Toward a Seamless Integration of Computing, Experimental, and Observational Science Facilities: A Blueprint to Accelerate Discovery

About the ASCR Integrated Research Infrastructure Task Force

There is growing, broad recognition that integration of computational, data management, and experimental research infrastructure holds enormous potential to facilitate research and accelerate discovery.[1] The complexity of data-intensive scientific research—whether modeling/simulation or experimental/observational—poses scientific opportunities and resource challenges to the research community writ large.

Within the Department of Energy's Office of Science (SC), the Office of Advanced Scientific Computing Research (ASCR) will play a major role in defining the SC vision and strategy for integrated computational and data research infrastructure. The ASCR Facilities provide essential high end computing, high performance networking, and data management capabilities to advance the SC mission and broader Departmental and national research objectives. Today the ASCR Facilities are already working with other SC stakeholders to explore novel approaches to complex, data-intensive research workflows, leveraging ASCR-supported research and other investments. In February 2020, ASCR established the Integrated Research Infrastructure Task Force[2] as a forum for discussion and exploration, with specific focus on the operational opportunities, risks, and challenges that integration poses. In light of the global COVID-19 pandemic, the Task Force conducted its work asynchronously from April through December 2020, meeting via televideo for one hour every other week. The Director of the ASCR Facilities Division facilitated the Task Force, in coordination with the ASCR Facility Directors.

The work of the Task Force began with these questions: Can the group arrive at a shared vision for integrated research infrastructure? If so, what are the core principles that would maximize scientific productivity and optimize infrastructure operations? This paper represents the Task Force's initial answers to these questions and their thoughts on a strategy for world-leading integration capabilities that accelerate discovery across a wide range of science use cases.

[1] See "Pioneering the Future Advanced Computing Ecosystem: A Strategic Plan," a report by the National Science and Technology Council, published November, 2020; "AI for Science," a technical report published February 2020. See also the European Union's European Open Science Cloud initiative and the China Science and Technology Cloud initiative.
[2] The Task Force members are listed in Appendix 1.

Executive Summary

The Department of Energy, Office of Science operates world-leading facilities for experimental, observational, and computational science. DOE supercomputing facilities will reach performance at the scale of ExaFLOPs in the coming years, enabling new vistas of scale and precision for large scale simulations and data analysis. Experimental scientific facilities are undergoing similar upgrades that will lead to higher data rates and correspondingly larger computational demands, and will increase the need for near-real-time processing and resilient support for more complex workflows. A transformation of science is underway, with workloads at supercomputing facilities increasingly driven by this explosion of data from instruments and experimental facilities, as well as the accelerating use of Artificial Intelligence (AI) as a tool for scientific discovery.

A seamless integration of computing, networking, instruments, and experimental facilities is required to support these emerging workloads and open up a new frontier of U.S. leadership in scientific discovery. We propose to accomplish this by providing frictionless access to the ASCR supercomputing facilities. We describe our vision of combining the power of ASCR supercomputers and networking infrastructure into an integrated scalable fabric, available to end user scientists via interfaces that aim to automate and simplify access to high performance computing systems. This will enable unprecedented computational science capabilities for experimental and observational facilities, and will create new opportunities to combine large simulations and modeling with experimental facility data analysis. This blueprint for creating an integrated network of computational and experimental facilities will provide an enriched discovery environment and open doors for new scientific communities to access the DOE's world-leading computing and networking capabilities.

Vision

Our vision is to integrate across scientific facilities to accelerate scientific discovery through productive data management and analysis, via the delivery of pervasive, composable, and easily usable computational and data services. We illustrate this vision below from the perspective of a BES x-ray lightsource user, describing that user's experience and the impact on science of a seamless and integrated computing environment.

We contrast this vision with scientists' experiences today, where a disjoint set of capabilities is available at each facility and where experimental facility workflows cannot readily take advantage of the unique resources at ASCR facilities to accelerate their science without creating laborious custom solutions. The increment we envision between today's scenario and an integrated future is described visually in Figure 1, where we have identified eight interrelated areas of interaction between experimental and observational facilities with ASCR high performance computing (HPC) and high performance networking (HPN) facilities. Requirements that drive the eight areas of interaction are detailed in Appendix 3 (Collected Requirements from Exascale Crosscut Report).

