U.S. DEPARTMENT OF ENERGY

May 2024

2024 ADVANCED SCIENTIFIC COMPUTING ADVISORY COMMITTEE

FACILITIES SUBCOMMITTEE RECOMMENDATIONS

Ed Seidel
University of Wyoming, Chair

Amanda Randles
Duke University, Co-Chair

Rick Arthur
GE Aerospace

Keren Bergman
Columbia University

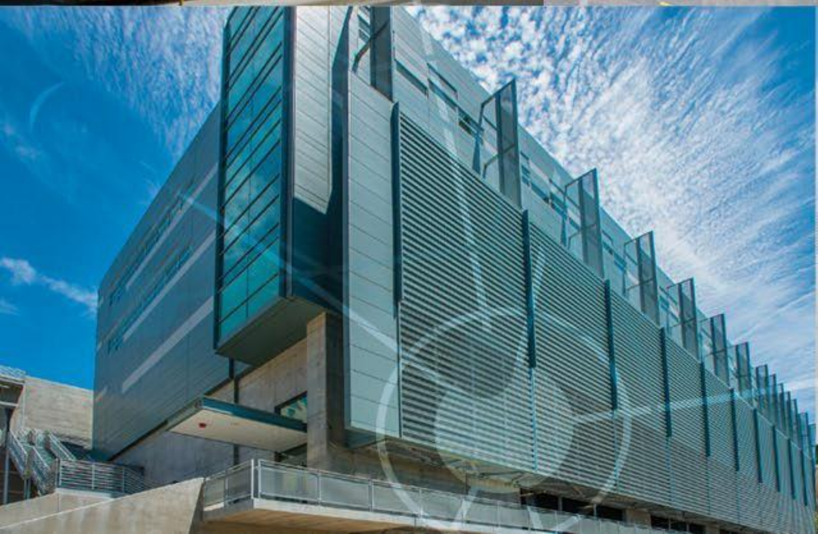
Bill Carlson
IDA Center for Computing Sciences

Ewa Deelman
University of Southern California

Ray Grout
National Renewable Energy Laboratory

Bruce Hendrickson
Lawrence Livermore National Laboratory

Dan Reed
University of Utah



On the cover

Clockwise, from left to right OLCF at Oak Ridge National Laboratory, ALCF at Argonne National Laboratory,
NERSC at Lawrence Berkeley National Laboratory, HPDF at Jefferson Lab, ESnet at Lawrence Berkeley National Laboratory.

DOI: https://doi.org/10.2172/2370379

CONTENTS

2024 Advanced Scientific Computing Advisory Committee
Facilities Subcommittee Recommendations

Membership
Ed Seidel, University of Wyoming; Amanda Randles, Duke University; Co-Chairs
Rick Arthur, GE Aerospace
Keren Bergman, Columbia University
Bill Carlson, IDA Center for Computing Sciences
Ewa Deelman, University of Southern California
Ray Grout, National Renewable Energy Laboratory
Bruce Hendrickson, Lawrence Livermore National Laboratory
Dan Reed, University of Utah

Executive Summary

We were given a charge to assess the necessity for new or upgraded facilities to ensure the Office of Science (SC) remains at the forefront of scientific discovery. This effort included evaluating five specific ASCR facilities for their potential to contribute to this goal and to rate the readiness for construction of each.

Our analysis led us to three overarching recommendations:

> **Recommendation 1: Ensure the continued support and development of all five ASCR computational facilities reviewed—ALCF, OLCF, NERSC, HPDF, and ESnet—as they are central and essential to all SC science programs and broader national science and engineering research programs.** Each facility provides distinct and critical functionality that are essential to achieve SC science goals. Significant R&D investment is necessary to sustain their crucial roles. A summary of our findings can be found in Table 1. Recommendations on individual facilities are found in Section 3.

> **Recommendation 2: Science demands integration. We advocate viewing ASCR facilities not as isolated entities, but as integral components of a single, larger integrated computational *ecosystem* (henceforth referred to as *Ecosystem)*, with a single governance model.** This effort will require new ways of governing and potentially funding the overall ***Ecosystem***, which should not be developed via individual site procurements. Rather it should be designed, developed, built, and operated as an integrated facility ***ecosystem*** for DOE science. It is critical for supporting SC science programs, along with additional software, algorithm, workforce, and science application components, to serve science and engineering research. *Further, this integrated ecosystem is required for programs of other agencies, and industry. Its critical role in bolstering national scientific and technological capabilities, as well as its status as a model internationally, cannot be overstated*.

> **Recommendation 3: A comprehensive, coordinated R&D program delivering multiple prototype computing systems over a five-year timescale must be mounted to inform**

pathways for this integrated ecosystem, operational by 2034, due to (a) rapidly evolving economic and technical landscapes of the semiconductor and computing industries and (b) changing research practices. Despite varying readiness for construction, ASCR facilities collectively require a comprehensive R&D strategy for their future development. With the end of Moore's Law and with vendors focusing on other markets, the future of high-end computing for science is highly uncertain. R&D is needed to chart this course, to influence vendors, and to prepare applications for future platforms. The changing nature of interdisciplinary research requires a deeper integration of facilities, workflows, algorithms, software and application tools. *The R&D program should involve a collaborative effort across DOE computational facilities spanning SC. To ensure a comprehensive, multidisciplinary approach, it is critical to include contributions from computing and cloud vendors, computer science, DOE experimental facilities, domain science and engineering communities, with deep collaborations with DOE NNSA and other federal agencies*. A new governance model will be required to manage this.

Our specific charge included analyzing facilities stewarded by ASCR, each of which provides unique and scientifically necessary capabilities as components of an evolving integrated national ecosystem. Our findings are summarized in Table 1,

Table 1. Summary of findings responding to the charge to our ASCR subcommittee from SC Director Berhe. Details are provided in our key findings and recommendations below and in Section 3.

Facility	Description	Importance to SC Science Mission	Readiness for Construction for 2034 deployment
ALCF	Leadership Class computing facility	Absolutely central; required for success of science mission	ALCF-4: Ready to initiate construction ALCF-5: Significant scientific and engineering challenges to resolve before construction
OLCF	Leadership Class computing facility	Absolutely central; required for success of science mission	OLCF-6: Ready to initiate construction OLCF-7: Significant scientific and engineering challenges to resolve before construction
NERSC	High performance production scientific computing center	Absolutely central; required for success of science mission	NERSC-10: Ready to initiate construction NERSC-11: Significant scientific and engineering challenges to resolve before construction
ESnet	High performance networking; connects computing & experimental facilities across SC	Absolutely central; required for success of science mission	ESnet-7: Ready to initiate construction ESnet-8: Significant scientific and engineering challenges to resolve before construction
HPDF	Distributed data-focused facility with hub-and-spoke architecture. Will provide unique data storage and management capabilities.	Absolutely central; required for success of science mission	HPDF Hub: Significant scientific and engineering challenges to resolve before construction HPDF Spokes 1 and 2: Mission and technical requirements not yet fully defined

These findings are corroborated through our interviews with other SC facility subcommittee chairs, which included multiple meetings and the exchange of preliminary findings. They unanimously confirmed the critical role of ASCR facilities in achieving scientific objectives.

It is important to emphasize that *not only do other science agencies depend on ASCR facilities, but U.S. industry does as well, as an essential and growing component of the user base, and as a contributor to the technologies needed to support the development of the facilities*. A coordinated all-of-government approach is needed for success. Additionally, industry will need to be actively involved.

