DOE's Integrated Research Infrastructure

Biological and Environmental Research Advisory Committee Meeting
April 21, 2023

Ben Brown

Director, Facilities Division

Advanced Scientific Computing Research (ASCR)

My goal for this presentation is to provide context for DOE's Integrated Research Infrastructure (IRI) effort

- The development of the IRI Vision

- The IRI Framework (developed through the 2022 IRI Blueprint Activity)

- 2023 Look ahead

ASCR looks forward to engaging with BERAC and BER in any way that would be helpful.

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

Oak Ridge Leadership Computing Facility
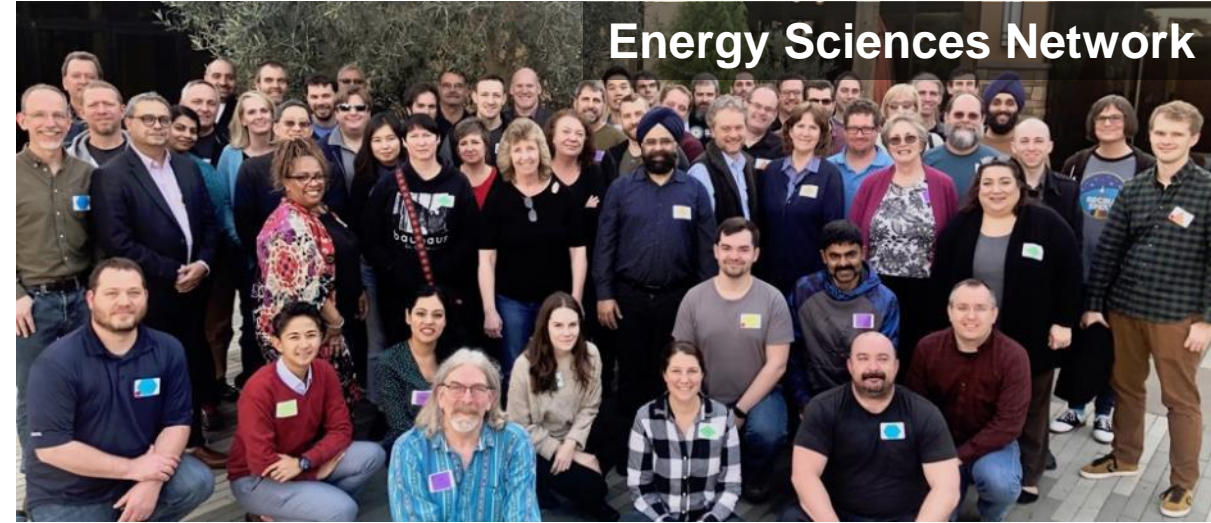
Energy Sciences Network
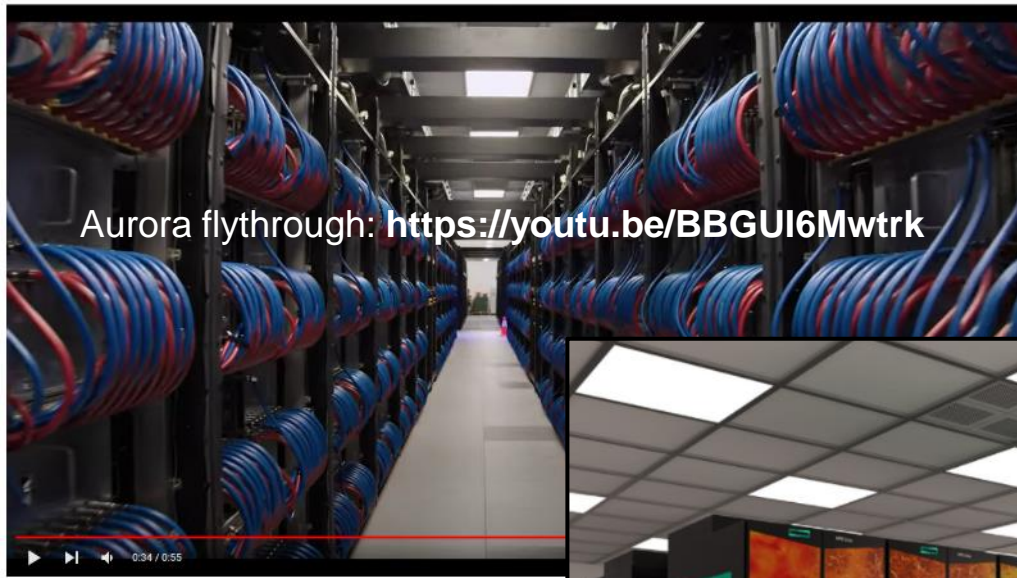
National Energy Sciences Supercomputing Center

Argonne Leadership Computing Facility

ASCR Facilities today:
Exascale systems!
Perlmutter!
ESnet6!

