Integrated Research Infrastructure – Architecture Blueprint Activity

Connecting instruments, data, and minds

Jini Ramprakash
Deputy Division Director, Argonne Leadership Computing Facility

May 10, 2023

Argonne National Laboratory

The U.S. Department of Energy's Argonne National Laboratory delivers world-class research, technologies, and new knowledge that aim to make an impact — from the atomic to the human to the global scale.

DOE User Facilities

- **Cutting-edge resources:** Advanced tools, instruments, and expertise
- **Open access:** Available to academia, industry, and government researchers
- **Collaboration:** Interaction with scientists from various fields
- **Education and training:** Workshops, seminars, and skill development
- **Economic impact:** Driving scientific advancements and innovations
- **U.S. competitiveness:** Supporting groundbreaking discoveries
- **National priorities:** Research in security, energy, and sustainability

Argonne National Laboratory DOE User Facilities

Advanced Photon Source

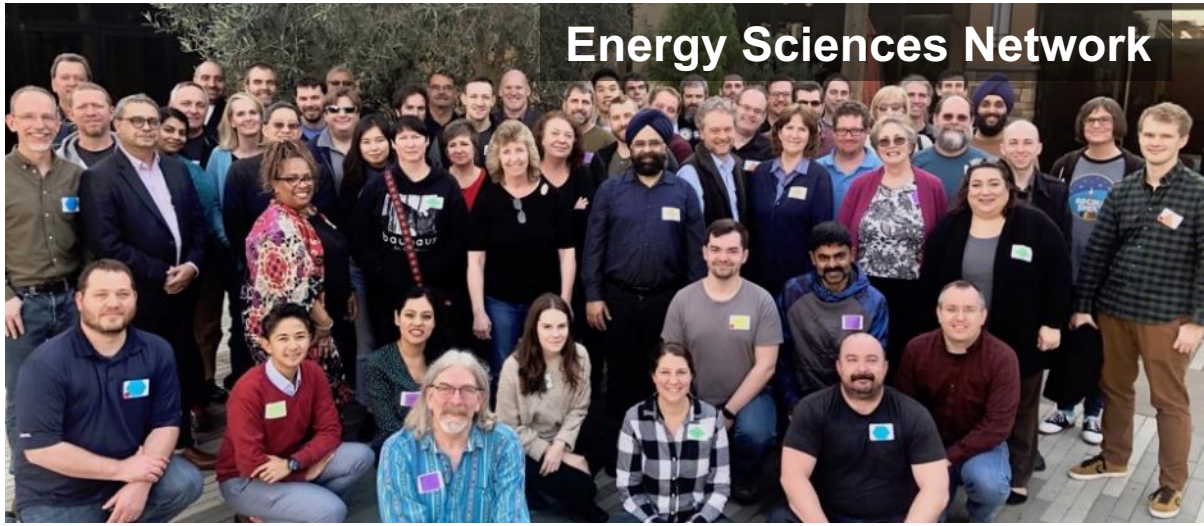
Argonne Tandem Linear Accelerator System

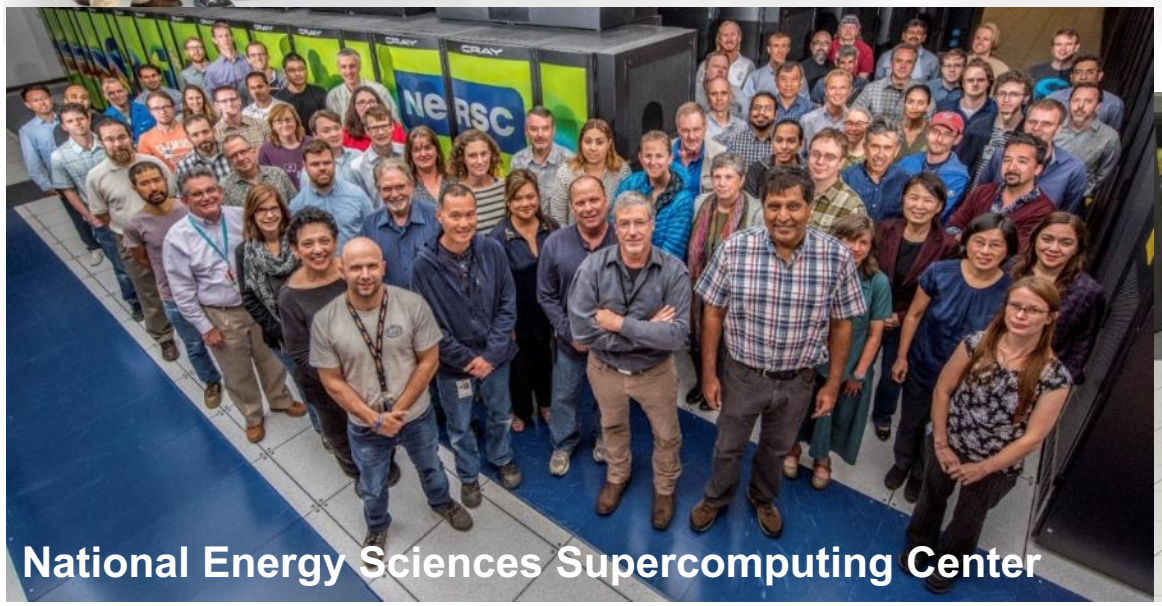
Argonne Leadership Computing Facility

Center for Nanoscale Materials

Atmospheric Radiation Measurement – The Southern Great Plains

The people of the ASCR Facilities: Providing high performance Research Computing, Data, and Networking for DOE and the Nation



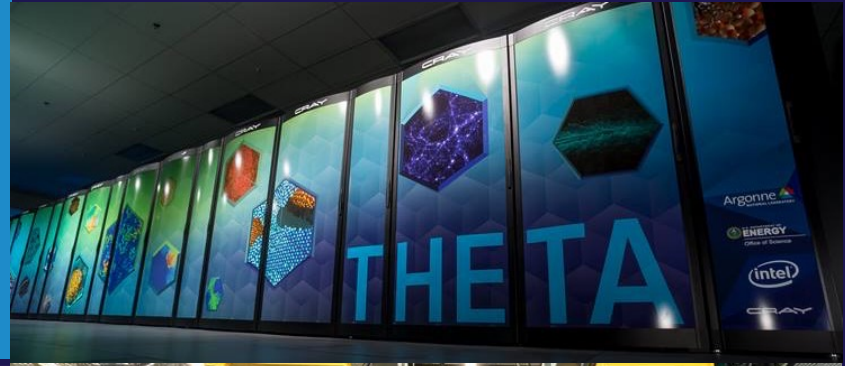


DOE SC
Advanced Scientific Computing Research User Facilities

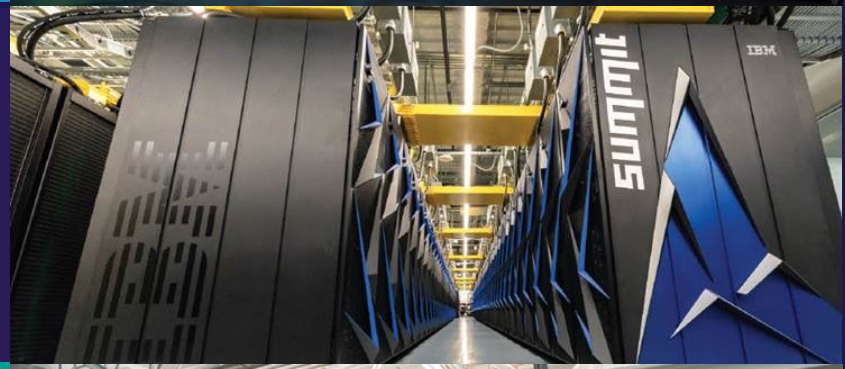
The Advanced Scientific Computing Research (ASCR) program leads the nation and the world in supercomputing, high-end computational science, and advanced networking for science.