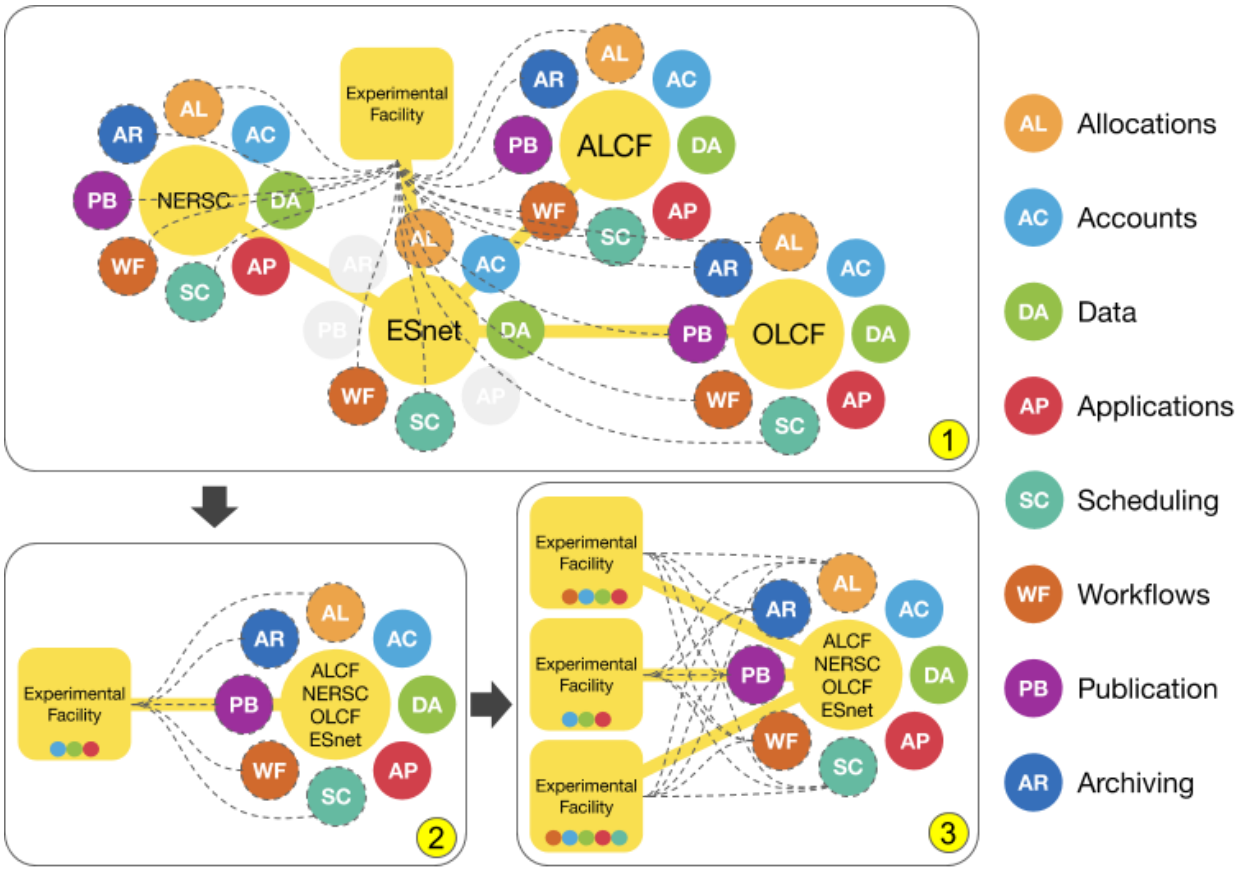

Figure 1. *Depiction of the integration of experimental facilities with computational facilities, across the range of services provided, in contrast with the one-to-one approach required today. 1. Today, an experimental facility must arrange separate bespoke interactions with individual HPC/HPN facilities. 2. A future paradigm with common interfaces could simplify integration of an experimental facility with multiple HPC/HPN facilities. 3. In turn, these common interfaces could support expansion and integration across multiple experimental facilities and HPC/HPN facilities.*

A future vision of a scientist's journey

Imagine a team of scientists are allocated time on a beamline at a premier x-ray light source facility within the DOE SC complex. The team's project, the accurate characterization of a biochemical process of national importance, is ambitious, and their allocated time on the beamline is limited. Accordingly, they plan to use near-real-time computational analysis and Exascale simulations to provide on-the-fly feedback to monitor the quality of their data and make adjustments and changes to their experiment as needed.

During beam time, the raw data from the experiment is filtered by computing resources local to the beamline, and the filtered data is sent via ESnet to ASCR HPC facilities for large-scale near real-time analysis. The end-user scientists performing the analysis use their home institution's credentials to access both the beamline and HPC facilities where appropriate accounts and data repositories are already available. The computing at the HPC facilities and the guaranteed network service are managed through

appropriately-authenticated application programming interface (API) calls made by the beamline software environment, and charged to an allocation agreed to by the involved facilities and their DOE programs. The beamline control software can move the workload between facilities as needed (e.g., because of maintenance, outage, or a fully-occupied system) using API calls. Prompt analysis results are returned to the beamline software environment for easy visualization to the experimentalists, giving them confidence in their experimental methods and sample characterizations.

The scientists may adjust their methods based on the prompt analysis data and on a concurrently-running simulation of the experiment environment, allowing the collection of a detailed data set which completely describes a new biochemical process. This real-time improvement would not be possible without the concurrent use of the unique beamline, the ability to refine the experiment in the moment as guided by rapid feedback from HPC, and the Exascale simulation which guides the scientists in focusing on the most relevant aspects of the samples.

After the beam time is done and the scientists return home, they conduct in-depth analysis of the data set. They use the provenance captured by the workflow to integrate the simulation results with the experimental results, and they subsequently publish their work, which provides critical decision support to policymakers while also breaking new scientific ground. The entire data set (experimental data, simulation output, simulation code, and published data) is made available in a data portal that allows other scientists to discover the data and integrate it into subsequent analyses.

The scientist's journey in this vignette is presented in more detail, and with references to specific technologies, in Appendix 2: User Journey Map.

While this vision may seem far off, many individual research and pilot projects have already demonstrated the power of integrating facilities for the benefit of science. The challenge now is to explore pathways from labor-intensive proof-of-principle demonstrations to seamlessly-executed automated services, whose complexity is hidden from the end user.