Summary of Key Findings

1. **ASCR advanced computing systems continue to be critical for SC to remain at the forefront of scientific discovery as science becomes more interdisciplinary, integrated, and digital.** DOE science programs require large experimental facilities, theory, and leading-edge computation, data analysis and storage, and advanced networking. This includes new applications and integrated workflows for data management, artificial intelligence (AI)/machine learning (ML), and physics-based simulation and modeling. The success of programs managed by ASCR, BES, BER, HEP, FES, and NP depends on the success of ASCR facilities, a point reinforced by interviews with other SC subcommittees.
2. **A combination of complementary facilities is needed to support DOE mission science, and through their integration should be thought of as a single overarching ecosystem with multiple components.** These components provide complementary functions and are integrated and synergistic. Together, they span the set of capabilities needed to support DOE mission science in the coming decade. We do not see how one can be funded over another without risking science goals across all SC programs.
3. **ASCR facilities are also important to other organizations, including but not limited to, NNSA, NSF, NIH, NIST, NASA, NOAA, and DoD, and to U.S. industrial competitiveness.** The impact of ASCR facilities extends beyond SC to support scientific discovery, U.S. industry, and global economic competitiveness. By virtue of its investment and expertise, DOE is, *de facto*, the lead federal agency for advanced scientific computing. If DOE fails to lead R&D in the face of a changing computing world, the rest of the country will suffer as well.
4. **Continued success of DOE mission science requires an “all hands on deck” approach to developing next-generation computing infrastructure.** Because ASCR facilities operate in rapidly evolving economic and technical landscapes of the semiconductor and computing industries and changing research practices, *progress cannot be business as usual*. This committee views it as critical for ASCR to leverage expertise beyond the boundaries of the Office of Science to conduct the necessary R&D and attract vendors in R&D partnerships.

In summary, ASCR and the Office of Science are international leaders in both the capabilities of their computational and networking facilities and the breadth of the science programs that depend vitally on these resources, spanning DOE and other agencies. However, given the rapidly changing landscape of interdisciplinary research, international competition, and the highly uncertain and disruptive technological future of high-end computing, our committee unanimously agrees that to maintain this scientific leadership, "business as usual" cannot continue. By "business as usual," we refer to the isolated procurement of individual computing systems based on variations of commercial market product evolution; *a new approach is needed*. This new approach should treat computational, data management, and network capabilities as an integrated ecosystem. It should include a comprehensive R&D program involving all programs of SC and other agencies to inform future facility deployment, and adopt a more sustainable and collaborative approach to workforce development and retention.

Failure to follow these strategies risks loss of international leadership in science/engineering research and in advanced computing, with long-term implications on US national security, technological leadership, and economic competitiveness. Our recommendations offer opportunities for positioning SC to achieve science goals, enhancing beneficial partnerships across US and select international science agencies, and sustaining U.S. international leadership in key areas.

Section 1 Introduction and Considerations

Science and engineering research are evolving rapidly, addressing more complex questions that require a deeper integration of disciplines (e.g., physics, chemistry, biosciences, mathematics, computer science, and engineering) and methodologies (experimental facilities, theory, data analytics, and computational approaches). While DOE science areas each may have specific and different kinds of facilities on which they depend (e.g., accelerators, light sources, and other instruments), they all share a common need for significant and growing computational and data analysis resources. Instruments are producing data at exponentially growing rates, theory requires high-end modeling and simulation to make specific predictions, and new computational methods (e.g., AI/ML), are revolutionizing our ability to draw scientific conclusions, manage the flow of data, and serve our diversifying science and engineering communities.

While science grows in complexity, requiring the integration of multiple approaches, a related trend is that the underlying computational technologies that enable this science are themselves experiencing unprecedented and disruptive changes. This implies that *there are no clear pathways for building the computational facilities that will be needed a decade from now,* which has profound and broad implications for science and for DOE programs going forward. A number of studies have emphasized different aspects of this point and its various implications. In particular, key findings of prior reports (Dongarra[1]/Yelick[2]/Giles[3]) relevant for our work include (a) business as usual will not yield the computing facilities required; (b) a more sustainable model is required to support personnel; (c) the computing technology roadmap is very unclear and major R&D programs will be needed to determine successful pathways for

[1] Can the United States Maintain Its Leadership in High-Performance Computing?

[2] Charting a Path in a Shifting Technical and Geopolitical Landscape: Post-Exascale Computing for the National Nuclear Security Administration

[3] https://science.osti.gov/-/media/ascr/ascac/pdf/reports/2013/ASCAC_facilities_statement_final.pdf

effective production computing facilities of the next decade; (d) this must involve a comprehensive approach of science and engineering domains, computer science teams, vendors, and facilities, to attain the critical mass required for success; and (e) the approach must include hardware, algorithmic, scientific, and software components. *The work of our subcommittee reaffirms and builds on these key points.*

In what follows, we address the specific charge we were given, explore these themes as motivation for our three recommendations, discuss opportunities for DOE and national science and engineering research, and describe the risks of not following the recommendations.

Section 1.1 The charge summary recommendations

The charge from Director Berhe is provided in full in Appendix 2, but in summary, it says:

- Consider what new or upgraded facilities will be necessary to position the SC at the forefront of scientific discovery, including a list of five specific ASCR facilities.

- The potential of each to contribute to world-leading science in the next decade. These should be considered in terms of the readiness for construction; the sufficiency of R&D performed to date to ensure the technical feasibility of the facility; the extent to which the cost to build and operate the facility is understood; and site infrastructure readiness. Please place each facility in one of three categories:
    (a) Ready to initiate construction.
    (b) Significant scientific/engineering challenges to resolve before construction.
    (c) Mission and technical requirements not yet fully defined.

Responding directly to the charge, we repeat **Recommendation 1**, summarized in Table 1 and detailed in Section 3:

> **Recommendation 1: Ensure the continued support and development of all five ASCR computational facilities reviewed—ALCF, OLCF, NERSC, HPDF, and ESnet—as they are central and essential to all SC science programs and broader national science and engineering research programs.** Each facility provides distinct and critical functionality that are essential to achieve SC science goals. Significant R&D investment is necessary to sustain their crucial roles. A summary of our findings can be found in Table 1. Recommendations on individual facilities are found in Section 3.

We regard Recommendation 1 as necessary but not sufficient for success. Additional Recommendations 2 and 3 are discussed and elucidated below. In a nutshell, they are that the five ASCR computational facilities should be viewed and further developed as an integrated **ecosystem**. As the *de facto* high-end national computing infrastructure, they together serve other parts of DOE and many other national research agencies. Finally, a comprehensive multi-agency R&D program will be needed, with vendors and industry, if we are to be successful in maintaining international leadership in science and technology.

Section 1.2 Science-driven imperative

Our recommendations are grounded in the urgent needs voiced by scientific communities, as evidenced by our discussions with other SC subcommittees. The urgency for new capabilities in research, "*...by yesterday…*" was frequently cited. This need is driven by a significant shift in how scientific advances increasingly depend on advanced experimental facilities and extreme computational power to handle surging data rates and theory probing. The fusion of experimental, theoretical, computational, and scientific disciplines is transforming traditional methods. Innovations such as real-time experiment control and interdisciplinary research that synthesizes geographically distributed data are becoming essential. This shift necessitates innovative workflows, AI/ML algorithms, and software tools. There is a critical need for substantial improvements in networking, as demonstrated by ESnet, and in capabilities for managing large data volumes, a challenge HPDF is well positioned to tackle. These advancements are crucial to handle the complexity and scale of data that modern science demands.

However, as scientific methods rapidly evolve, a noticeable expertise gap exists. Science communities may lack detailed knowledge of computational possibilities, while computational experts face emerging technological challenges without a clear roadmap for future systems. This gap highlights a disconnect between the aspirations of scientific inquiry and the realities of computational capabilities.

We are at a juncture where the methodologies in science, the technological landscape, and the expertise of both scientific and computing communities are evolving swiftly. It is vital that these groups work closely and urgently to delineate the pathways that will guide their progress.