ALCF and OLCF make up the DOE Leadership Computing Facility

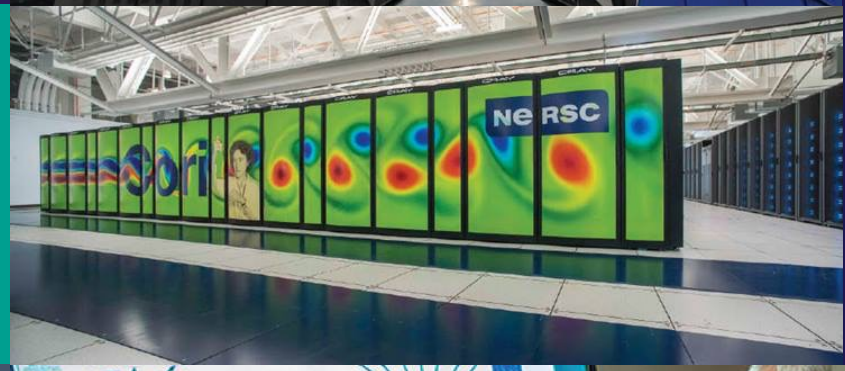
Argonne Leadership Computing Facility (ALCF)



Oak Ridge Leadership Computing Facility (OLCF)



National Energy Research Scientific Computing Center (NERSC)



Energy Sciences Network (ESnet)

DOE's Integrated Research Infrastructure (IRI) Vision:

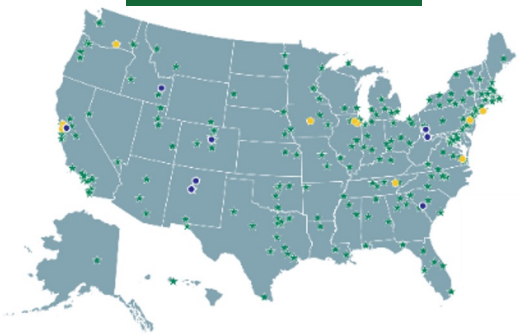
To empower researchers to meld DOE's world-class research tools, infrastructure, and user facilities seamlessly and securely in novel ways to radically accelerate discovery and innovation

Advanced Networking

Edge Sensors

Advanced Computing

Researchers



Local Campus Computing

Computing Testbeds

Cloud Computing

Data Management

High Performance Data Facility

Data Repositories PuRE Data Assets

Software

Software and Applications

AI Tools Digital Twins

New modes of integrated science

Rapid data analysis and steering of experiments

Novel workflows using multiple user facilities

AI-enabled insight from integrating vast data sources

Timeline of key IRI activities, 2020-22

FY 2021 President's Budget Request (PBR) includes Integrated Computation and Data Infrastructure Initiative

ASCR IRI Task Force launch

ASCR IRI Task Force report

FY 2022 PBR includes Integrated Computation and Data Infrastructure Initiative

SC User Facilities COVID era report

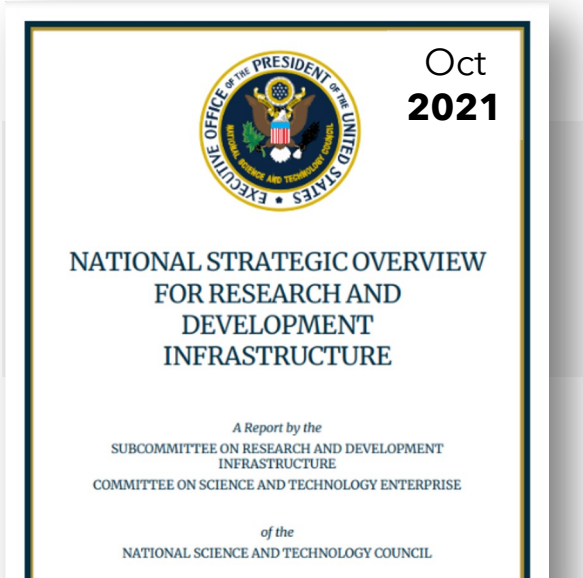
FY 2023 PBR Advanced Computing Initiative includes Integrated Computation and Data Infrastructure