Aurora at ANL

Aurora flythrough: **https://youtu.be/BBGUI6Mwtrk**

Power/cooling/space: **https://youtu.be/HmQjMEPhfDM**

The promise of exascale: **https://vimeo.com/720503659**

Frontier at ORNL

Perlmutter at LBNL

ESnet6, managed by LBNL

ESnet's role: **https://vimeo.com/757334469**

U.S. DEPARTMENT OF **ENERGY** | Office of Science
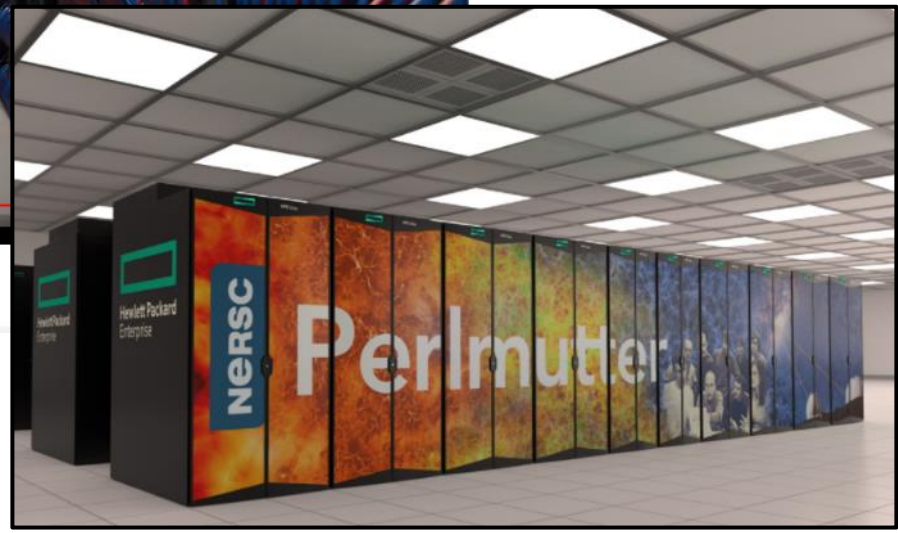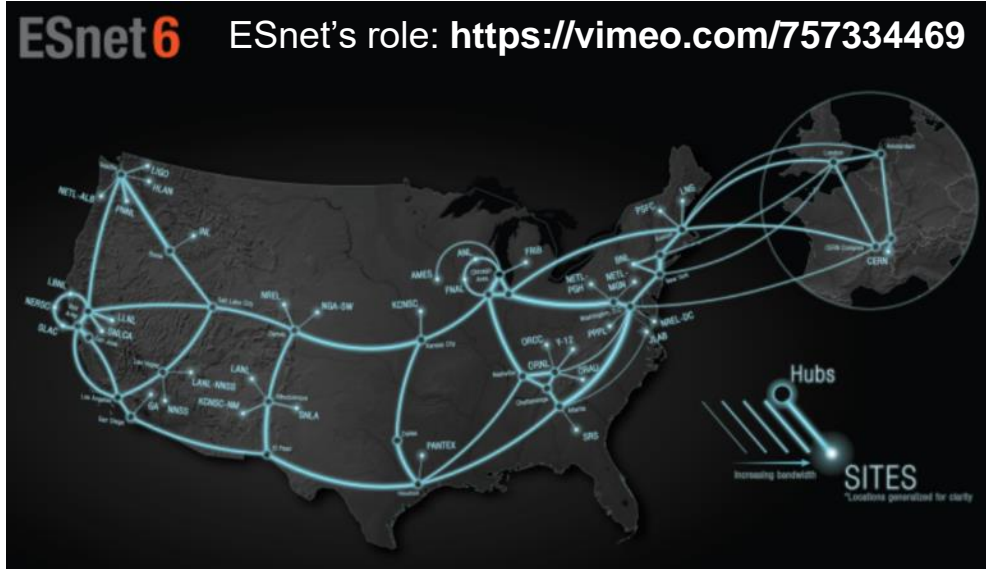
DOE is positioned to lead the new era of integrated science within the US Government and the world.

Linking distributed resources is becoming paramount to modern collaborative science.

The challenges of our time call upon DOE and its national laboratories to be an open innovation ecosystem:

Accelerating discovery & innovation

Democratizing access

Drawing new talent

Advancing open science

An IRI case: Pharmaceutical research for COVID vaccines

This artist's rendering depicts x-ray crystallography at SLAC's Linac Coherent Light Source. LCLS partnered with NERSC and ESnet to perform real-time image analysis for research of the SARS-CoV-2 virus structure.
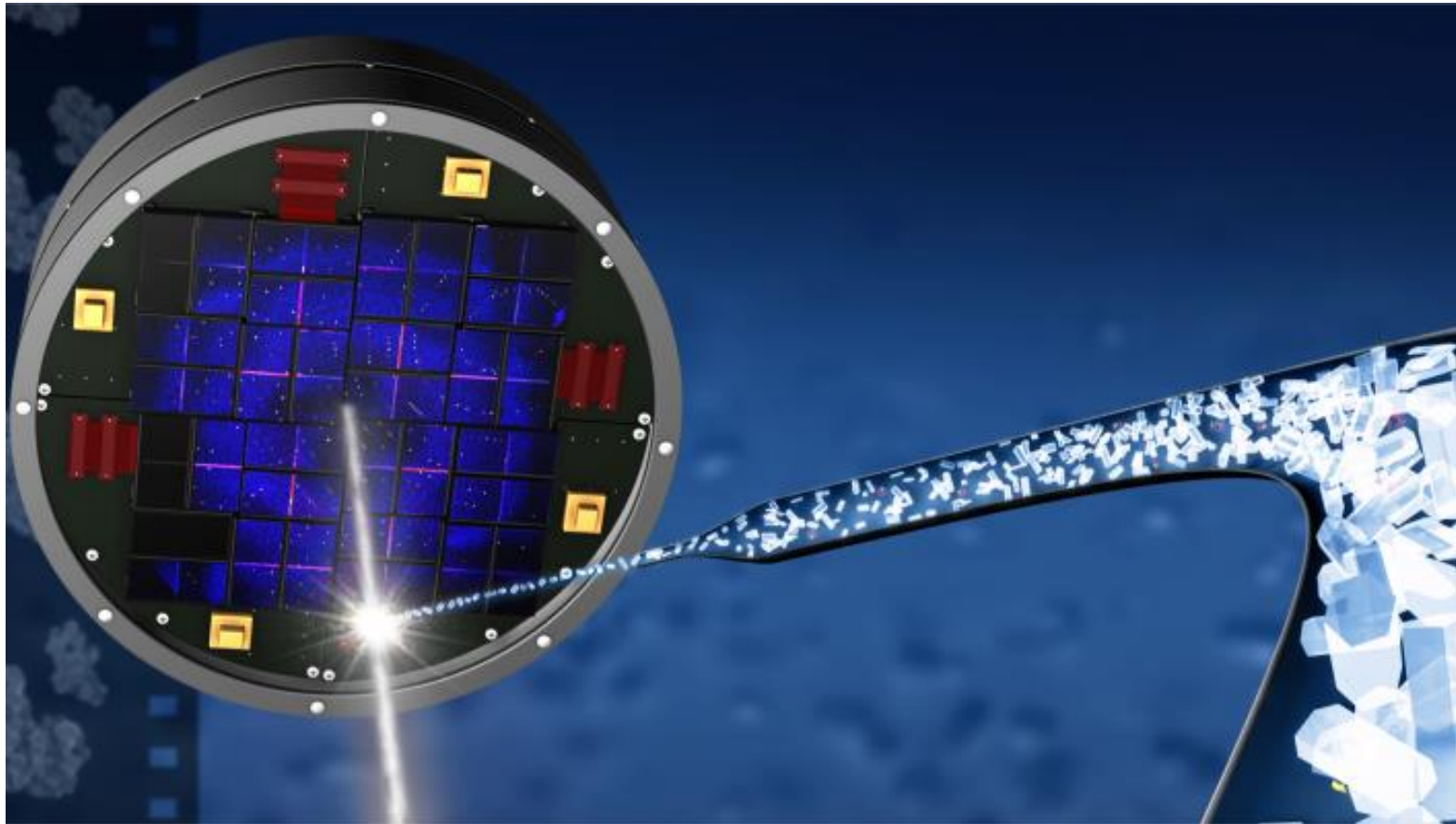
Linac Coherent Light Source at SLAC