Beyond SC, communities within NNSA and other agencies with related missions are facing increasing demands for computational power and data analysis capabilities, especially in critical areas like climate research, epidemiology, cancer research, and transportation safety. These needs highlight the importance of shared datasets, software tools, networks, and computational infrastructures. Additionally, the DOE's role in promoting an innovation economy is crucial. Investments in microelectronics, energy independence, and infrastructure are essential for national security and economic prosperity. It is vital to support these broad computational and data demands to maintain U.S. international competitiveness.

The pathway forward must be sculpted through an exhaustive research and development program that prioritizes integration. The future involves not just expanding existing facilities but conceptualizing a unified facility that merges multiple components, creating a cohesive environment for scientific exploration. This approach will support a versatile workforce skilled in AI/ML, modeling/simulation, data management, and HPC, among other areas. Such an integrated framework is crucial for tackling the complex challenges ahead and will serve as a foundation for our vision of an integrated research infrastructure, facilitating unprecedented collaboration and innovation in scientific discovery.

Section 2 Toward an Integrated Computing and Data Facility (the *Ecosystem*): Vision and R&D Pathways

We have responded to the subcommittee's charge regarding the construction readiness of five specific computational facilities (ALCF, OLCF, NERSC, HPDF, ESnet), with **Recommendation 1** that all are required for DOE SC mission science success. In our view, this must be accompanied by additional recommendations, motivated by the science discussion above:

> **Recommendation 2: Science demands integration. We advocate viewing ASCR facilities not as isolated entities, but as integral components of a single, larger integrated computational *Ecosystem*, with a single governance model.** This will require new ways of governing and potentially funding the overall ***Ecosystem***, which should not be developed via individual site procurements. Rather it should be designed, developed, built, and operated as an integrated facility ***Ecosystem*** for DOE science. It is critical for supporting SC science programs, along with additional software, algorithm, workforce, and science application components, to serve science and engineering research. *Further, this integrated **Ecosystem** is required for programs of other agencies, and industry. Its importance to the entire national scientific and technological capability, and its importance as a model internationally, cannot be overstated*.

Further, the ***Ecosystem*** is also central to serve the array of experimental facilities, previously laid out by ASCR as the Integrated Research Infrastructure (IRI), also described in Section 2.1. In addition, a comprehensive R&D program must be undertaken to develop both the future computing components and their integration (because the technologies, algorithms, and workflows require it).

> **Recommendation 3: A comprehensive, coordinated R&D program delivering multiple prototype computing systems over a five-year timescale must be mounted to inform pathways for this integrated ecosystem, operational by 2034, due to (a) rapidly evolving economic and technical landscapes of the semiconductor and computing industries and (b) changing research practices.** Despite varying readiness for construction, ASCR facilities collectively require a comprehensive R&D strategy for their future development. With the end of Moore's Law and with vendors focusing on other markets, the future of high-end computing for science is highly uncertain. R&D is needed to chart this course, to influence vendors, and to prepare applications for future platforms. The changing nature of interdisciplinary research requires a deeper integration of facilities, workflows, algorithms, software and application tools. *The R&D program should involve a collaborative effort across DOE computational facilities spanning SC. To ensure a comprehensive, multidisciplinary approach, it is critical to include contributions from computing and cloud vendors, computer science, DOE experimental facilities, and domain science and engineering communities, with deep collaborations with DOE NNSA and other federal agencies*. A new governance model will be required to manage this.

In this section, we further elucidate these additional conclusions.

Section 2.1 Envisioning the Integrated ASCR Facilities *Ecosystem*: A Unified Framework for Scientific Collaboration

The *Ecosystem*. Driven by needs of science, we believe the five ASCR facilities (ALCF, OLCF, NERSC, HPDF, and ESnet) should be viewed as, deeply integrated into, and operated as an integrated ASCR facilities ecosystem. This transformation of five component facilities into a single integrated *ecosystem* has begun, but substantial R&D will be required as it evolves over the next decade. For simplicity, we will refer to this group of facilities and their ongoing integration as the ***Ecosystem***.

It is important to recognize that the ***Ecosystem*** is, and must continue to be, the most advanced science-driven computational, data, and network facility on the planet. The ***Ecosystem*** must be dynamic and able to adapt to changing needs of SC (and properly funded). It serves—and is absolutely essential to—all science programs in SC, and well beyond. The ***Ecosystem*** has multiple user communities and partners and provides unique capabilities needed for their science. The ***Ecosystem***:

- Serves theoretical and computational science efforts across SC.
- Serves important needs of user communities of other U.S. research agencies. Even as these agencies provide some of their own computing and data facilities, ASCR's ***Ecosystem*** and its components provide unique capabilities supporting science that cannot be found elsewhere. It is essential to the entire national scientific enterprise.
- Serves U.S. industry, helping sustain global competitiveness for the nation. The industry needs for the ***Ecosystem*** will continue to grow over the coming decade.
- Serves and provides international leadership for the global scientific community. ASCR's global leadership role cannot be overlooked and must be sustained in the future.

These points all show how the ASCR ***Ecosystem*** stands on its own to support national and international scientific, industry, and technological leadership, but there is an additional and very specific way that the ASCR ecosystem and its further integration is required specifically for SC.

- The ***Ecosystem*** serves, and will be essential to, the array of DOE SC experimental facilities that are generating exponentially growing data volumes, in what is referred to as the Integrated Research Infrastructure (IRI), described below.

Future development of this ***Ecosystem*** will require a shift in how DOE and ASCR approach its computing facilities. Lessons can be learned from the Exascale Computing Program (ECP). ECP was enabled by a collaborative R&D program that spanned SC and NNSA laboratories, with vendor and research community partnerships. Collectively, these partnerships fueled the innovation needed to produce exascale computing hardware, software, and applications, *and also the needed workforce to achieve its goals* (a point that we feel must be strongly emphasized and adopted in our proposed R&D effort). ECP also enhanced the culture of collaboration across the SC and NNSA laboratories. We will explore this point further below.

We have stopped short of calling the ***Ecosystem*** itself a *named facility* although we considered doing so and urged SC to consider it. However, we wish to emphasize that the ***Ecosystem,*** *as we see it, will* require new active and intentional management and funding strategies that highlight its role in providing integrated and interconnected services to all SC programs and facilities, as well as to other U.S. agencies and even to international users and research partners. Although our subcommittee envisions the ***Ecosystem*** under ASCR, we strongly urge SC leadership to work with ASCR and other SC offices on how best to manage and fund the ***Ecosystem*** and, with NNSA and key agencies outside DOE, on how best to collaborate on use and R&D needed to further develop it. Considering the evolving nature of computing facilities, we suggest the formation of a higher-level coordination body that could include several lab directors and ensure that future generations are distinct and innovative, addressing risks of uniformity. The ***Ecosystem*** should be dynamic; if major new computing facilities are contemplated beyond the five we considered, we urge that their integral role be considered.

Integrated Research Infrastructure, or IRI. We see the ***Ecosystem*** as vital yet distinct from DOE SC's *Integrated Research Infrastructure* (IRI), which aims to further integrate ASCR's computational systems with SC's experimental facilities handling rapidly increasing data volumes. ASCR has already started collaborating with all five SC science programs (BER, BES, HEP, NP, FES) to develop the IRI, which will connect all 28 SC user facilities across DOE's national laboratories seamlessly.

The vision for IRI has evolved significantly since its initiation in FY21 with a collaborative taskforce from ALCF, OLCF, NERSC, and ESnet. By FY22, more than 170 experts from DOE national laboratories had joined to draft the IRI Architecture Blueprint Activity (ABA), laying the groundwork for a coordinated SC-wide strategy for this integrated research infrastructure. Notably, the High-Performance Data Facility (HPDF), spanning two sites, emerged as a crucial node in linking the ***Ecosystem*** and supporting all SC experimental and science programs.

Initial IRI planning has aligned with key aspects of our recommendations, particularly regarding testbed activities to find optimal scientific solutions and a common governance model for the ASCR Ecosystem's facility components. This planning confirms the perceived value of both the ***Ecosystem*** and IRI to the science programs, emphasizing the importance of involving all SC science as well as computer and computational science communities in future developments to maintain their support.