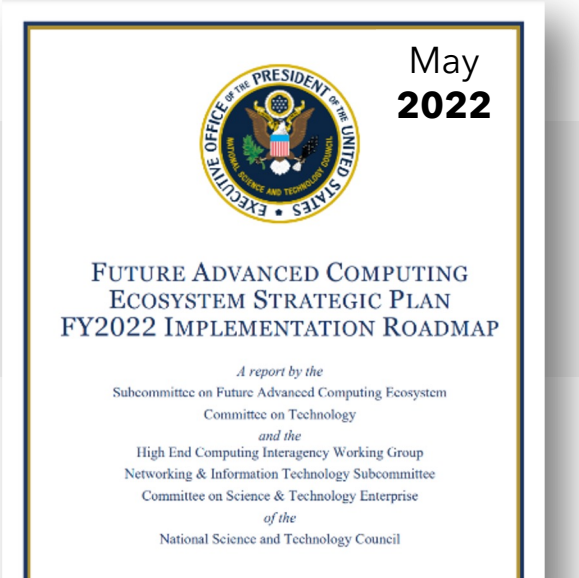
SC IRI Blueprint Activity launch

IRI Blueprint Activity results

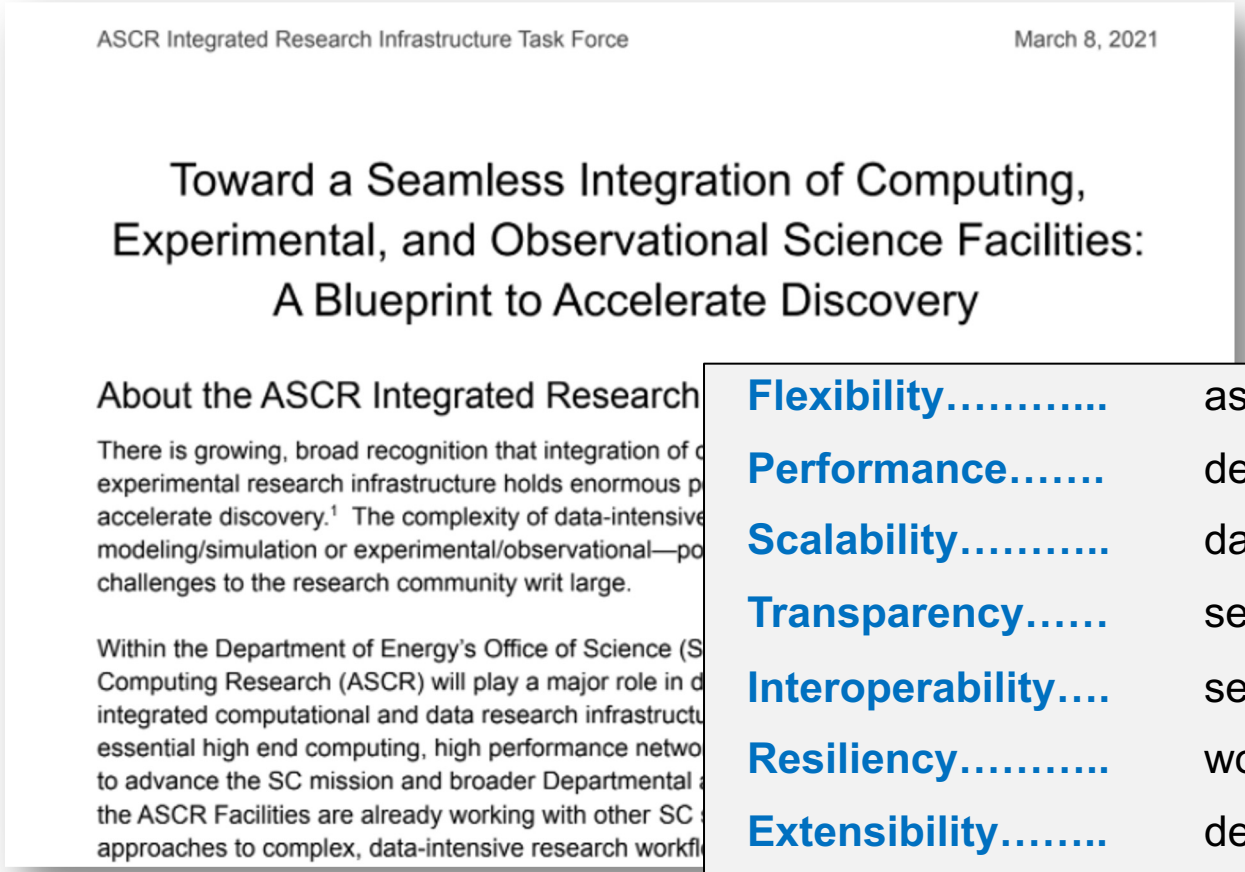
Jan **2020** Mar Jan **2021** Mar Oct Jan **2022** Mar Sep Jan **2023**

Integration of instrumentation, data, and computing infrastructure are essential requirements for national R&D objectives





In 2020-21, the ASCR IRI Task Force organized the ASCR Facilities' thinking and approach



ASCR Integrated Research Infrastructure Task Force March 8, 2021

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

About the ASCR Integrated Research

There is growing, broad recognition that integration of c
experimental research infrastructure holds enormous p
accelerate discovery.[1] The complexity of data-intensive
modeling/simulation or experimental/observational—po
challenges to the research community writ large.

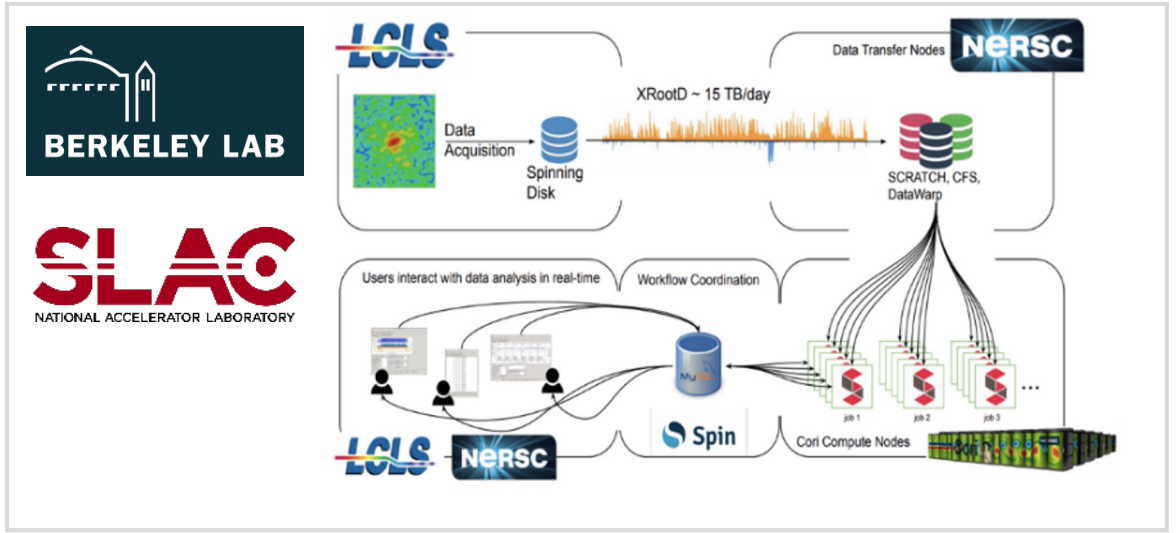
Within the Department of Energy's Office of Science (S
Computing Research (ASCR) will play a major role in d
integrated computational and data research infrastructu
essential high end computing, high performance netwo
to advance the SC mission and broader Departmental
the ASCR Facilities are already working with other SC
approaches to complex, data-intensive research workfl