NERSC at LBNL

ESnet, managed by LBNL

SLAC National Accelerator Laboratory

U.S. DEPARTMENT OF ENERGY | Office of Science

DOE's Integrated Research Infrastructure (IRI) Vision:
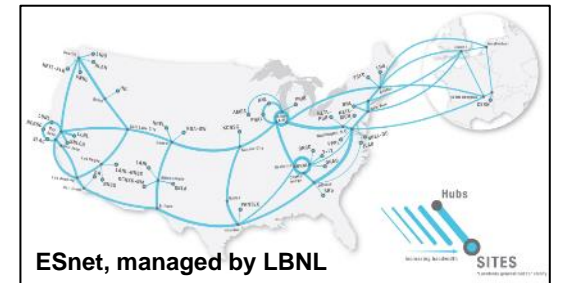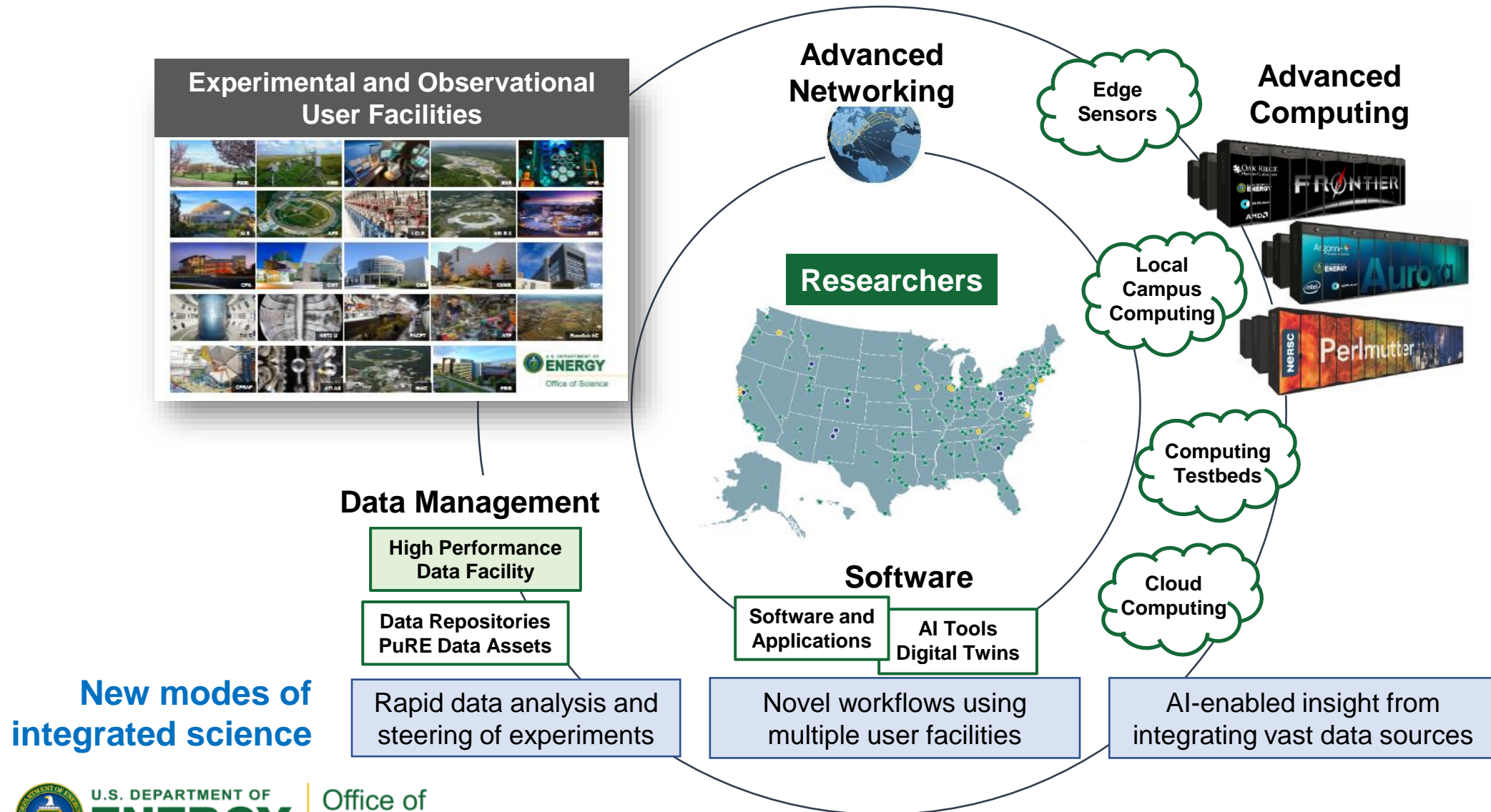*To empower researchers to meld DOE's world-class research tools, infrastructure, and user facilities seamlessly and securely in novel ways to radically accelerate discovery and innovation*

Experimental and Observational User Facilities

Advanced Networking

Edge Sensors

Advanced Computing

Researchers

Local Campus Computing

Computing Testbeds

Data Management

High Performance Data Facility

Data Repositories PuRE Data Assets

Software

Software and Applications

AI Tools Digital Twins

Cloud Computing

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

U.S. DEPARTMENT OF **ENERGY** | Office of Science
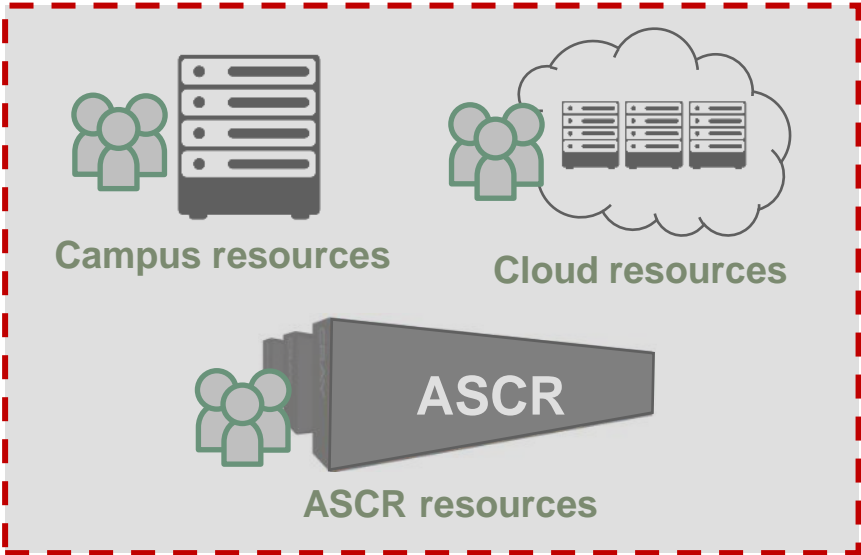
BER is a longstanding leader in integrated science!