Our goal is to provide science teams with capabilities that would otherwise be unavailable, and thus provide a clear leadership advantage to DOE facility users, while expanding ASCR leadership in large-scale simulation, data analysis, and AI domains. Achieving the vision will promote maximum productivity of DOE-funded research infrastructure and promote US global leadership in science. While the example above is focused on the obvious synergy between BES light sources and ASCR computing and networking facilities, there are numerous examples from other SC Program Offices with similar requirements documented in a variety of recent formal workshops spanning the last several years. Appendix 3: Collected Requirements from the Exascale Crosscut Report, provides a detailed summary of these requirements.

In the rest of this white paper, we provide a high level blueprint for a path forward, beginning with a set of guiding principles and a discussion of identified gaps, risks, and opportunities.

Principles

In defining the components of an integrated research infrastructure, we adhere to a set of guiding principles. Enabling these principles in all components of the infrastructure will ensure that we build a system that can grow, extend, and adapt to new science use cases. The principles will also allow the independent constituents of the integrated infrastructure to retain a common core upon which to build a unified solution, while maintaining their individual operational integrity.

- **Flexibility**
  Distributed HPC computing resources should be exposed as simple consumable services that can be easily assembled together while concealing the complexity of the system. The overhead to customize capabilities for new communities should be small.

- **Performance**
  The default behavior should be performant, without requiring arcane options or complicated scripting. For example, the location of computing, experiments, and data should not be an immediate obstacle to high-performance campaigns.

- **Scalability**
  The infrastructure should support ***at-scale*** data capabilities without requiring excessive human effort or undue customizations as data scale increases. Data transfer and data access in place, at scale, are both critical capabilities.

- **Transparency**
  The infrastructure should provide transparent mechanisms to enable resilient workflows, seamless data transfer, and easy access to all facilities. The security, authentication, authorization, and related policies and technologies should support (cross-facility) automation.

- **Interoperability**
  Services deployed within the infrastructure should be interoperable across facilities, and should extend to analogous services outside the environment (e.g., at non-DOE-funded supercomputer centers).

- **Resiliency**
  Infrastructure should be reliable and resilient, in order for projects to meet their mission needs. Workloads may be moved in response to planned and unplanned events. Moving computing to data and moving data to computing should both be supported as first-class capabilities.

- **Extensibility**
  The infrastructure should be designed to accommodate future needs and be able to adapt with minimal disruption. Enabled by flexibility and interoperability, this design principle will support initiatives such as new AI for Science methodological and capability deployments.

- **Engagement**
  Frameworks and processes should be established to enable user facilities to cooperatively innovate and co-design, prototype, and productionalize end-to-end architecture solutions. Strong user engagement and partnerships will drive solutions aligned with user needs.

- **Cybersecurity**
  The solutions undertaken to improve scientific productivity must be secure, both for facilities and for scientific users.

Gaps

The vision above is a significant step beyond what is possible with today's infrastructure. By evaluating the ability of facilities to meet the needs of this vision, we have begun to identify gaps that exist today; these are described in the table below, with potential responses that aim to close the gaps. This summary is derived in part from gaps identified in numerous past workshops and requirements reviews. (The requirements from the Exascale Requirements Review Crosscut Report are identified and cross-linked in Appendix 3.) The list below is not comprehensive but merely a first pass of the gaps that are immediately visible to the Task Force.

Area	**Gap**	**Potential Response**
Allocations	Science projects and facilities lack multi-site, multi-year computing and data storage allocations.	Re-evaluate existing allocation programs. Establish a model to support multi-site, multi-year allocations for both compute and data storage at ASCR facilities.
Accounts/Access	Users have to establish accounts and identities at each site.	Establish a federated identity model that can be supported at all user facilities while still adhering to facility cyber-security policies.
Data/Archives/ Publishing	Projects lack the ability to transparently access large scale datasets across multiple facilities, lack access to distributed data sharing mechanisms, and lack the tools that address the full life-cycle of the data.	Establish multi-site storage allocations and agreements on long-term storage policies, coupled with sufficient storage resources. Collaborate and explore solutions that would enable a distributed archive or data repository across SC User Facilities and accompanying solutions for the data management life cycle.

Workflows/Applications /Scheduling	Science projects cannot seamlessly execute and schedule workflows across facilities for resiliency and other purposes.	Develop approaches to enable projects to execute complex workflows seamlessly across facilities and adapt scheduling policies to address the full range of workflow patterns.
Overarching functional areas		
Policies and Governance	Cross-facility governance, policies, and metrics are not yet aligned with the proposed integrated infrastructure.	Update facility metrics and policies to align and reward a broader set of capabilities unique to ASCR and other SC User Facilities that align with a more integrated vision. Ensure that cyber-security concerns and related site policies are addressed in the integrated infrastructure.
Engagement and Partnerships	Multi-facility workflows and next-generation data analysis techniques don't always have a group of cross-cutting experts to bring these capabilities into science collaborations broadly.	Expand, extend, and enhance the collaborations that support the integration of facility advancements into scientific workflows, science collaborations, and science programs. Explicitly support engagement, knowledge sharing, and coordination between facility engineering staff and researchers, as well as with facility users.