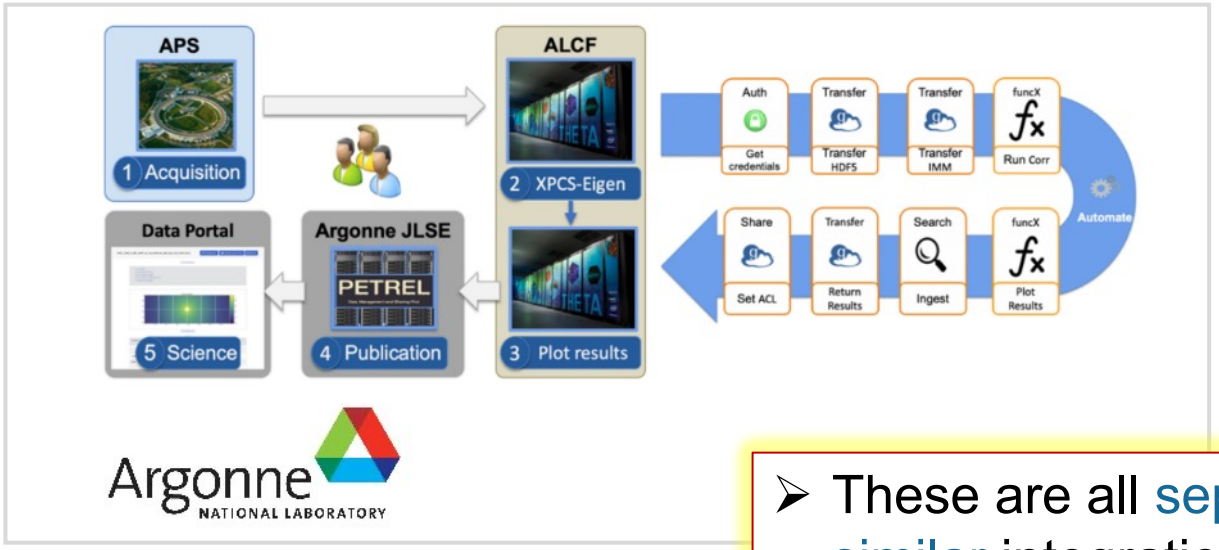
Corey Adams
Katie Antypas
Debbie Bard
Shane Canon
Eli Dart
Chin Guok
Ezra Kissel
Eric Lancon
Bronson Messer
Sarp Oral
Jini Ramprakash
Arjun Shankar
Tom Uram

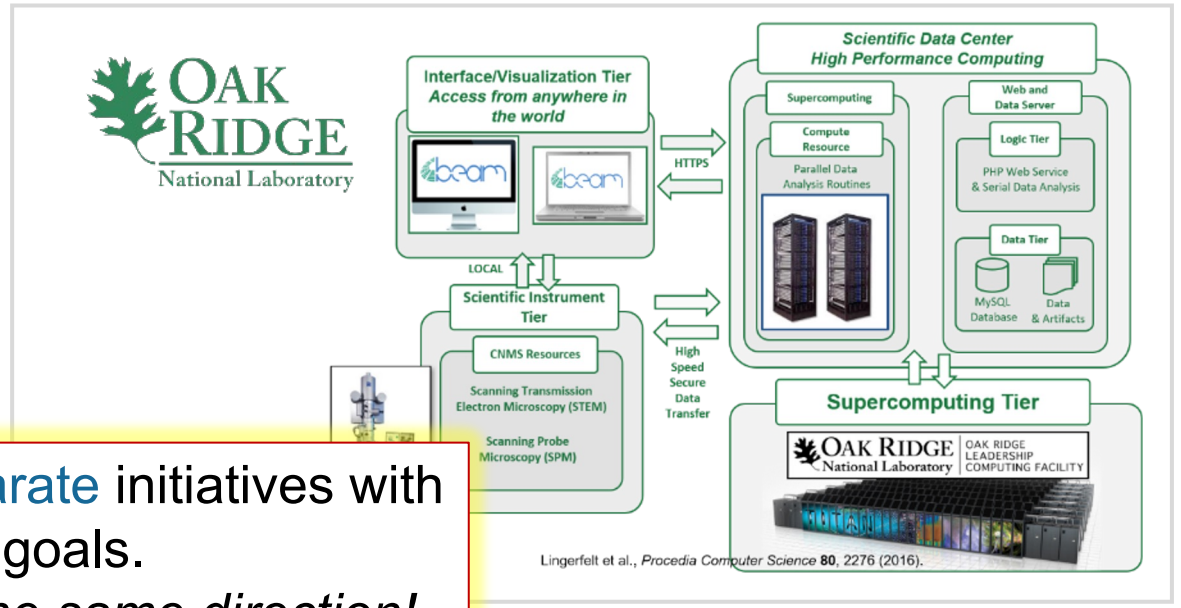
Flexibility…….	assembly of resource workflows is facile; complexity is concealed
Performance……	default behavior is performant, without arcane requirements
Scalability…….	data capabilities without excessive customizations
Transparency……	security, authentication, authorization should support automation
Interoperability….	services should extend outside the DOE environment
Resiliency…….	workloads are sustained across planned and unplanned events
Extensibility……..	designed to adapt and grow to meet unknown future needs
Engagement……	promotes co-design, cooperation, partnership
Cybersecurity…...	security for facilities and users is essential

Finding: Across DOE, innovators have been taking concerted steps towards integration through research, partnerships, and lab-level projects

LBNL's Superfacility project
ORNL's INTERSECT initiative
ANL's Nexus work
NERSC-LCLS LLANA software project
ECP ExaWorks & ExaFEL projects
BES DISCUS Light Source Data Working Group project
BES-ASCR CAMERA applied math center
BER joint EMSL-JGI FICUS joint-allocation program
… and more