Technologist Perspective

Researcher Perspective

photo credit: Didrik Johnck

https://www.flickr.com/photos/deetrak/623632892

U.S. DEPARTMENT OF ENERGY | Office of Science
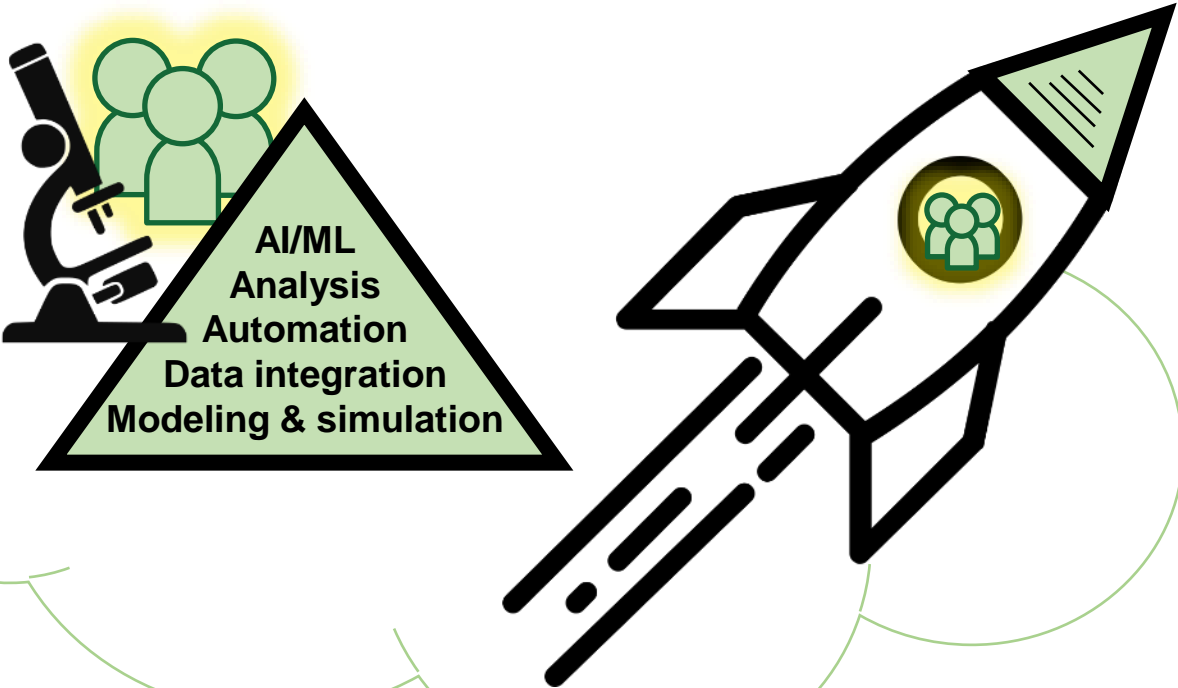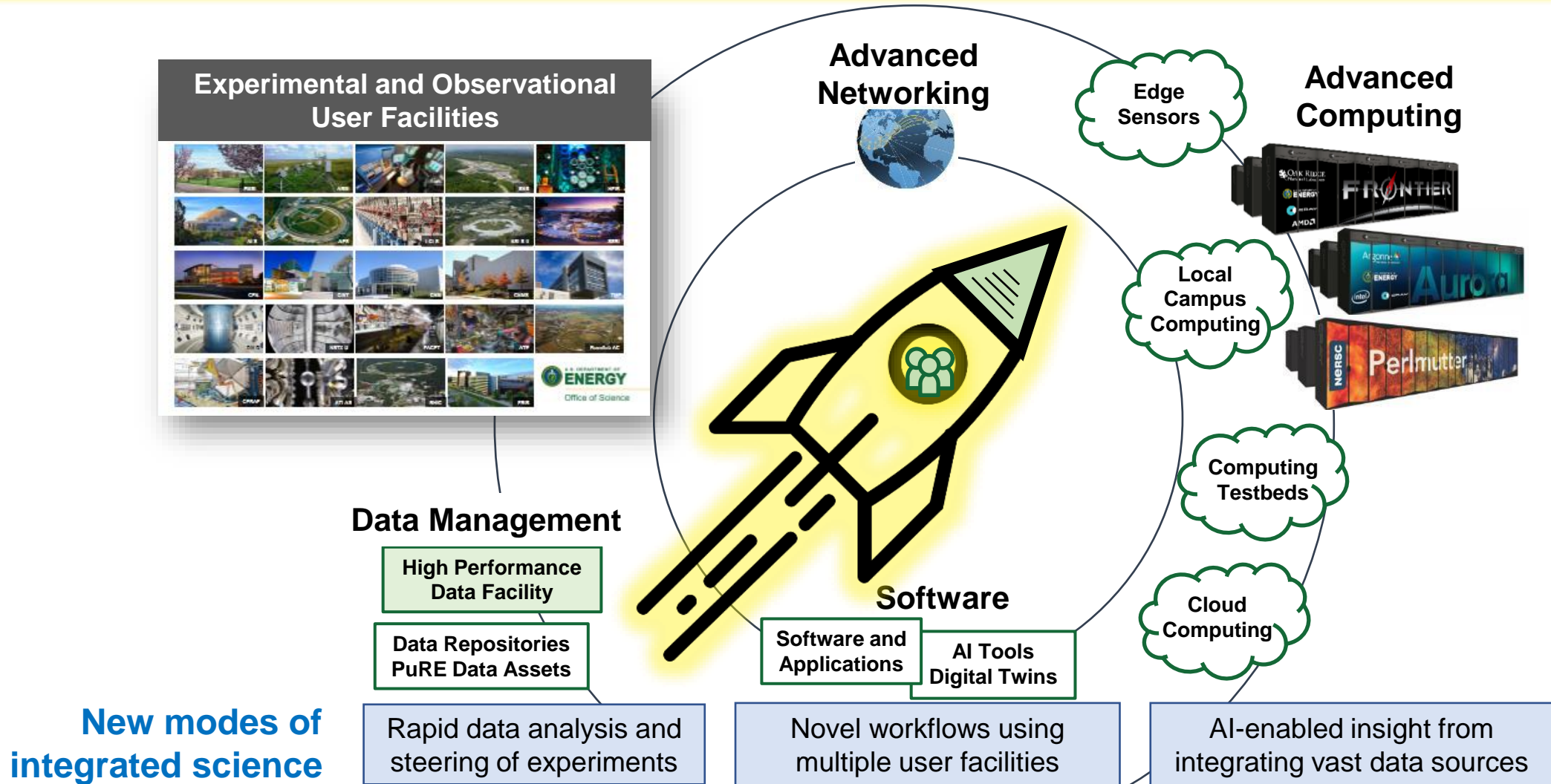
From the researcher view, "integration" – melding specialized infrastructure and resources – is about risk and reward.