Blueprint for Integrated Infrastructure

This is a ***notional*** blueprint meant to illustrate the types of activities and coordination that could be required to deliver an integrated research infrastructure. Developing a comprehensive blueprint that considers the available resources and balances priorities would require further consideration and should be done in partnership with other stakeholders across the Office of Science enterprise.

In the near term, several goals can be achieved which would lay the groundwork for a broader implementation of the vision. We note that many of these activities are already being done piecemeal by individual teams or facilities. The aim of the proposed effort is to integrate these capabilities across multiple facilities and bring them into production. A cross-cutting "smart automation and integration" approach will underlie the design, which will also further enable AI for Science activities.

- **Develop an allocation mechanism that awards time to projects across facilities and for multiple years.** The mechanism should include allocations for both compute and storage. We suggest a pilot program within ALCC.

- **Begin the deployment of a federated identity solution across ASCR facilities and other DOE experimental and observation facilities.** The technologies to achieve this solution are widely available and deployed in academia and industry.

- **Begin to establish common baseline APIs for cross facility data management and job scheduling**, and leverage these APIs to enable cross-facility workflows that can be monitored at experiment-time.

- **Begin to develop a distributed data management system** that facilitates cross-facility work, through replication or distributed metadata.

- **Explore how to support and measure performance and resilience,** which might include continuous integration infrastructure to test the readiness of applications across compute facilities.

- **Identify new partnerships and continue deep engagement with existing science partners** through workshops, requirements gathering, and regular communication. This effort will help ensure the proposed infrastructure tracks the needs of the science community, and will help experimental and observational partners to adapt to using the new framework.

- **Reexamine and expand the ASCR HPC facility metrics** to better align with more diverse workloads and applications. In particular this should expand the definition of ‘capability’ from solely considering the number of compute nodes used to a broader set of criteria, while continuing to emphasize large-scale simulation science.

- **Establish a governance panel** that would include delegates from the ASCR Facilities and a representative number of experimental SC User Facilities to draft an agreement for how cross-site authentication, allocations, APIs, distributed services, etc would be maintained and supported.

- **Target a small number of pilot projects** that use the various functional pieces defined in this document, in production, in their workflows across multiple facilities.

In Figure 2 below, we provide a notional planning roadmap based on the blueprint for illustrative purposes. As noted above, the development of a comprehensive roadmap would require further consideration and should be done in partnership with all stakeholders. Blank cells in years 3-5 indicate graduation or acceptance of a capability into operations or into a stewardship model that is yet to be determined.