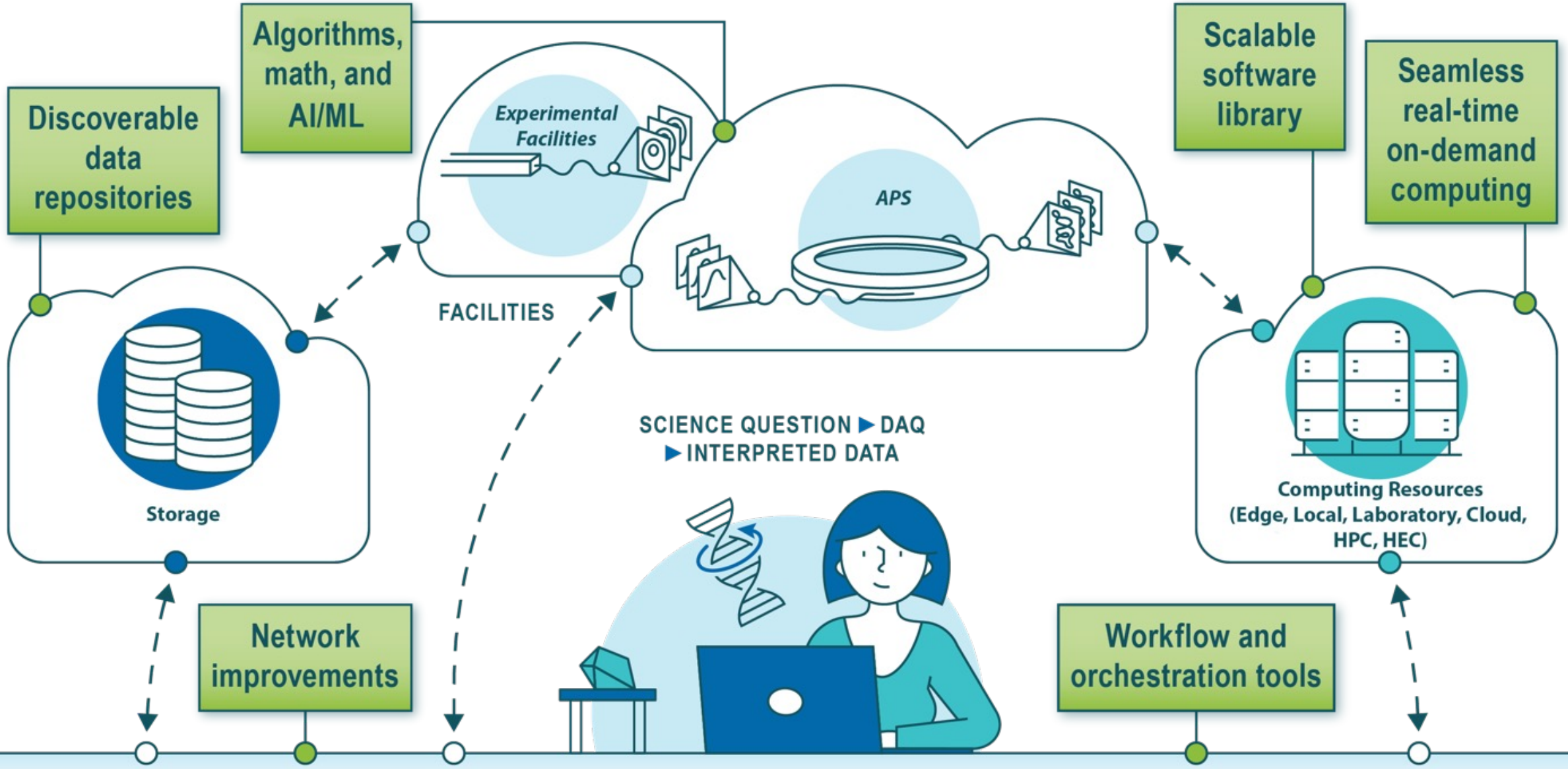




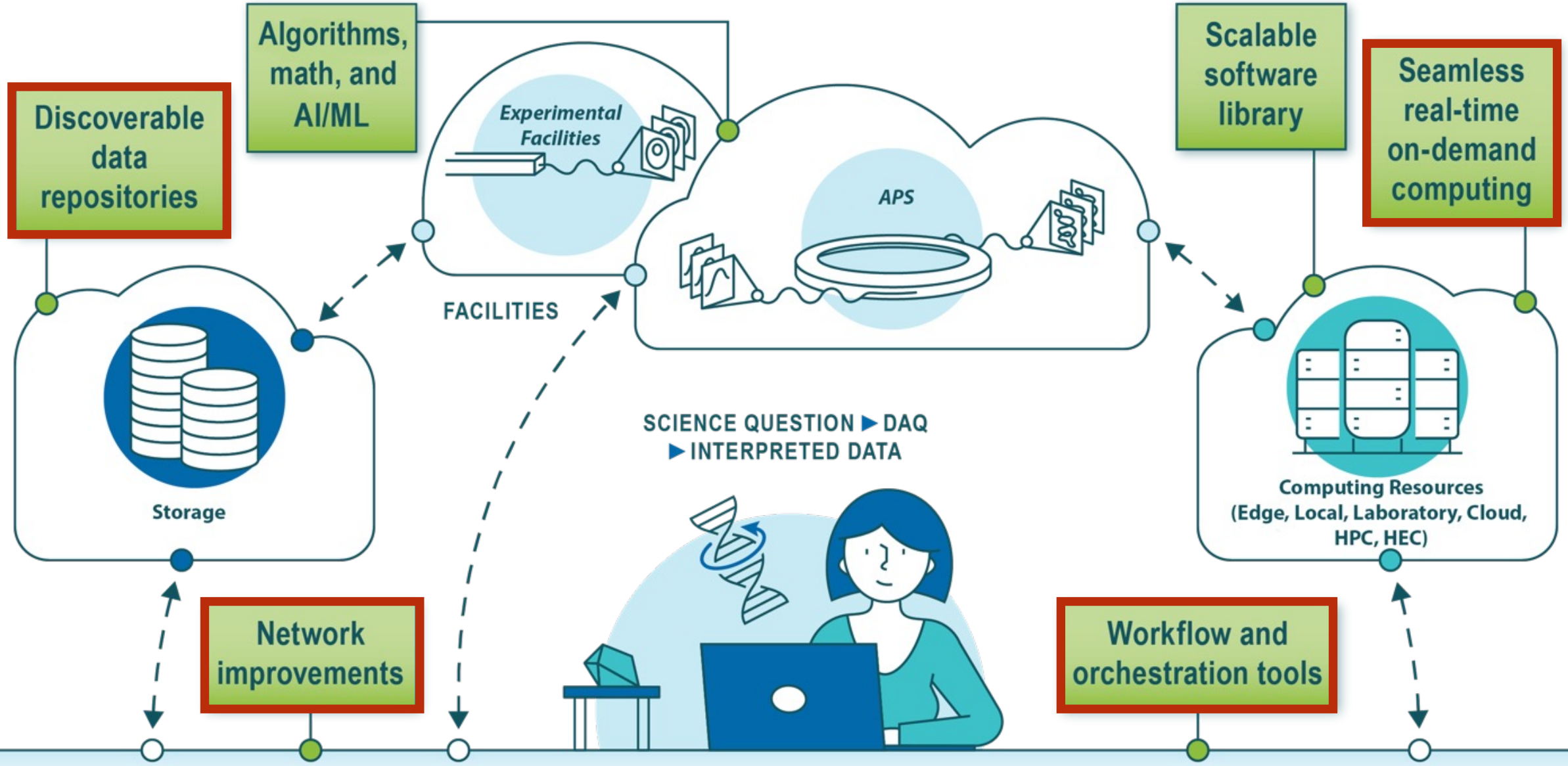


➢ These are all separate initiatives with similar integration goals.
Let's now row in the same direction!