We concur, but we recommend that the proposed R&D program needed to develop this ecosystem should extend beyond SC, including other parts of DOE and other science agencies that rely heavily on these facilities. We do not believe SC alone has the critical mass to carry out all R&D needed for success, nor can it shoulder the responsibility alone to develop such an ***Ecosystem*** that serves so many other science and engineering research activities. Agencies outside SC should be brought into alignment, to achieve critical mass, including carrying out essential research workforce development, and to maintain international leadership. *While interagency R&D programs can be complex to manage, it is imperative that barriers to collaboration across the agencies be broken down and effective collaboration established, as nothing less than U.S. national scientific leadership and economic competitiveness are at stake.*

Appropriate and coordinated funding can help lower such barriers, via a coordinated national strategy.

Our analysis of the various facilities in Section 3 should be taken in the context of the integrated ***Ecosystem*** view of the components as part of the bigger whole (the whole is most definitely greater than the sum of its parts), as more and better interdisciplinary science and engineering research will be possible through such an integrated facility.

Section 2.2 Proposed comprehensive R&D program to inform final construction of the integrated facility on the decadal timescale

As other national studies have emphasized, the computing landscape, particularly in advanced computing, is undergoing profound change. This shift is driven by remarkable advancements in generative artificial intelligence (AI) and a complex array of economic, technical, and geopolitical factors. The "free lunch" of Dennard scaling and Moore's Law performance increases has ended. Semiconductor foundry costs are rising, the financial locus of computing innovation has shifted to AI and cloud hyperscalers, and U.S. dependence on chip manufacturer TSMC in Taiwan for leading-edge semiconductors poses increasing geopolitical risks.

Against this backdrop of change and response to these challenges, a new model of public–private partnerships is essential to design, develop, and deploy a next generation of advanced computing infrastructure that can effectively support DOE's differentiated missions. This will necessitate a shift from the current model of periodic vendor system procurements to one that leverages collaborative research and development initiatives to shape the future. These must span all of DOE, leveraging expertise across SC and NNSA while also fostering broad, whole-of-government partnerships with other federal agencies that have related missions and dependencies.

Simply put, this model requires a more in-depth and foundational R&D approach than previous Path Forward initiatives, establishing long-term collaborative partnerships with a variety of technology partners—including traditional hardware vendors, startups, and cloud hyperscalers—well before making any procurement decisions. The necessary collaborative R&D program should focus not only on hardware and software development but also integrate tools and techniques for handling complex, distributed workflows and enabling multidisciplinary discovery. These elements are crucial for the effective development and integration of leading-edge computing facilities such as ALCF, OLCF, NERSC, ESnet, and the High-Performance Data Facility (HPDF), as well as supporting the broader needs of SC experimental facilities.

This work may well include building substantial hardware–software prototypes to test ideas and help de-risk promising technology paths for component technology providers and product vendors. Only once the feasibility and integration of these hardware–software prototypes have been validated should the procurement process commence.

Furthermore, this R&D approach—crucial for developing technologies that meet the specific needs of DOE SC and scientific computing—will help attract and retain the necessary talent to keep DOE and the nation globally competitive and secure. *Workforce development is vital for* *sustaining these capabilities and propelling the ecosystem forward. Developing and retaining top talent within DOE and the broader U.S. scientific community is crucial for maintaining global competitiveness and security. This requires not only the creation of targeted technologies and systems but also the implementation of a workforce strategy that attracts, empowers, and retains skilled professionals (as emphasized here and in previous reports).*

To summarize, the proposed R&D partnership focuses on two main areas. First, it requires identifying and prototyping viable computing technologies before committing to significant future procurements. Second, it involves developing methods to integrate these technologies into a cohesive, multidisciplinary computing, data, and network infrastructure designed to support DOE's SC science programs and other agency programs reliant on ASCR facilities. Workforce development is a critical crosscut across both. This comprehensive approach will ensure the development and integration of individual facility components and their synergistic operation across multiple sites, creating a unified ecosystem that meets the research community's needs.

This all-inclusive, collaborative R&D strategy also demands a new governance model inspired by the Exascale Computing Project (ECP), which coordinated software and applications across all of DOE's SC (including but not limited to the five ASCR facilities covered in this charge) and NNSA laboratories. Extending this model to include hardware and finding common ground with technology developers—such as computing vendors, startups, and cloud hyperscalers—is crucial and will require discussions with vendors, given DOE's limited direct procurement influence. The R&D phase will need to precede and then run concurrently with initial procurements, ensuring that the technologies developed are both viable and well- integrated into the broader ecosystem.

Furthermore, this R&D strategy places a significant emphasis on involving "outside" participants, including universities and private sector researchers, from the early stages of planning. *DOE simply cannot "go it alone."* This inclusive strategy aims to harness diverse expertise and foster innovation by integrating academic research and industrial practicality into the framework's foundational stages. Universities will play a crucial role in pushing the boundaries of theoretical aspects and providing fresh research perspectives, while private sector entities will contribute practical insights and scalability solutions, enhancing the framework's applicability and robustness.

Balancing risk, reward, and investment within this new framework will require a finely tuned governance model that ensures all parties have clear expectations and defined roles. Risk-sharing mechanisms and investment models should be structured to incentivize participation while protecting stakeholders from undue exposure. This balance will be critical to maintaining engagement and driving the collective pursuit of technological advancements.

If DOE fails to take this approach of research and pathfinding for its scientific computing ecosystem, the risks to the country's computational capabilities and the advanced computing ecosystem are both large and potentially debilitating. These risks include, but are not limited to (a) loss of U.S. global leadership in advanced computing, (b) further destabilization of the computing hardware vendor ecosystem due to premature technology choices, (c) the inability to achieve DOE's science objectives, as well as collateral science effects at other agencies that depend on DOE, and (d) new generations of systems with even lower efficiency, with concomitant scientific, technical, and political risks. Given the severity of these risks, failure to adopt a long-term, integrated R&D program may lead to erosion or loss of program funding. Success in this approach should lay the foundations for success across all such critical areas.

Application software. For facilities to have the desired scientific impacts, they must be able to run advanced software suites that include modeling and simulation, data management and analysis, artificial intelligence - and all the libraries and tools that they depend upon. Much of this software is not owned by the facilities themselves, but rather by scientists and scientific communities. The ability to run these assorted software stacks efficiently is already a core element of procurement decisions. But beyond procurements, the facilities have a keen interest in ensuring the continued development and support for this broad range of capabilities.

Traditionally, research grants have paid for the development of new capabilities, but the Office of Science has not consistently supported maintenance and support for software. This challenge is acute at this point in time, with uncertainty about how to sustain and evolve the vast body of software developed under the Exascale Computing Project. It is worth considering a facilities-like model for supporting software with a finite-duration development phase followed by a sustained, lower investment in operation and sustainment. Whatever model is selected, the existing ASCR facilities are major stakeholders and need to play a central role in the conversation.

Section 3 Assessment of the need and readiness for construction of the individual facilities

In this section, we respond directly to the charge from Director Berhe for each of the five individual facilities, reviewing the vision and science-driven need, the role in the ASCR ecosystem, required R&D and prototyping needed to inform the pathways for development and deployment on a decadal timescale, and then, specifically (a) the potential for each component to contribute world-leading science in the next decade, as well as (b) the readiness for construction for each. As stressed throughout, we regard each of these as components of an overarching data and computing facility that serves all of SC and other agencies.