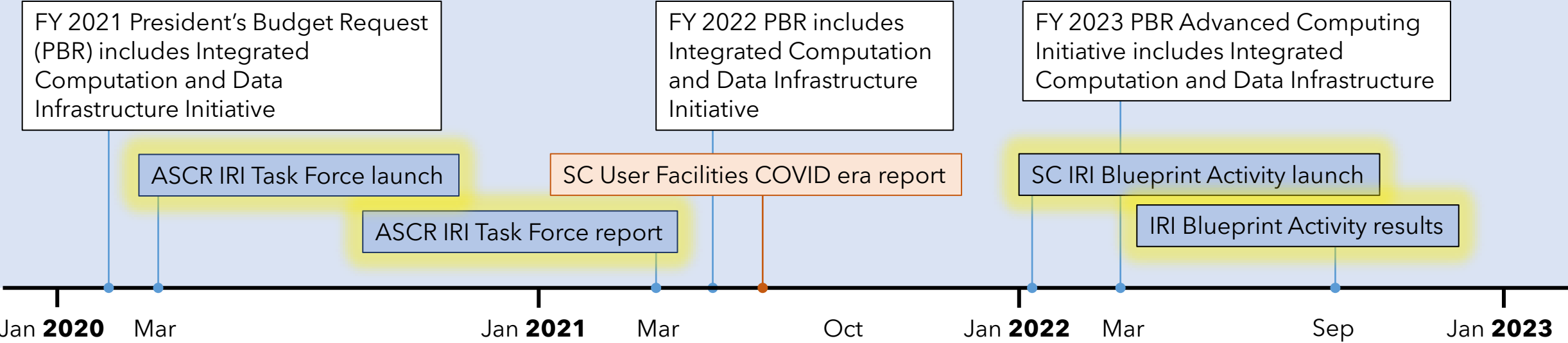**Access to computing and data resources could accelerate our research!**

AI/ML
Analysis
Automation
Data integration
Modeling & simulation

Campus resources

Cloud resources

ASCR

ASCR resources

U.S. DEPARTMENT OF **ENERGY** | Office of Science

"Simple and powerful" is the mantra: researchers will benefit from an operational environment that is intuitive and simple to use yet extraordinarily powerful in accelerating discovery.

Experimental and Observational User Facilities

Advanced Networking

Edge Sensors

Advanced Computing

Local Campus Computing

Computing Testbeds

Data Management

High Performance Data Facility

Data Repositories PuRE Data Assets

Software

Software and Applications

AI Tools Digital Twins

Cloud Computing

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

FY 2022 PBR includes Integrated Computation and Data Infrastructure Initiative

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

ASCR IRI Task Force launch

SC User Facilities COVID era report

SC IRI Blueprint Activity launch

ASCR IRI Task Force report

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

Oct **2021**

NATIONAL STRATEGIC OVERVIEW FOR RESEARCH AND DEVELOPMENT INFRASTRUCTURE

A Report by the
SUBCOMMITTEE ON RESEARCH AND DEVELOPMENT INFRASTRUCTURE
COMMITTEE ON SCIENCE AND TECHNOLOGY ENTERPRISE

of the
NATIONAL SCIENCE AND TECHNOLOGY COUNCIL

May **2022**

FUTURE ADVANCED COMPUTING ECOSYSTEM STRATEGIC PLAN FY2022 IMPLEMENTATION ROADMAP

A report by the
Subcommittee on Future Advanced Computing Ecosystem
Committee on Technology
and the
High End Computing Interagency Working Group
Networking & Information Technology Subcommittee
Committee on Science & Technology Enterprise
of the
National Science and Technology Council

U.S. DEPARTMENT OF ENERGY | Office of Science

11

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 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 integrated computational and data research infrastructure 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 workflow