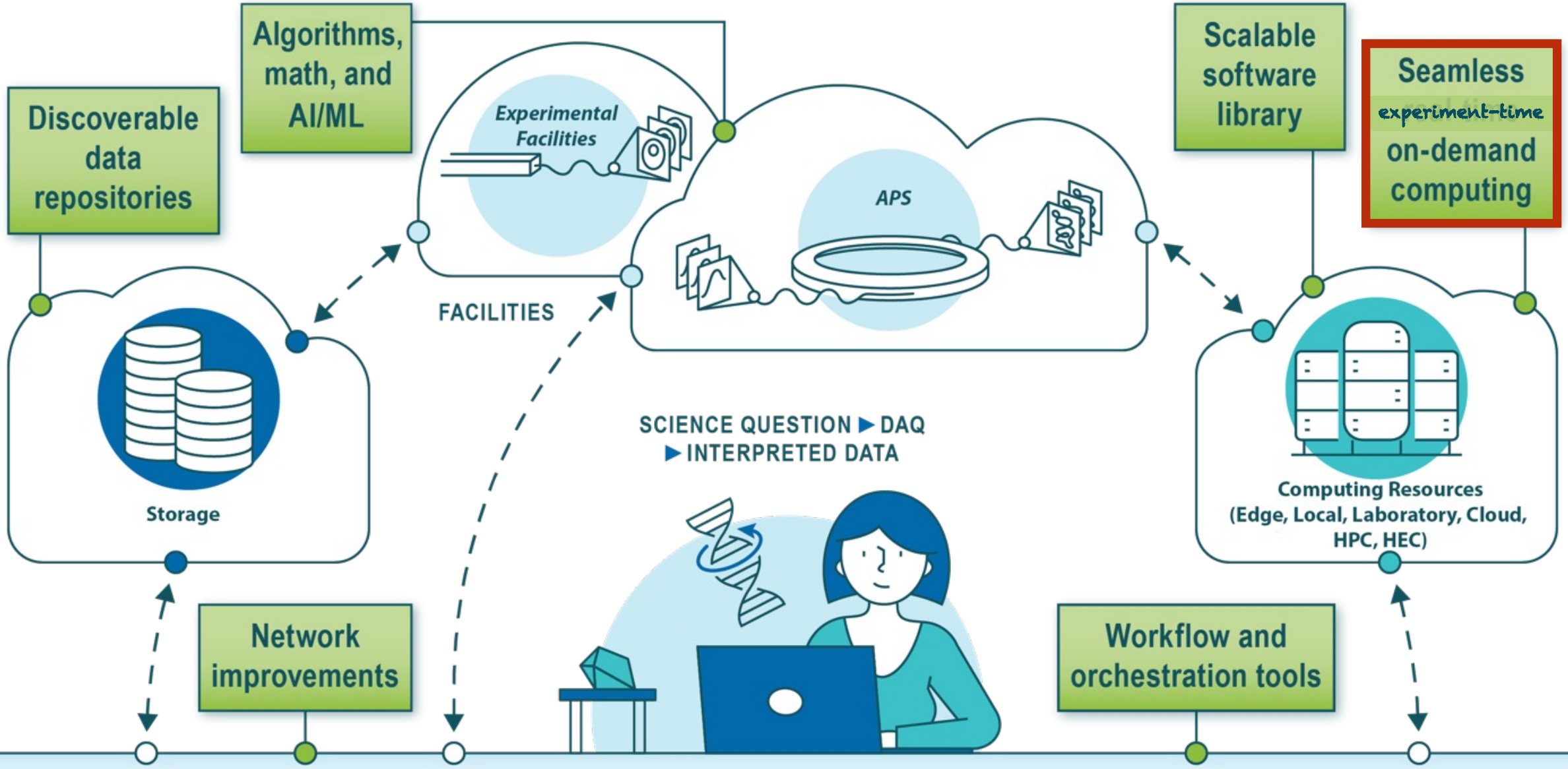
Integrated Research Infrastructure



Integrated Research Infrastructure



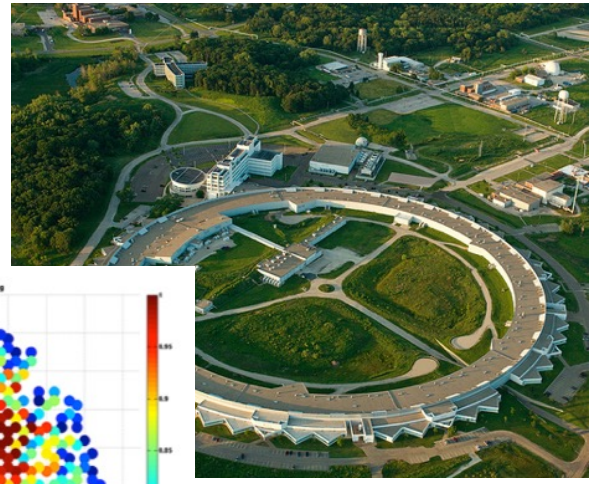
Integrated Research Infrastructure

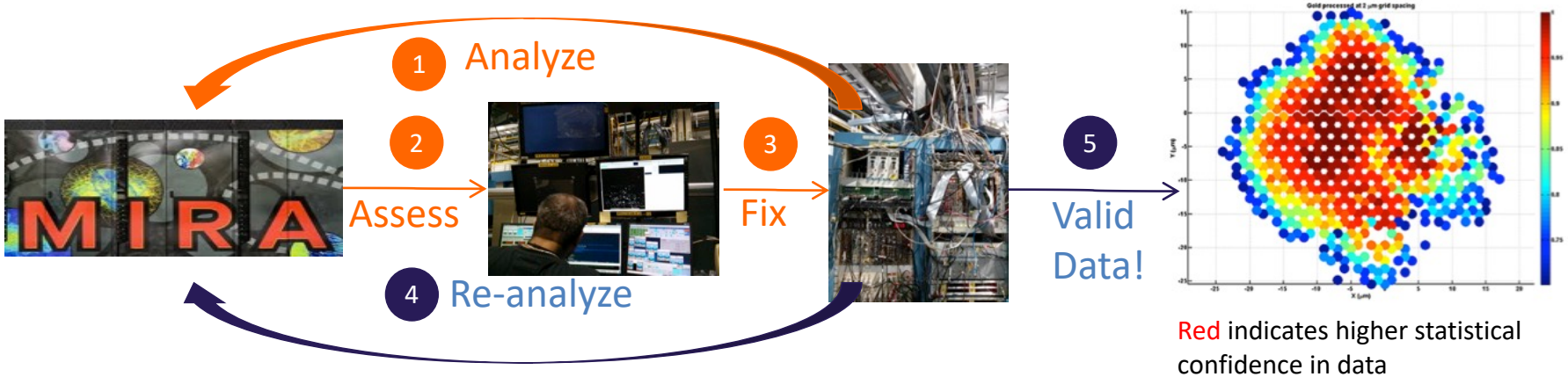


Argonne National Laboratory's **NEXUS**

Experiments Integrating Research Infrastructure

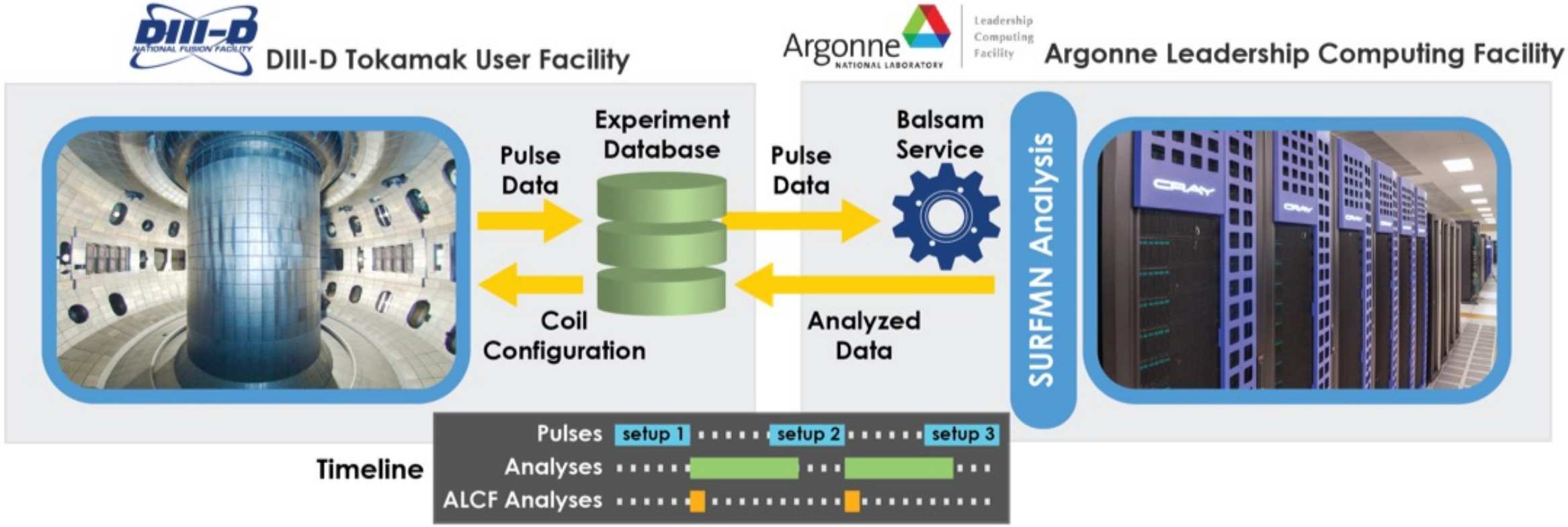
Impact	Accomplishments	Status
• APS scientists use Mira to process data from live HEDM experiments, providing real-time feedback to correct or improve in-progress experiments	• Real-time analysis of experimental steering • **Cable flaw was found and fixed at start of experiment,** saving an entire multi-day experiment and valuable user time and APS beam time.	• Workflow is established • Augmenting real-time scheduling





Red indicates higher statistical confidence in data

Experiments Integrating Research Infrastructure



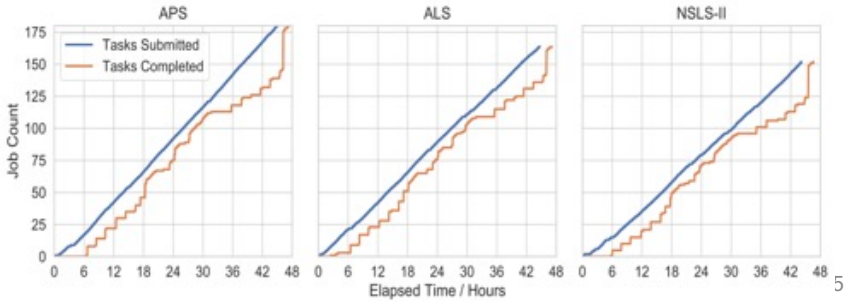