As there are a number of common threads among the 10 individual facilities, we organize our thoughts along groupings. The first is functional (Leadership Computing [LCF], NERSC, HPDF, and ESnet) and the second is by the "era" in which the facilities will be deployed. The first era are those facilities to be deployed in the next few years (ALCF-4, OLCF-6, NERSC-10, HPDF, and ESnet-7), while the second will be deployed after 2030 (ALCF-5, OSCF-7, NERSC-11, HPDF [no versioning yet!] and ESnet-8). This is because our assessment of scientific contribution is best described by the function of the facility, while there are many common aspects of the readiness assessment that apply broadly to either the earlier or the later deployment group.

As noted in Section 2.2, the shift in the computing landscape necessitates a novel approach to R&D, emphasizing active collaboration with both current and new vendors to explore experimental hardware systems and new software technologies. This partnership model aims to co-develop and test innovative computing solutions, ensuring they are ready for next-generation deployment. It underscores a transition from vendor-driven risk to a shared model of technological advancement, crucial to keep pace with rapid technological evolution and the expanding demands of scientific research. The changes include broadening the pool of vendors (to include some of the hyperscalers and startup companies), and investigating novel procurement models with which to engage the vendors.

The ”early” facilities are largely ”ready for construction,” but R&D is recommended as a part of the construction process. It is our feeling that these technologies and strategies will be insufficient for the ”later” facilities and the proposed research program is urgent and must deeply inform this later set. Therefore, we choose to rate these as needing to overcome ”science and engineering hurdles.“ The proposed research program is our view of what is needed to overcome these hurdles.

Section 3.1 Leadership-Class Computing Facilities

3.1.1 Vision and science-driven imperative

The Argonne and Oak Ridge Leadership-Class Computing Facilities (LCFs) provide SC researchers and partners with highly capable computing resources—often the fastest in the world. It is noted that ASCR has conventionally referred to the LCF as one facility with two sites. For purposes of our assessment, we will treat them as individual facilities. These platforms are used by many different programs and impact every SC office. They allow computational studies at length and time scales otherwise inaccessible, and enable a broad range of scientific discoveries and insights. These facilities are absolutely central to the success of science in the next decade.

3.1.2 Role in the ASCR *Ecosystem*

The LCFs allocate computing cycles via INCITE, a competitive process that is oversubscribed by a factor of 3–4. Significant allocations are provided to national laboratories, universities, and industry. The facilities also play a central role in sustaining U.S. preeminence in advanced computing amid fierce international rivalry.

Leadership-class platforms *uniquely* serve to assess the potential to computationally model problems at the frontier of science and engineering. It is often the nature of grand-challenge problems to require a degree of scale that will sufficiently encompass some minimal formulation of crucial phenomena (including resolution, feature size, physics, and fidelity). In effect, solutions are not valuable beneath some threshold of capability. Therefore, the top-tier facility will define state of the art in applying computational methods to barrier problems, discovering and validating problem resolution for needed solution confidence, and consequently guiding requisite scale of system procurements.

The two LCFs support not only two leading-edge machines but also various smaller platforms as well as technical expertise in advanced computing. This approach ensures a wide-ranging perspective on future computing trends, and also includes exploratory areas such as quantum computing and AI. At a time of enormous uncertainty in the future of scientific computing, this internal SC expertise is critical. In addition, the LCFs also serve as regional hubs for industry collaboration and talent attraction, bolstering local economic and technological development. *It is a strong conclusion of the subcommittee that two LCF facilities serve to anchor multiple technology pathways, build and sustain critical workforce, and enhance the U.S. economy in ways that one facility simply cannot.*

3.1.3 Required R&D

As discussed above, the LCF facilities will require R&D to be successful, especially towards OLCF-7 and ALCF-5. This is covered elsewhere and is not repeated here.

3.1.4 Site-specific recommendations

For the ALCF and OLCF facilities, an emphasis on workforce development and training is crucial. Both centers are actively engaged in extensive training efforts, which include webinars, onsite programs, and workshops designed to enhance user expertise in navigating the complexities of modern computational systems and software. At ALCF, the Argonne Training Program on Extreme-Scale Computing (ATPESC) exemplifies this commitment by providing hands-on training to researchers on the skills necessary for leveraging the full capabilities of leading-edge computational resources. Similarly, OLCF continues to expand its educational outreach through a variety of training sessions that focus on practical applications, system optimization, and emerging technological trends. Both facilities are dedicated to fostering a skilled workforce that is well-prepared to meet the challenges of an evolving scientific landscape. This approach not only enhances the scientific productivity and technological proficiency of the community but also ensures that both ALCF and OLCF remain at the forefront of computational science and engineering innovation.

3.1.5 Assessment of potential to contribute to science and readiness for construction

OLCF-6, ALCF-4, OLCF-7, and ALCF-5 have very high potential to contribute to a broad range of fundamental and applied scientific challenges. OLCF-6 and ALCF-4 are in the early phases of the CD process, with anticipated completion in the later 2020s. Objectives for these machines include

- 5x-10x improvement in applications performance over existing systems.
- Improved energy efficiency to address the growing environmental concerns and operational costs associated with high-performance computing.
- Integration of AI, modeling and simulation, and data-intensive capabilities in a single resource.
- Support for DOE's plans for an Integrated Research Infrastructure.

OLCF-7 and ALCF-5 are scheduled for completion in the early to mid-2030s. Given the rapid flux in computing technologies, their required capabilities and the associated scientific opportunities are much harder to predict, though the need for computing resources is certain to be very high.

Oak Ridge and Argonne have enhanced their campus infrastructure and conducted a requirements assessment for OLCF-6 and ALCF-4. These procurements are largely ready for construction using existing and foreseeable technologies, albeit with continuing R&D. Conversely, later procurements will need to be informed by a thoughtful, robust, and collaborative R&D program, given both the uncertainties in the computing landscape discussed above, and in applications needs.

Section 3.2 NERSC

3.2.1 Vision and science-driven imperative

NERSC envisions a transition from its current emphasis on modeling and simulation, AI training and inference, and data analytics to become a workflow-driven facility. This transformation aims to facilitate seamless workflows within the Integrated Research Infrastructure (IRI), enabling pervasive AI utilization and leveraging quantum and other beyond-Moore technologies. As the SC mission center, delivering capabilities ranging from modeling and simulation through time sensitive workflows, NERSC is mentioned by the other subcommittees as critical for their respective science progress. NERSC seeks to balance leveraging technological disruptions with ensuring the user community's engagement. Deploying a system that caters to the user community's needs is crucial to maintaining scientific progress.

3.2.2 Role in the ASCR *Ecosystem*

NERSC serves as the SC mission computing center, playing an indispensable role in executing scientific workflows that support various application programs across the office. Among ASCR user facilities, NERSC stands out in supporting a diverse range of DOE open science computational workflows, benefiting approximately 10,000 scientists engaged in 1,000 projects per year.

Through ESnet, NERSC maintains tight integration with local and distributed SC user facilities, providing real-time analytics and data management capabilities. Success at NERSC is attributed to deep partnerships with the user community, fostering workforce development, technological evolution, and tool production for broader dissemination.

3.2.3 Required R&D

As with the leadership facilities, the roadmap for NERSC provides opportunities to leverage the outcomes of a coordinated R&D program for the longer time horizon facility refreshes (NERSC-11), whereas the next facility project (NERSC-10) is more evolutionary and can proceed to construction using primarily facility-specific R&D activities. These necessary activities are primarily centered on creating workflow system components, APIs, and federated IDs that anchor NERSC’s integration with the ***Ecosystem***. NERSC is committed to advancing its infrastructure to adeptly deploy and leverage emerging architectures, ensuring they are used to their fullest potential.

3.2.4 Site-specific recommendations

NERSC provides three principal services: 1) facilitating large-scale applications in simulation, modeling, and data analysis; 2) enabling complex workflows for experiments and AI applications; and 3) offering computing for time-critical and interactive operations. About 80% of NERSC's computational resources are committed to supporting the SC program areas' missions directly.

