researchers use them more effectively, NSF can help ensure that its software and systems remain relevant to its science portfolio, that researchers are prepared to use the systems, and that investments across the foundation are aligned with this future.

RECOMMENDATION 6: NSF should also invest modestly to explore next-generation hardware and software technologies to explore new ideas for delivering capabilities that can be used effectively for scientific research, tested, and transitioned into production where successful. Not all communities will be ready to adopt radically new technologies quickly, and NSF should provision advanced computing resources accordingly.

SUSTAIN THE INFRASTRUCTURE FOR ADVANCED COMPUTING

Expertise and other long-lived assets, such as the physical infrastructure for computing centers, are an essential part of a robust and sustainable advanced cyberinfrastructure. In recent years, NSF has adopted a strategy for acquiring computing facilities and creating centers and programs to operate and support them that relies on irregularly scheduled competition among host institutions roughly every 2 to 5 years and on equipment, facility, and operating cost sharing with those institutions. Mounting costs and budget pressures suggest that a strategy that relies on state, institutional, or vendor cost sharing may no longer be viable. Repeated competition can lead to proposals designed to win a competition rather than maximize scientific returns. Moreover, it is important to ensure the development and retention of the talent that is needed to effectively manage systems, support users, and evolve software to make effective use of today's and tomorrow's architectures.

RECOMMENDATION 7: NSF should manage advanced computing investments in a more predictable and sustainable way.

> **RECOMMENDATION 7.1:** NSF should consider funding models for advanced computing facilities that emphasize continuity of support.

> **RECOMMENDATION 7.2:** NSF should explore and possibly pilot the use of a special account (such as that used for Major Research Equipment and Facilities Construction) to support large-scale advanced computing facilities.

> **RECOMMENDATION 7.3:** NSF should consider longer-term commitments to center-like entities that can provide advanced computing resources and the expertise to use them effectively in the scientific community.

> **RECOMMENDATION 7.4:** NSF should establish regular processes for rigorous review of these center-like entities and not just their individual procurements.

Managing its advanced computing investments in a more predictable and sustainable way, as it does for other long-term (10 years or more) infrastructure, would not only benefit the researchers currently supported by NSF's advanced computing programs but also would provide opportunities to apply the same expertise more broadly within NSF, such as the large-scale science projects that have long-term needs for advanced computing. It would also create new opportunities for NSF's advanced computing programs to address long-term storage, preservation, and curation challenges for data.

Copies of this report are available free of charge from http://www.nap.edu.