Mark Kostuk, Thomas D. Uram, Todd Evans, Dmitri M. Orlov, Michael E. Papka and David Schissel, *Automatic Between-Pulse Analysis of DIII-D Experimental Data Performed Remotely on a Supercomputer at Argonne Leadership Computing Facility*, **Fusion Science and Technology**, 74(2), 2018.

Experiments Integrating Research Infrastructure

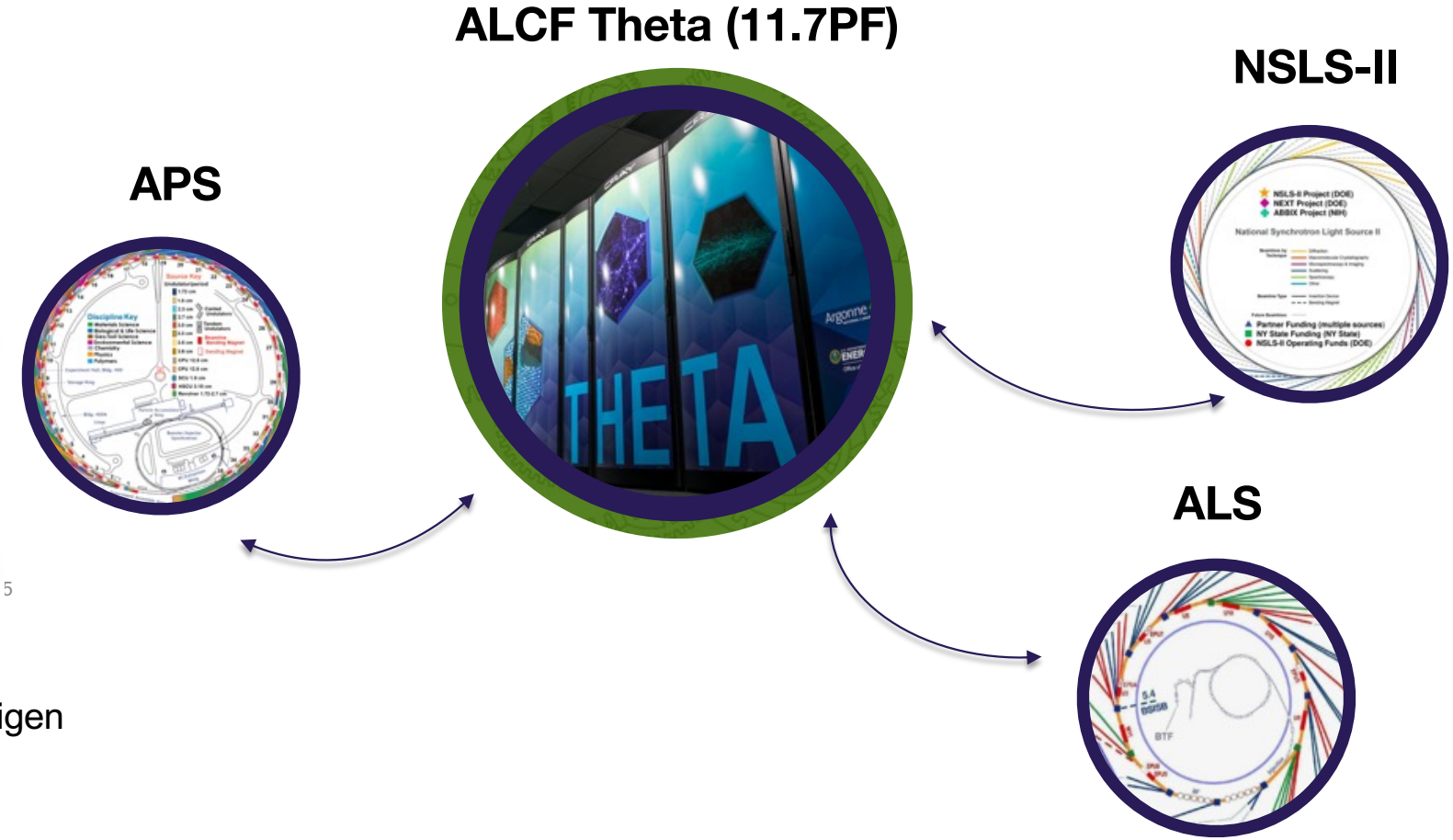
Utilizing leadership computing facility for continuous near real-time XPCS data processing

Balsam enables transparent access to remote resources

- Identity management
- Simplified scheduler API
- High-speed transfers via Globus



- Continuously executed for 48+ hours
- Process data in near real-time with XPCS-Eigen using special queue on Theta



Michael Salim, Tom Uram, J. Taylor Childers, Venkat Vishwanath, Michael E. Papka, *Balsam: Near Real-Time Experimental Data Analysis on Supercomputers*, **2019 IEEE/ACM 1st Annual Workshop on Large-scale Experiment-in-the-Loop Computing (XLOOP)**, 2019, pp. 26-31.