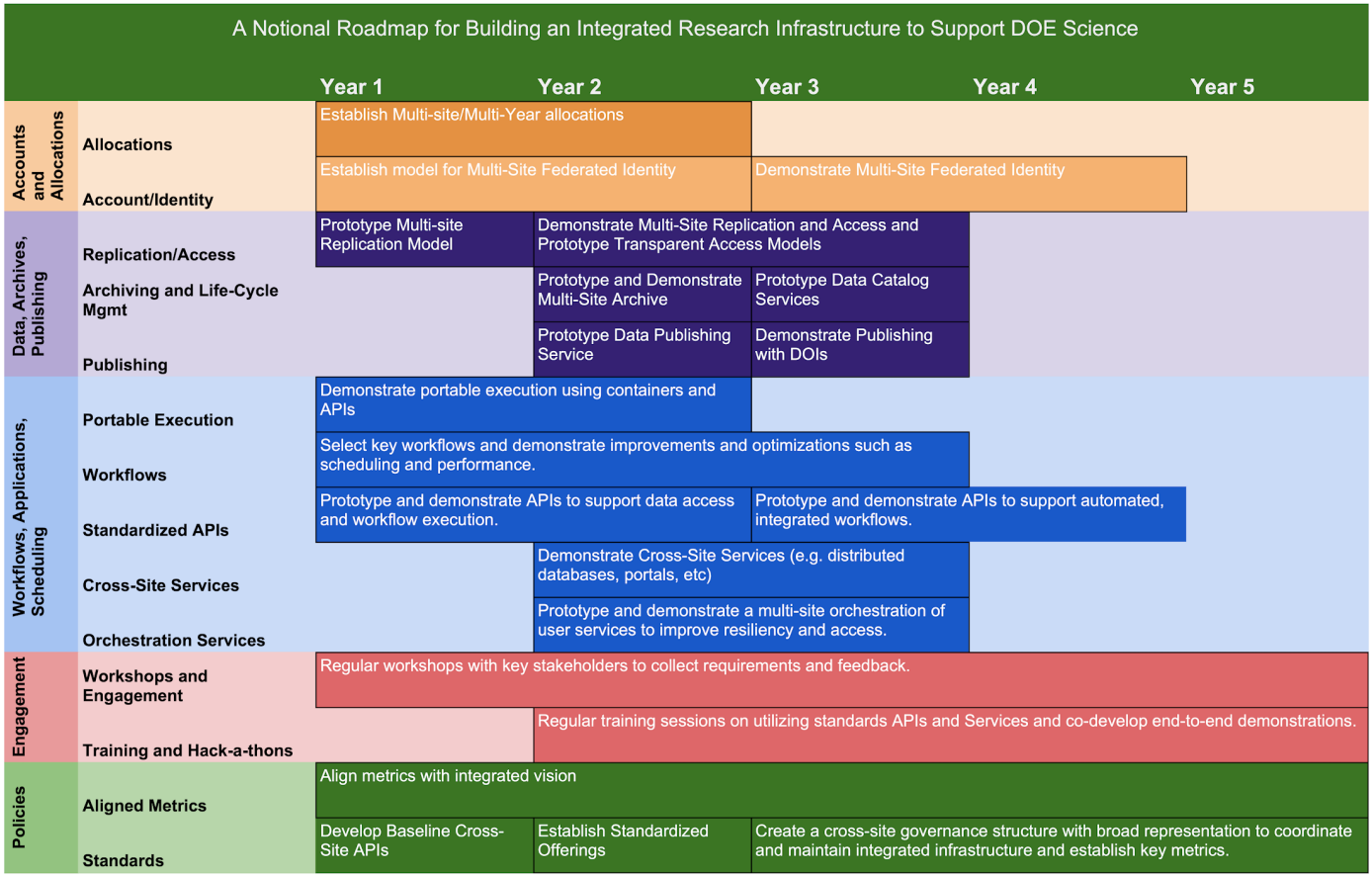

Figure 2. *This notional roadmap illustrates the types of activities and coordination that would be required to deliver an integrated research infrastructure. Developing a comprehensive roadmap that considers the available resources and balances priorities requires further consideration and should be done in partnership with other stakeholders.*

Activities Already Underway

Many of the activities and priorities described in this document are in various stages of progress and have benefited from ASCR research funding. For example, each ASCR HPC facility, as well as ESnet, has demonstrated integrated workflows for use cases at a variety of SC User Facilities, including light sources, nanoscience centers, high energy physics experiments, etc. Each ASCR HPC facility has deployed workflow software to facilitate scientific pipelines from experimental facilities and all of the ASCR HPC facilities have solutions that allow users to share, archive, and manage data. The Distributed Computing and Data Ecosystem (DCDE) project recently demonstrated the feasibility to execute workflows and move data across various labs with a single identity.

A key goal of this blueprint is to describe a path forward from these individual partnerships and distinct deployments and demonstrations, to a broader long-term strategy with common and reusable software, APIs, and components across facilities and science areas. A sampling of current activities and projects is provided in Appendix 4: Existing Projects.

Risks and Challenges

We have identified a number of risks and challenges that underlie our proposal. Without an integrated research infrastructure, we expect there will be missed opportunities and growth risks to the DOE science enterprise. We also acknowledge that this work will be difficult, and we outline here the main challenges we foresee.

Risks

- If experimental facilities cannot leverage ASCR resources, they may deploy their own larger local computational resources to meet their mission needs, leading to siloing and inefficiencies.
- If experimental facilities come to rely on computational resources or capabilities available at a single ASCR facility without backup/failover resources, those experimental facilities may experience service outages impacting their workflows, or deploy their own non-integrated resources to mitigate risk. Similarly, if users cannot rely on multiple ASCR facilities for data services, then disruptions in data availability or even loss of data may occur.
- If an integrated research infrastructure is not created to address the growing HPC needs of various stakeholders, then individual stakeholders will create local and potentially suboptimal solutions to fill the gap. DOE/SC/ASCR is well positioned to lead both nationally and internationally with the creation of such an infrastructure.
- If common and scalable capabilities are not available to experimental facilities, they may be left to implement what they can using the tools and systems they have available, incurring significant cost and disruption for upgrades.
- If there is no way to reliably transfer data from experiments to HPC, valuable experimental data will be kept in ad-hoc storage (e.g. hard drives in boxes or filing cabinets). This “dark data” will be unavailable to the wider scientific community, and may represent a significant opportunity cost or require experiments to be re-run in order to make the data more broadly available.

Challenges

- Differences in system architectures, storage systems, services, and cybersecurity policies may complicate or restrict an integrated approach.
- There is a broad range of workflows of high mission importance; care must be taken to establish common components and services and thus avoid divergent bespoke solutions that lack a common foundation.
- The ASCR HPC Facilities have unique capabilities and must be allowed to innovate, which must be balanced with the need for standardization and interoperability.
- Enabling this integrated infrastructure and the related scientific workflows should not imply that existing world-leading scientific workloads at HPC facilities will be prioritized any less. The expansion demands for the HPC facilities’ limited resources could exacerbate challenges related to resource oversubscription and user competition.

- Expanding the diversity of workloads accommodated at ASCR HPC Facilities may cause inherent challenges with scalability of resources.
- High-performance end-to-end networking, which underlies many of the distributed data capabilities described in this document, requires the adoption of network designs, technologies, and policies (e.g. the Science DMZ model) that are not currently deployed at many DOE facilities. While the ASCR Facilities can (and do) provide assistance and expertise to help, the DOE National Laboratories and User Facilities complex as a whole will need to modernize its approach to high-speed networking, and solve the “last mile” problem at facilities, from edge to instrument. The ASCR Facilities will be called upon to help, which will require resources.