A key aspect of NERSC's approach involves integrating workload analysis and targeted workforce development, striving for a synergy between technological progress and scientific output. Reflecting this balance, NERSC's guiding principle is that "Sociology is as important as technology," emphasizing the equal importance of community dynamics and technological innovation in its operations.

NERSC's system strategy is integral to this framework, focusing on cultivating deep relationships with vendors and the scientific community, as illustrated by the NERSC Science Acceleration Program (NESAP) and collaborative efforts with vendors to stay abreast of technological developments. This engagement provides a linkage between emerging trends from the vendor community and workforce development opportunities across the broad Office of Science community that uses the NERSC facility. It also serves as a prototype for our envisioned ***Ecosystem***. The strategy includes a move toward high-density, liquid-cooled systems, which is in response to the center's evolving requirements, particularly in AI training and inference tasks, demonstrating a commitment to aligning their infrastructure with the latest technological trends and shared objectives.

3.2.5 Assessment of potential to contribute to science and readiness for construction

NERSC's role in advancing world-leading science over the next decade is absolutely central, with its contributions recognized as pivotal to the efforts funded and anticipated science programs across various SC offices. Each SC office involved plays a critical part in shaping NERSC's strategic direction, ensuring its alignment with the forefront of scientific research. The NERSC-10 mission (CD-0), rooted firmly in current mission requirements, is designed not only to meet immediate needs but also to anticipate and enable future scientific endeavors. This forward-looking approach ensures that the DOE and its computational users remain at the cutting edge of research, in sync with industry market dynamics and technological advancements while also providing a timely upgrade as NERSC-10 reaches its operational conclusion.

Regarding infrastructure readiness, NERSC-10 is scheduled for 2026 deployment. This exascale system is set to enhance capabilities in simulation and modeling, AI training and inference, and experimental data analysis. The project is on track, ready to begin construction according to its schedule. Looking ahead, NERSC-11, expected around 2030, aims to build on the achievements of NERSC-10. It will expand the center's connectivity within the DOE and to a broader emerging technological ecosystem, encompassing initiatives like IRI/HPDF, pervasive AI, and advanced computing paradigms. Currently in its preliminary planning phase, NERSC-11 is strategically positioned to leverage the technological strides and insights gained from NERSC-10 and current R&D (e.g., similar to PathForward/FastForward programs) for continued leadership in computational science.

Section 3.3 High Performance Data Facility (HPDF)

3.3.1 Vision and science-driven imperative

HPDF is a first-of-its-kind facility in DOE SC that will support current and future DOE SC data-intensive applications. The need for a data-focused facility is critical to the SC mission and a cornerstone of the IRI vision. HPDF’s mission is to enable and accelerate scientific discovery by delivering state-of-the-art data management infrastructure, capabilities, and tools. It thus fills a critical need for a data storage and management solution that will greatly enhance the support for DOE science. HPDF is envisioned to be at the heart of IRI providing a stable high-performance data management solution. The overall structure of HPDF is built on a hub and spoke model. The distributed facility consists of a hub hosting centralized resources and services while enabling high-priority DOE mission applications at multiple spokes supported through the orchestrated distributed infrastructure.

3.3.2 Role in the ASCR *Ecosystem*

HPDF fills a critical need in the ecosystem to be able to receive and store data from instruments and simulations flowing from SC user facilities. The HPDF Hub and associated spokes will become the leadership facilities for stewardship of the scientific data life cycle, advancing DOE’s commitment to findable, accessible, interoperable, and reusable (FAIR) data principles. HPDF will include data services, software tools, and technologies to support the full data life cycle, including capture, staging, processing, management, and archiving.

3.3.3 Required R&D

The HPDF Hub will be physically located at JLab and LBNL interconnected to an extensible network of spokes via ESnet's high-performance networks.

HPDF will support three computational patterns identified in the IRI Blueprint.[4] In particular, HPDF aims to support time-critical patterns enabling novel use of DOE SC user facilities by providing timely feedback during experimentation, ingesting experimental and observational data and enabling further real-time or near-real time processing. In addition to time-critical patterns, HPDF will also address data-integration-intensive and long-term campaign patterns by providing data wrangling, storage, curation, and long-term archival services.

Unlike storage capabilities provided by NERSC and the LCFs, HPDF addresses the needs of DOE applications and will need to provide novel solutions to support the *entire data lifecycle management* that is integrated with other ASCR user facilities. In particular HPDF plans to provide:

- Dynamic and scalable data management.
- Data capture ready to receive streaming data.
- Dynamically allocatable data storage and edge computing at the point of generation.
- Dynamic placement of data close to computing to enable efficient and time-critical processing.
- Archiving of data in accordance with FAIR principles.

Providing long-term data storage will require not only technical solutions but also data governance models and data storage policies that help decide which data needs to be preserved and for how long.

As with other ASCR user facilities, HPDF will provide critical expertise and hands-on help to researchers to enable them to leverage the resources in an effective way and enable the sharing and long-term archival of DOE-funded results in support of open science. Distributing the Hub between two geographically separated locations will provide a high-performance and resilient system with built-in disaster recovery architecture.

In its initial stages, HPDF is poised to become an essential part of the ASCR infrastructure, a world-leading facility setting the example for addressing the explosive growth in data across all domains. HPDF holds a great promise of democratizing data access and fostering collaborations across science disciplines, enabling seamless data sharing of well-curated artifacts. With its

[4] https://doi.org/10.2172/1984466

user-centered, community-specific spokes it can also potentially democratize science, enabling researchers with all levels of experience to access data and associated artifacts.

3.3.4 Site-specific recommendations

Although the high-level goals and approach of HPDF are well defined, the site selection for HPDF occurred only in October 2023 and the facilities design needs to be refined to reach the level of Critical Decision 1 (CD-1). As part of this process, HPDF needs to gather user requirements through community workshops and/or review of existing documents gathered by DOE SC. Based on these requirements HPDF needs to architect the system that meets the requirements, provides support for the three IRI patterns, and supports the experimental and observational data streaming paradigm, which is challenging to realize today. It involves real-time to near-real-time data calibration and processing, fast data transfer from instrument to storage, and dissemination among others.

The HPDF team has synthesized key design principles of experiment-friendly availability and data-driven agility. From these principles, they have defined a set of key architectural components that will anchor the system design and a set of well thought-out services that will be provided by these components. They have also developed a process for selecting the hardware and software needed to support the HPDF design.

HPDF is making excellent progress in the design phase of the facility and is on a good path to be ready for construction once CD-3 is reached.

The hub and spoke design of HPDF allows us to evaluate the readiness for the construction of the individual components. Being a central component on which spokes depend, the Hub is further along in the design phase.

3.3.5 Assessment of potential to contribute to science and readiness for construction

Given its importance to DOE science, the subcommittee places HPDF as absolutely central. With the increases in data volumes and rates from both the computing and experimental facilities, the changing nature of interdisciplinary science and scientific workflows, and advances in AI, the HPDF is absolutely central to the vision of the **Ecosystem,** and we do not see how the science goals of SC and the broader community can be achieved without HPDF.

The subcommittee assessed the readiness for construction of the HPDF Hub to be (b) significant scientific/engineering challenges to resolve before initiating construction.

HPDF spokes will provide customizable and dedicated access to HPDF capabilities for high-performance data management. Because of their ability to be customized to community needs the spokes can grow and evolve as science requirements evolve, new instruments come online, and new communities form and dissolve.

The subcommittee assessed the readiness for construction of HPDF Spokes 1 and 2 to be (c) mission and technical requirements not yet fully defined.

Section 3.4 ESnet

3.4.1 Vision and science-driven imperative

Initially, networking was viewed as conduit for data transfer, but it has evolved over the decades into a dynamic array of services that are now central to the DOE's scientific enterprise, particularly within the integrated facility framework of HPDF and IRI. Throughout, ESnet has been at the forefront, not just in envisioning this future but also in working closely with diverse scientific user communities across all SC segments to understand research directions and predict future data movement needs. ESnet has been instrumental in prototyping and then implementing novel, science-driven services and techniques as production services, and is now absolutely essential to the integrated ***Ecosystem***.