In late 2021, SC Leadership charged ASCR to devise and lead the Office of Science IRI Architecture Blueprint Activity

Devised, organized, and implemented the IRI ABA

HQ Executive Leadership

Ben Brown
Director
ASCR Facilities Division

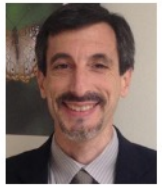
IRI-ABA Leadership Group

Debbie Bard
Group Lead for Data Science Engagement
NERSC, LBNL

Amber Boehnlein
Chief Information Officer
JLab

Kjiersten Fagnan
Chief Informatics Officer
JGI, LBNL

Chin Guok
Group Lead for Planning and Architecture
ESnet, LBNL



Bill Miller
Senior Technical Advisor
ASCR Facilities Division

Eric Lancon
Director, Scientific Data and Computing Center
BNL

Jini Ramprakash
Deputy Division Director
ALCF, ANL

Arjun Shankar
Section Head, Advanced Technologies
OLCF/NCCS, ORNL

Nicholas Schwarz
Group Leader, Scientific Software Eng. & Data Mgmt.,
APS, ANL

BER	Paul Bayer, Jay Hnilo, Resham Kulkarni
BES	Tom Russell
FES	Josh King, Matt Lanctot
HEP	Jeremy Love, Eric Church
IP	Kristian Myhre
NP	Xiaofeng Guo, Jim Sowinski

- *Exchanged on urgent IRI needs, priorities, & commonalities across programs.*
- *Engaged and provided feedback at key points on Activity progress and outputs.*

Convened over **150 DOE national laboratory experts** from **all 28 SC user facilities** across **13 national laboratories** to consider the **technological, policy, and sociological challenges** to implementing IRI.

IRI Blueprint Activity Key Results

We now possess a reference framework to inform a coordinated, SC-wide strategy for IRI.

The key organizing elements of the IRI Framework are Science Patterns and Practice Areas:

> **IRI Science Patterns** that represent integrated science use cases across DOE science domains and

> **IRI Practice Areas** that will support the realization of a DOE-integrated IRI ecosystem.

The IRI Framework: Science Patterns

IRI Science Patterns are broad classes of integrated research workflows with common driving features. Each Science Pattern represents a spectrum of DOE science domains and will benefit from a strategic and coordinated approach to design and solution. A given workflow case may span several Science Patterns.

Time-sensitive pattern has *urgency*, requiring real-time or end-to-end performance with high reliability, e.g., for timely decision-making, experiment steering, and virtual proximity.

Data integration-intensive pattern requires combining and analyzing data from multiple sources, e.g., sites, experiments, and/or computational runs.

Long-term campaign pattern requires sustained access to resources over a long period to accomplish a well-defined objective.

The IRI Framework: Practice Areas

IRI Practice Areas are cross-cutting communities of practice whose efforts will be essential to advance robust and extensible IRI designs and solutions.

User experience practice will ensure relentless attention to user perspectives and needs through requirements gathering, user-centric (co)-design, continuous feedback, and other means.

Resource co-operations practice is focused on creating new modes of cooperation, collaboration, co-scheduling, and joint planning across facilities and DOE programs.

Cybersecurity and federated access practice is focused on creating novel solutions that enable seamless scientific collaboration within a secure and trusted IRI ecosystem.

Workflows, interfaces, and automation practice is focused on creating novel solutions that facilitate the dynamic assembly of components across facilities into end-to-end IRI pipelines.

Scientific data life cycle practice is focused on ensuring that users can manage their data and metadata across facilities from inception to curation, archiving, dissemination, and publication.

Portable/scalable solutions practice is focused on ensuring that transitions can be made across heterogeneous facilities (portability) and from smaller to larger resources (scalability).

IRI in the FY 2024 President's Budget Request

ASCR Facilities

"In FY 2024, the ASCR facilities will continue planning and begin implementation to advance DOE's Integrated Research Infrastructure (IRI) so that researchers can seamlessly and securely meld DOE's unique data, user facilities, and computing resources to accelerate discovery and innovation."

High Performance Data Facility (HPDF) project

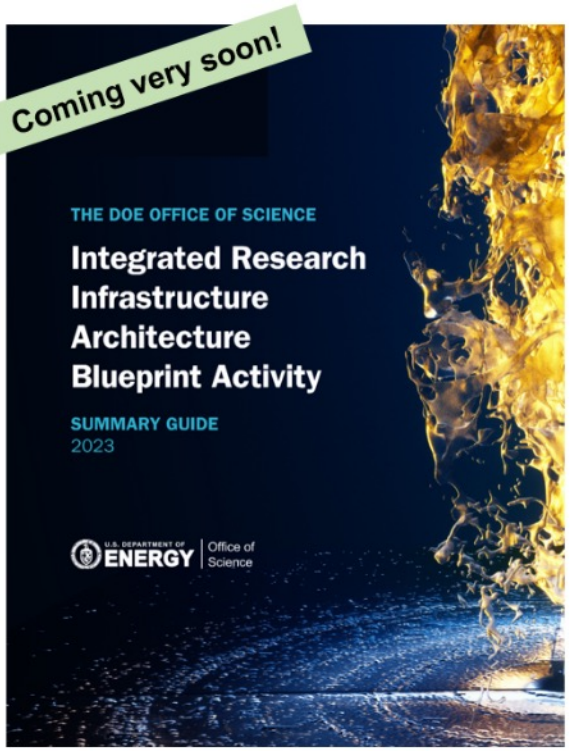
"The proposed HPDF will serve as a foundational element in enabling the DOE Integrated Research Infrastructure; will provide crucial resources to Office of Science programs to attack fundamental problems in science and engineering that require nimble shared access to large data sets, increasingly aggregated from multiple sources; will partner and operate in concert with other ASCR Facilities and potentially other DOE laboratory computing resource providers to provide a high availability high performance computing ecosystem for a wide variety of applications; will serve as a 'Hub' enabling 'Spoke' sites to deploy and orchestrate distributed infrastructure to enable high priority DOE mission applications."

See DOE Lab Funding Announcement LAB 23-3020 for more information.

https://science.osti.gov/grants/Lab-Announcements/Open

IRI Look Ahead for 2023

- Release of **IRI Blueprint Activity** final report
- Release of **ESnet Requirements Reviews IRI meta-analysi**
- Release of the ASCR Facilities' **IRI Testbed** whitepaper
- **High Performance Data Facility project**
- **What's next**
  - Early IRI partnerships: identifying and forging
  - Steering/governance of the IRI Program: growing into our shoes
  - IRI Testbed: describing what it is and how to engage




Thank You