Role of the ASCR Research Community

While we have identified a number of near term opportunities, our vision cannot be achieved without the engagement of the ASCR research community and an investment in the longer term challenges. A 2019 workshop, “DOE Computational Facilities - Research Workshop” held at Argonne National Laboratory with participants from both ASCR research and ASCR facilities aimed to understand the requirements and challenges of the DOE computational facilities, and to brainstorm research directions with the ASCR research community to positively impact DOE computational facilities. A number of the research directions identified at this and other workshops in recent years have direct relevance to achieving our vision of a more integrated ecosystem across ASCR facilities. They include:

- Identifying and managing workflow patterns and resource scheduling for emerging workloads, resource types, and architectures
- Determining primitives and abstraction levels for scheduling, managing, and executing workflows within and across facilities
- Enabling end-to-end workflow monitoring, modeling, optimization, and automation
- Characterizing workflows for optimal resource allocation to maximize system efficiency
- Developing failure prediction to reduce system and network disruption through proactive maintenance
- Detecting new threats for emerging HPC and high-performance network architectures
- Addressing OS, system, and network management needs for emerging workloads using shared resources
- Providing more automated and intelligent storage systems
- Ensuring robust, searchable, automatable metadata
- Addressing the storage and I/O needs of emerging workloads

Summary and Next Steps

We have described a vision for an integrated ecosystem spanning DOE SC’s ASCR facilities and experimental facilities. This integrated fabric will create new scientific opportunities for the broader Office of Science community. It will provide a competitive advantage to DOE facility users and will expand ASCR leadership in computational science as experimental data grows

and initiatives such as AI for Science unfold. We have identified both near term opportunities, as well as longer term goals and have provided a notional blueprint for a path forward, as well as ASCR research needed to implement this vision. Our immediate next step is to solicit feedback from key stakeholders on the proposed vision and blueprint, and define the concrete succeeding next steps.

Appendices

Appendix 1: Task Force roster

Name	**Affiliation**
Corey Adams	ALCF
Katie Antypas	NERSC
Debbie Bard	NERSC
Shane Canon	NERSC
Eli Dart	ESnet
Chin Guok	ESnet
Ezra Kissel	ESnet
Eric Lancon	SDCC
Bronson Messer	OLCF
Sarp Oral	OLCF
Jini Ramprakash	ALCF
Arjun Shankar	OLCF
Tom Uram	ALCF

ALCF Argonne Leadership Computing Facility at Argonne National Laboratory
ESnet Energy Sciences Network at Lawrence Berkeley National Laboratory
NERSC National Energy Research Scientific Computing Center at Lawrence Berkeley National Laboratory
OLCF Oak Ridge Leadership Computing Facility at Oak Ridge National Laboratory
SDCC Scientific Data and Computing Center at Brookhaven National Laboratory

Ben Brown, ASCR, served as the facilitator for the Task Force.

Appendix 2: User Journey Map

User Action	**Technologies and Capabilities Used**
User is allocated beam time, issued an account tied to a successful proposal, completes safety training, arrives onsite, etc. The user has 48 hours of continuous beam time. The user and their primary student will alternate shifts.	Light source facility proposal acceptance, allocation, scheduling, intake, training, etc.
User starts up the beamline environment, clicks “pre-flight check” for tomography in the UI. User is planning to use the MondoPixel SuperCamera, with an output data rate (after zero suppression and compression by on-camera FPGA or beamline edge computing) of 25Gbps.	Beamline control software reaches out to HPC facility to provision the experiment environment. Light source account is matched against HPC facility UID/role/permissions (federated ID). Analysis containers with appropriate tomography software are staged. HPC facility scheduler makes appropriate resources available. Network bandwidth reservations are made and bound to the beamline DTN and the HPC facility compute nodes or network environment (as appropriate).
User console “goes green” indicating the environment is ready for data taking.	HPC and Network facility environments are ready, with control APIs listening for messages from beamline control software.
User places a sample on the stage, closes the hutch, and engages the interlock.	Beamline control software notices that the hutch is locked, and makes the API call to launch the prepared computing environment. Containers are launched at the HPC facility. API call returns, indicating to beamline control software that the computing environment is live.
User takes data using the beamline UI	Data comes out of the MondoPixel SuperCamera at 25Gbps, traverses the network to the container environment at the HPC facility, and is ingested and analyzed.
User finishes data taking, and clicks stop.	HPC environment finishes prompt analysis, publishes thumbnail images on the web service corresponding to this user’s work, and sends the URL back to the beamline environment.
User views thumbnails, decides that the sample is done, and switches out the sample.	Beamline environment informs HPC environment of user actions/events. The HPC environment notices that the sample change

	time is short enough that there is no reason to try and schedule other jobs on the nodes between beamline runs.
User runs through seven more samples using the same procedures as the first two.	HPC environment takes the data, produces prompt analysis, stores raw data and prompt analysis products in storage system, with appropriate naming, permissions, etc.
After nine samples, the user sees what appears to be some sort of artifact in the data. They go outside for some fresh air with a colleague, talk about it, and then go find a conference room to whiteboard the issue.	HPC environment notices that it's not being used, pauses the containers, tells the beamline that it's paused so the beamline can indicate pause state to the user, and schedules some jobs that have been previously marked as interruptible from the batch queue on the nodes. The user is still charged at some reduced rate for their reservation of the allocated nodes.
User comes back, sees the beamline is paused, and clicks "resume." User puts sample 10 in the hutch, and engages the interlock.	HPC environment fires back up when the user clicks resume, or when the hutch is locked. The interruptible batch jobs are killed, reverting to their last checkpoint.
User processes 5 more samples.	Normal operations as above. HPC scheduler notices that a full-system job is going to be run in 20 minutes. HPC environment issues API calls to a different HPC facility, and tells it to stage the environment. Users are mapped (using federated ID), containers are staged, network reservations are made, etc. The beamline software is notified when all this is ready.
The user puts in sample 16, and engages the interlock.	Beamline environment switches to the backup HPC facility.
User processes sample 16. The prompt analysis displays images in an expected manner.	The pipeline is now running at the backup HPC facility. Prompt analysis sends back URLs to images and metadata in the normal way, which shows up in the web interface in the normal way, despite the user now viewing prompt analysis results from two separate facilities. Primary HPC facility launches a full-system job for Dr. Wowzer's Gordon Bell submission - the job will run for 24 hours.
User goes to bed. Primary student takes over.	Pipeline continues to run at backup HPC facility. After 24 hours, Dr. Wowzer's job at