Modern instruments at DOE science facilities are producing data at unprecedented and exponentially increasing rates. This data must be dynamically connected to distributed computing facilities, potentially accessing data in real-time from global locations for scientific analysis, distributed user access, and dynamic instrument control during experiments. Moreover, data from various instruments, archived and disseminated globally, requires processing with increasingly sophisticated and demanding algorithms for scientific inference.

As a leading research and service organization, ESnet continuously assesses evolving scientific needs across ASCR and all SC programs. It has developed roadmaps to address the challenges posed by rapidly expanding data streams and complex workflows involving multiple instruments and distributed computing systems. ESnet's historical approach includes extensive community engagement through its requirements review program and a co-design process that has been foundational in developing and deploying network services from ESnet-1 through ESnet-5 over approximately 35 years. The requirements review process is anchored in understanding science needs, workflows, and usage patterns that the ESnet team translates into the quantitative requirements to allow data to flow freely through the DOE complex. This process has produced tools and services, like perfSONAR and the Science DMZ model, which have become standards adopted by universities worldwide. With ESnet-6 now operational and connecting all current DOE labs and user facilities—handling nearly two exabytes of data last year—ESnet has proven its capability, expertise, and innovative spirit to support the demanding scientific landscape anticipated in the next decade.

3.4.2 Role in ASCR *Ecosystem*

ESnet is crucial to the DOE SC science programs, evidenced by its extensive engagement in preparing for future challenges. In its latest three-year requirements gathering cycle, ESnet conducted 88 case studies involving over 450 participants from all five SC science programs. These studies provide deep insights into current and future networking needs as science disciplines evolve and next-generation instruments come online.

Currently, ESnet is well-connected to activities and facilities in all SC science programs, as demonstrated by dataflows between various facilities across all SC programs. These dataflows will increase dramatically in the future, as significant upgrades in experimental instruments are leading to exponentially higher data generation rates. Examples include the LHC upgrades resulting in tenfold increases in data rates, and the forthcoming Vera Rubin Observatory expected to produce vast amounts of data with its high-capacity imaging capabilities. Similarly, upgrades in BES light sources and new capabilities in NP and BER for handling complex data streams necessitate robust network solutions. In FES, real-time control of fusion experiments across continents highlights the need for dynamic and high-speed networking.

These scenarios underscore the essential role of ESnet in supporting integrated and interdisciplinary scientific research, characterized by rapid data analysis, novel workflows, and AI-optimized processes across geographically dispersed locations. The use case scenarios for these examples will require significant shifts in the way science is carried out, with several common characteristics for new modes of integrated and interdisciplinary science: (a) rapid (and real-time) data analysis and steering of experiments, where the computational and experimental facilities, and their users are in widely distributed geographic locations; (b) novel workflows using multiple user facilities; and (c) AI-enabled and optimized workflows, and drawing scientific inferences from widely distributed, huge data sources, requiring dynamic and very high speed networking connections. Many of these will require distributed computing scenarios that will simply be impossible without advanced and highly reliable networking services. ESnet is clearly on the critical path for most DOE mission science objectives across all science programs, that will not be possible if ESnet is unable to fulfill its networking objectives.

3.4.3 Site-specific recommendations

The vision for ESnet includes significant system upgrades to meet the demands of the HPDF and IRI frameworks. ESnet-6 is currently laying the groundwork for these advancements. ESnet-6 will be the underlying networking framework for prototyping IRI pathways that will be needed in exploring new workflows of the types described above. We stress that while ever-increasing bandwidth and automated network orchestration will need to be in place to accommodate raw data rates and volumes as a baseline requirement, the dynamic services and AI-based queries of multiple distributed datasets will provide much greater challenges.

ESnet-7 is envisioned for the 2027+ timeframe, which expects to deploy much advanced net and data services, which will require advances in AI techniques and significant wireless services needed by BER, for example. Building on pathfinding work for IRI, it will need to provide more nimble, and intelligent network functions and data-centric services to serve evolving science scenarios. Further, AI-controlled network management services for dynamically defined network configurations will be required for the degree of automation foreseen for experiments controlled by computation to rapidly cycle through parameter studies, for example.

ESnet-8, foreseen in the 2032+ timeframe, will continue these advances and see an extensive infrastructure overhaul, and will also need to deal with optical fiber IRU lease renewals. This in particular needs to be taken into account, and pricing for lease renewals and possibly new fiber is uncertain. Further development and advanced deployment of AI technologies to control networks for IT operations will be needed.

Thus, ESnet's role is central now as a force multiplier for computing and experimental facility investments and will become increasingly critical as science and technology evolve, demanding more sophisticated and integrated computational and data infrastructures.

3.4.4 Required R&D

The networking technologies needed to support the integrated scientific computing facility are sufficiently well defined that we can say that ESnet is both essential to the vision of DOE science, and that we have a high degree of confidence that the future networking capabilities required can be successfully developed and deployed.

However, the profound step change in the data generation rates and volumes produced by next-generation experimental and computing facilities, and dramatic shifts in the modalities of research, mean that a significant R&D program will be needed to develop both the dynamic networking technologies and the science and engineering research applications. While the uncertainties in the basic future network technologies needed for SC mission science objectives are less pronounced and carry less risk than for the computing systems described above, a significant R&D program still will be required for IRI, HPDF, and the vision of science and engineering research to be successful. Hence, ESnet must be an integral partner in the proposed R&D program described in this document.

In terms of ESnet services and R&D in the context of supporting research beyond the DOE labs, including involvement with other agencies such as NSF and at the campus where many of the users live and work, ESnet has a very strong track record of working with such partners.

3.4.5 Assessment of potential to contribute to science and readiness for construction

The subcommittee assessed ESnet and its planned upgrades to ESnet-7 and ESnet-8, and found them to be ***absolutely central*** not only to realizing the vision of the integrated data and computing facility discussed here, but also to delivering on DOE SC science mission goals.

In terms of readiness for construction of the network capabilities needed for the integrated facility on a decadal time scale, **the subcommittee places ESnet-7 as well enough along in its development to place it in category (a) ready to initiate construction**.

Our subcommittee places ESnet-8 in category (b), having significant science/engineering challenges to resolve before initiating construction.

However, as laid out above, the challenges for the networking components are less severe than those for the computing components, and we are highly confident that these challenges can be resolved through the proposed R&D and pathfinding/prototyping programs.

Section 4 Summary

Our subcommittee has made a thorough study of both current ASCR computational facilities (ALCF, OLCF, NERSC, HPDF, and ESnet) and plans for them going forward toward 2034. Our subcommittee is unanimous in the following findings, recommendations, and the risks of not following them.

We found that ASCR and Office of Science (SC) programs are internationally leading in both the capabilities of the computational and networking facilities and the breadth of the science programs that depend vitally on them, both across DOE and other agencies. Further, we have looked deeply into trends in science and engineering communities across all SC programs, BER, BES, HEPAP, NP, FES, and ASCR itself, as well as across other agencies. Deep changes in research are anticipated, due to the increasing integration of research disciplines needed to address leading complex problems, the profoundly changing research practices, and the introduction of new experimental facilities, with exponentially growing data streams that will be deployed on this timescale.

Building on prior reports (Dongarra[1]/Yelick[2]/Giles[3]), it is clear that with computing technology roadmaps themselves at a crossroads, the current procurement process for high-end computing facilities will not yield the needed computational facilities either to serve SC science goals or to maintain national science and technology leadership. Changes must be made to the approach, which includes both a deeper integration of ASCR computational facilities into a single coherent ***Ecosystem*** and a comprehensive R&D program, spanning all of SC, NNSA, and a significant number of other government agencies that also depend critically on these ASCR facilities for their work. *A coordinated all-of-government approach is needed for success*.