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

U.S. DEPARTMENT OF ENERGY | Office of Science

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

Bill Miller
Senior Technical Advisor
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

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

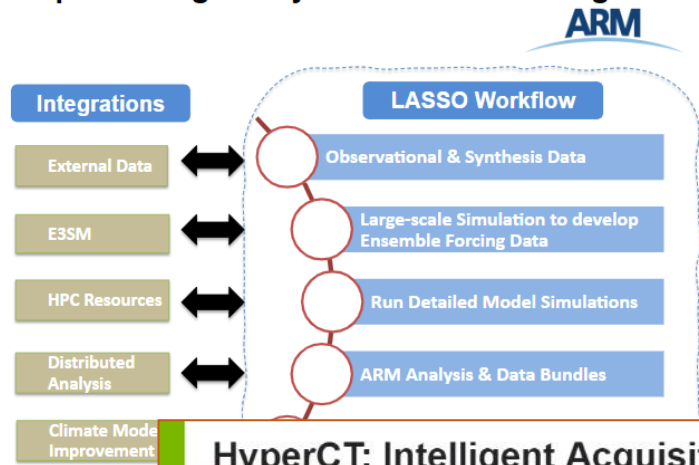|	
BER	Paul Bayer, Jay Hnilo, Resham Kulkarni
BES	Tom Russell
FES	Josh King, Matt Lanctot
HEP	Jeremy Love, Eric Church
IP	Kristian Myhre
NP	Xiaofeng Guo, Jim Sowinski

- *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.
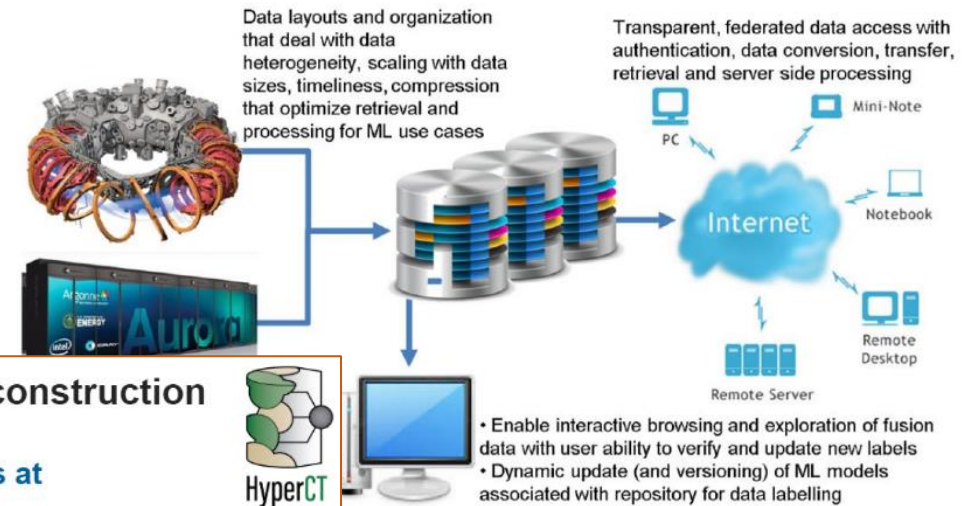
# Finding: SC Programs are grappling with integration in many venues

Courtesy Paul Bayer, BER

Courtesy Matt Lanctot, FES

Courtesy Tom Russell, BES

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 ALCF-APS Balsam software project
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