	the primary HPC facility completes. Primary HPC facility calls the backup HPC facility API, and the process of moving back to the primary facility takes place, in the same way that the transition to the backup HPC facility took place.
Student is heads-down processing samples.	Unbeknownst to the student, the processing pipeline has moved back to the primary HPC facility. Primary HPC facility launches a data transfer job to migrate data from the backup facility so that the data set for the project is in one place. Metadata database sync and other relevant operations are also done. Once those operations have completed (including integrity verification), the backup facility cleans up.
After 48 hours of hard work, the user and the student both head back to their hotel rooms and sleep for 12 hours. The next day, they get on a plane and fly home.	When the beam time ends, the container environment is torn down (though provenance information is preserved), and the HPC facility returns the nodes to the batch queue.
Three days later, the user sorts through their data. There are 31 runs that look good enough to be inputs to publication. The student's PhD thesis includes special analysis code that runs on a third HPC facility (it has special accelerators). The 31 data directories are transferred using a simple transfer tool. In total, the transfer volume is 273TB. The transfer completes overnight.	This is normal stuff that the HPC and network facilities do today. It's important, but there isn't significant new work to be done to support this step.
Once the analysis is done, the results are moved back to the primary HPC facility. The user and the student share the results with the student's PhD committee, but with nobody else.	Data portal which supports very large data sets is used to support access to analysis products by specific individuals. Those individuals do not have HPC facility accounts or any access other than to view the data shared with them in the portal.
The paper is submitted, accepted, and published! User and student make the results public in the portal.	Data portal now shows their results in searches by other scientists which match the metadata of their results.
A scientist in another field has a large compute allocation for characterizing detector behavior in an attempt to aid the design of	The data portal API supports a scalable data transfer tool (e.g. Globus). If network reservations need to be made, they are

next-generation detectors. This scientist searches through the data portal, and selects all the samples from this and 93 other high-rate tomography experiments. The scientist transfers this data set (12.7PB in size) to an HPC facility with an Exascale machine.	made. If the portal DTNs can handle it on their own, they do that. If the job should run in the background, it can. There is sufficient space in the scientist's storage allocation at the Exascale site to receive the data set, and the data transfer completes over a few days (2-3 years from now) or over the weekend (5-8 years from now) or overnight (10 years from now).

Appendix 3: Collected Requirements from the Exascale Crosscut Report

Below is a table of requirements generated from the Exascale Requirements Reviews Crosscut Report.[3] The requirements are grouped into 8 areas that reflect how Experimental and Observational facilities interact with ASCR HPC and HPN facilities.

	ASCR	BER	BES	FES	HEP	NP
Allocations - Experiments need sturdy multi-year allocations						
Long-term multi-year allocations			BES 4.2	FES 3.1.2.4.3	HEP 1.2	
Inter-facility transferable allocations			BES 3.6, 3.7.2.2			NP 3.1.4, 3.2.3.2
Accounts - Account processes should be simple, fast, and common across facilities						
Common access (e.g., Federated ID)			BES 3.6, 3.7.2.2, 4.3			NP 3.1.4, 3.2.3.2
Faster access to compute resources for development, debugging, analysis, or real-time computing	ASCR 3.2.1.1, 3.4.2.1	BER 3.2.1.1.4	BES ES.1, 3.1.4	FES ES.4	HEP 4.1.2	NP ES.1
Data - Data should be easy to find, access, and share within and across facilities						
Seamless access to data across facilities		BER 3.2.3.4				
Persistent large-scale storage		BER 3.1.4.4.1	BES 3.6.2.3.1		HEP 4.1.2	NP 3.1.4
Data management, curation, and publishing		BER 2.1	BES 3.6.2.3	FES 3.4.2.2 3.4.2.4		NP 3.2.3.2
Common data standards and federated databases			BES 3.3.4	FES 4.3		
Seamless movement of simulation and experimental data to a compute resource					HEP ES, C.2	FES 3.1.4

[3] Published January, 2018 https://doi.org/10.2172/1417653.