Given the rapidly changing landscape of interdisciplinary research, international competition, and the highly uncertain and disruptive technological future of technical computing, our committee is unanimous in saying that *to maintain scientific leadership, "business as usual" cannot continue; a new approach is needed* to address these issues, and also to provide a more sustainable and collaborative approach to workforce development and retention.

The work of our subcommittee results in the following three recommendations:

> **Recommendation 1: Ensure the continued support and development of all five ASCR computational facilities reviewed—ALCF, OLCF, NERSC, HPDF, and ESnet—as they are central and essential to all SC science programs and broader national science and engineering research programs.** Each facility provides distinct and critical functionality that are essential to achieve SC science goals. Significant R&D investment is necessary to sustain their crucial roles. **A summary of our findings can be found in Table 1. Recommendations on individual facilities are found in Section 3.**
>
> **Recommendation 2: Science demands integration. We advocate viewing ASCR facilities not as isolated entities, but as integral components of a single, larger integrated computational *Ecosystem*, with a single governance model.** This will require new ways of governing and potentially funding the overall ***Ecosystem***, which should not be developed via individual site procurements. Rather, it should be designed, developed, built, and operated as an integrated facility ***Ecosystem*** for DOE science. It is critical for supporting SC science programs, along with additional software, algorithm, workforce, and science application components, to serve science and engineering research. *Further, this integrated **Ecosystem** is required for programs of other agencies*

and industry. Its importance to the entire national scientific and technological capability, and its importance as a model internationally, cannot be overstated.

Recommendation 3: A comprehensive, coordinated R&D program delivering multiple prototype computing systems over a five-year timescale must be mounted to inform pathways for this integrated ecosystem, operational by 2034, due to (a) rapidly evolving economic and technical landscapes of the semiconductor and computing industries and (b) changing research practices. Despite varying readiness for construction, ASCR facilities collectively require a comprehensive R&D strategy for their future development. With the end of Moore's Law and with vendors focusing on other markets, the future of high-end computing for science is highly uncertain. R&D is needed to chart this course, influence vendors, and prepare applications for future platforms. The changing nature of interdisciplinary research requires a deeper integration of facilities, workflows, algorithms, software, and application tools. *The R&D program should involve a collaborative effort across DOE computational facilities spanning SC. To ensure a comprehensive, multidisciplinary approach, it is critical to include contributions from computing and cloud vendors, computer science, DOE experimental facilities, domain science and engineering communities, with deep collaborations with DOE NNSA and other federal agencies*. A new governance model will be required to manage this.

Recommendation 1 is necessary but not sufficient for success. Recommendations 2 and 3 are also required if we are to serve SC and broader national science and engineering research programs, and to remain in a global science and technology leadership position with very strong international competition increasing year after year. Further, we regard this as critical for continued economic leadership as well as for national security.

Failure to follow these strategies risks loss of international leadership in science/engineering research and in advanced computing, with long-term implications on U.S. national security, technological leadership, and economic competitiveness. Our recommendations offer opportunities for positioning SC to achieve science goals, enhancing beneficial partnerships across U.S. and select international science agencies, and sustaining U.S. international leadership in key areas.

Appendix 1 Charge from Office of Science Director Berhe

To: CHAIRS OF THE OFFICE OF SCIENCE FEDERAL ADVISORY COMMITTEES:

Advanced Scientific Computing Advisory Committee
Basic Energy Sciences Advisory Committee
Biological and Environmental Research Advisory Committee
Fusion Energy Sciences Advisory Committee
High Energy Physics Advisory Panel
Nuclear Science Advisory Committee

The Department of Energy's Office of Science (SC) has envisioned, designed, constructed, and operated many of the premiere scientific research facilities in the world. More than 38,000 researchers from universities, other government agencies, and private industry use SC User Facilities each year—and this number continues to grow.

Stewarding these facilities for the benefit of science is at the core of our mission and is part of our unique contribution to our Nation's scientific strength. It is important that we continue to do what we do best: build facilities that create institutional capacity for strengthening multidisciplinary science, provide world class research tools that attract the best minds, create new capabilities for exploring the frontiers of the natural and physical sciences, and stimulate scientific discovery through computer simulation of complex systems.

To this end, I am asking the SC advisory committees to look toward the scientific horizon and identify what new or upgraded facilities will best serve our needs in the next ten years (2024-2034). More specifically, I am charging each advisory committee to establish a subcommittee to:

1. Consider what new or upgraded facilities in your disciplines will be necessary to position the Office of Science at the forefront of scientific discovery. The Office of Science Associate Directors have prepared a list of proposed projects that could contribute to world-leading science in their respective programs in the next ten years. The Designated Federal Officer (DFO) will transmit this material to their respective advisory committee chairs. The subcommittee may revise the list in consultation with their DFO and Committee Chair. If you wish to add projects, please consider only those that require a minimum investment of $100 million. In its deliberations, the subcommittee should reference relevant strategic planning documents and decadal studies.
2. Deliver a short letter report that discusses each of these facilities in terms of the two criteria below and provide a short justification for the categorization, but do not rank order them:
    a. The potential to contribute to world-leading science in the next decade. For each proposed facility/upgrade consider, for example, the extent to which it would answer the most important scientific questions; whether there are other ways or other facilities that would be able to answer these questions; whether the facility would contribute to many or few areas of research and especially

whether the facility will address needs of the broad community of users including those whose research is supported by other Federal agencies; whether construction of the facility will create new synergies within a field or among fields of research; and what level of demand exists within the (sometimes many) scientific communities that use the facility. Please place each facility or upgrade in one of four categories: (a) absolutely central; (b) important; (c) lower priority; or (d) don't know enough yet.

b. The readiness for construction. For proposed facilities and major upgrades, please consider, for example, whether the concept of the facility has been formally studied; the level of confidence that the technical challenges involved in building the facility can be met; the sufficiency of R&D performed to date to assure technical feasibility of the facility; the extent to which the cost to build and operate the facility is understood; and site infrastructure readiness. Please place each facility in one of three categories: (a) ready to initiate construction; (b) significant scientific/engineering challenges to resolve before initiating construction; or (c) mission and technical requirements not yet fully defined.

Many additional criteria, such as expected funding levels, are important when considering a possible portfolio of future facilities, however, for this assessment I ask that you focus your report on the two criteria discussed above.

I look forward to hearing your findings and thank you for your help with this important task.
I appreciate receiving your final report by May 2024.

Sincerely,

Asmeret Asefaw Berhe
Director, Office of Science

Appendix 2 Our Process

Our subcommittee followed a rigorous process of information gathering, discussions, interviews with all key computing laboratories reported on below, interviews with all other science program subcommittee chairs, and consultation with key stakeholders including ASCR, ASCAC, other agencies, OSTP, and others.

- Our Process
    - January–February: Subcommittee holds virtual meetings to assemble, develop strategy to respond to charge, and provide guidance to the five labs.
    - February 23: Subcommittee co-chairs met with Jack Dongarra, author of prior ASCR report, to gather additional context.
    - February 29 and March 1: Subcommittee conducts virtual interviews with leadership from the five labs.
    - March 7: Subcommittee meets in Washington D.C. to debrief interviews and begin drafting report.
    - March 27 and 29: Subcommittee meets with other Office of Science subcommittees to discuss commonalities in charges and findings.
    - March–April: Committee refines report.
    - April 19: Subcommittee meets at Argonne National Laboratory to continue refinement of report.
    - April 20–May 1: ASCAC member Roscoe Giles provides preliminary review and feedback.
    - April 20–May 17: Subcommittee finalizes report.
    - May 22: Final report delivered to ASCAC.
    - May 29: ASCAC met in Washington, D.C. and approved the report.
    - May 30 - June 3: minor changes made based on feedback from ASCAC.
    - June 3: Report delivered to Office of Science.