Predictable data transfers		BER 3.2.6.1.1			HEP 4.1.2	
Applications - Experimental codes need to be adapt to diverse architectures for performance						
Application portability (e.g., multi-HPC common tools and practices, port to new architectures)	ASCR 3.6.1	BER 3.2.4.4	BES 3.6, 3.7.2.2, 3.7.2.3	FES 4.2		NP 3.1.4, 3.2.3.2
Common software ecosystem	ASCR 4.1.2	BER 4.2	BES 4.2	FES 4.2	HEP 4.1, 4.2	NP 4.1
Use of HPC and non-HPC compute	ASCR 3.2.1.1	BER 3.1.4, 3.2.6	BES 3.6	FES 3.4	HEP 3.2, 4.1.2	NP 3.2, 4.1
Scheduling - Unified scheduler interface and multi-facility scheduling capabilities						
Real-time computing for decision making			BES ES 3.6, 3.6.2.4	FES 3.4.1.1		
Cross facility co-scheduling to connect experimental and HPC facilities during experiment		BER 3.1, 3.1.4.2.3	BES ES.1, ES.3.6, 3.6.2.1.3, 3.6.2.2	FES ES.4, 3.4.1.1, 4.4	HEP 4.1.2	NP ES.3.2, 3.2.1
Workflows - Management of job campaigns from simple sequence to complicated multi-facility job graphs						
Usability and simplicity for non-expert users of the resource (e.g., HPC, HPN, etc)			BES 3.5.1, 3.8.2			NP 3.4.4
Publication - Citable data products (at scale) improve science and conform to DOE policy						
Configuration of data resources (i.e., provenance information) are being published correctly.		BER 3.2.6.1.1				
Archiving - Cold storage for data likely needed in the future						
Data storage for large-scale post and reprocessing		BER 3.2.1.2		FES 3.1.4	HEP 3.1	

Appendix 4: Existing projects

Title	Dates	Objective	Status	Key components
Future Lab Computing Working Group - Distributed Computing Data Ecosystem (DCDE).	2018-	Federated ID for users of DOE facilities to work across each other's resources.	Active design work after 2019 focused exploratory pilot.	Federated ID, data transfer, remote access
BEAM Workflow	2015-2017	Connect CNMS to CADES	Custom workflow completed. Currently morphed into Pycroscopy and custom Jupyter workflows, and Data sharing workflows (DataFed).	Local processing, and data transfer workflows. Scalable computing on institutional resources and access to Titan/Summit supercomputers.
BER-ARM Operational Data Workflow	2016-	Ingest and process Atmospheric Radiation Monitoring Data on a nightly basis across CADES and prepare for OLCF runs.	Ongoing	Workflow across facilities.
SLATE/Summit Workflows	2019	Workflow nodes adjacent to summit	Active	Kubernetes-based orchestration.
LHC-ATLAS	2015-	Leverage leadership-scale machines for ATLAS simulation/analysis	Active	Job scheduler, data mover, data streamer (remote I/O, XRootD), data catalog
Light Sources and Computing Facilities Working	2019-	Identify integration points between light sources and computing	White paper written (included in	

Group		facilities to meet near-term computing needs	this google drive)	
LSST-DESC	ongoing	Online and offline processing of experimental data that requires leadership scale resources	Active	Containers, workflow software, online processing
DUNE/Neutrinos	2018-	Port Fermilab neutrino-experiment simulation and analysis tools to leadership-scale facilities. Long term: DUNE	Active	Containers, workflow software
QMC / QCD		Perform Quantum monte carlo simulations.	Active	Data transfer (Globus), parallel filesystems, computing, multi-site
Petascale DTN	2016-	High speed transfers between HPC facilities	Done	Globus, DTNs, parallel filesystems
BigData Express	2015-	Schedulable, high-performance data transfer	Active	mdtmftp, DTNs
Data Demos (SC14)	2014	Demonstration of potential impact of integrated data capabilities for DOE Science.	Done	Data movement, workflow, event/trigger from experiment, prompt data analysis
ZTF (Zwicky Transient Factory) pipeline	ongoing	Short turnaround data analysis pipeline, demonstrated ability to run at LBNL, NERSC and AWS	Active, with strategic trajectory	Data transfer, workflow, storage, computing pipeline, prompt analysis/results notification
SENSE	ongoing	Inter-domain provisioning for network and DTN resources	Active, pre-production ESnet service, active adoption in R&E	Resource modeling, API for resource negotiation and provisioning, end-to-end monitoring

			network collaborations (e.g., AutoGOLE)	
LLAna pilot project	10/19 - 10/20	Data analysis tools for LCLS-II	Active,	Workflow, multi-site data analysis w/ Jupyter
Superfacility API	ongoing	API at NERSC for automated tasks, ewg data management, user management, job submission, reservations etc	Active	Data transfer, workflow, automation, API for resource request and management
Balsam workflows	ongoing	Workflows software for optimal execution of large simulation campaigns or HTC workloads	Active	Jobs database, execution engine, improved throughput, provenance support
ALCC award: Towards Resilient and Portable Workflows across DOE's Facilities	06/20 - 06/21	ALCC award of time at ALCF, OLCF and NERSC to research the challenges in running EOD analysis pipelines at multiple sites (LBNL LDRD pending)	Active	Data management, containers, multi-site
LCLS-II data analysis	08/20 +	Data analysis performed locally (SLAC) and at NERSC	Active development, production running expected late 2020	Data management and transfer (via SENSE), full-machine realtime analysis
LBNL Superfacility project	01/19 - 01/21	Coordinating engineering and research at NERSC, ESnet and CRD to support connected facilities	Active	Data management and transfer, Jupyter, API, automation in compute and network, realtime computing, scheduling