Lingerfelt et al., *Procedia Computer Science* 80, 2276 (2016).
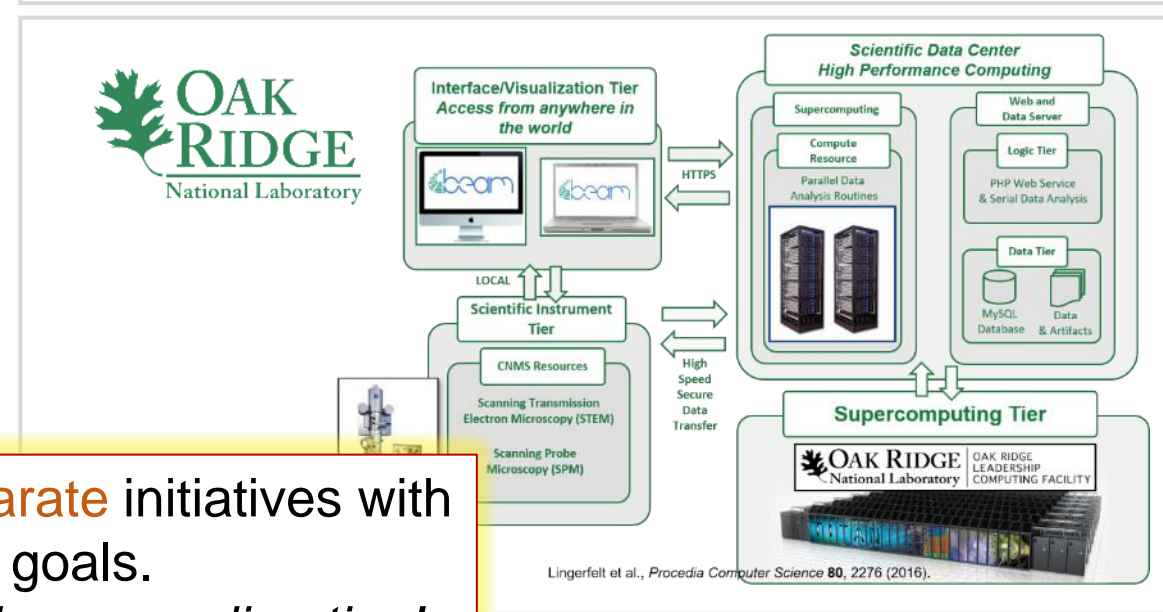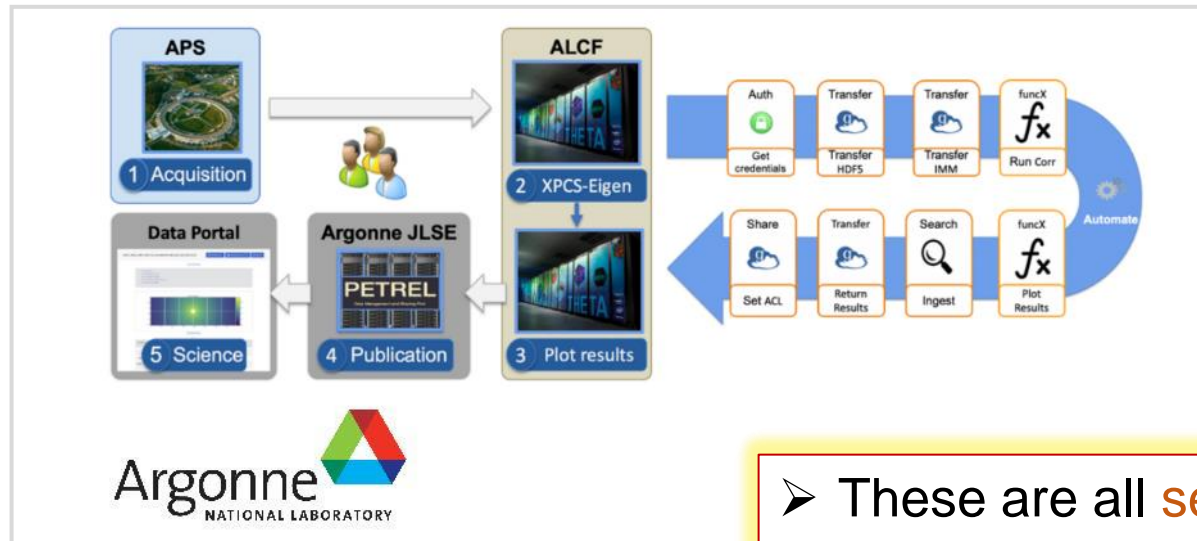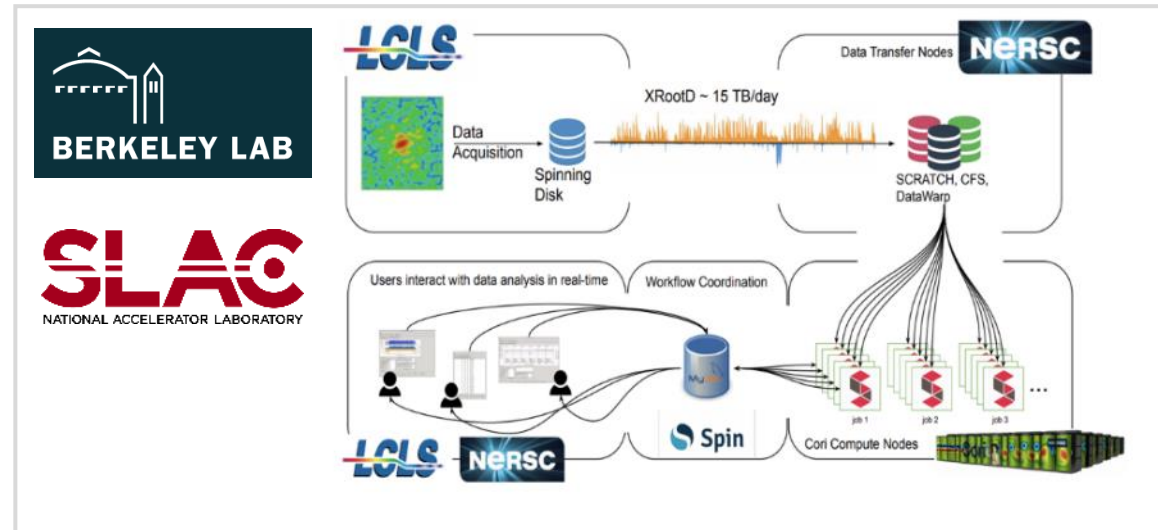
> ➤ These are all separate initiatives with similar integration goals.
> *Let's now row in the same direction!*

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.

Coming very soon!

THE DOE OFFICE OF SCIENCE

Integrated Research Infrastructure Architecture Blueprint Activity

SUMMARY GUIDE
2023

U.S. DEPARTMENT OF ENERGY | Office of Science
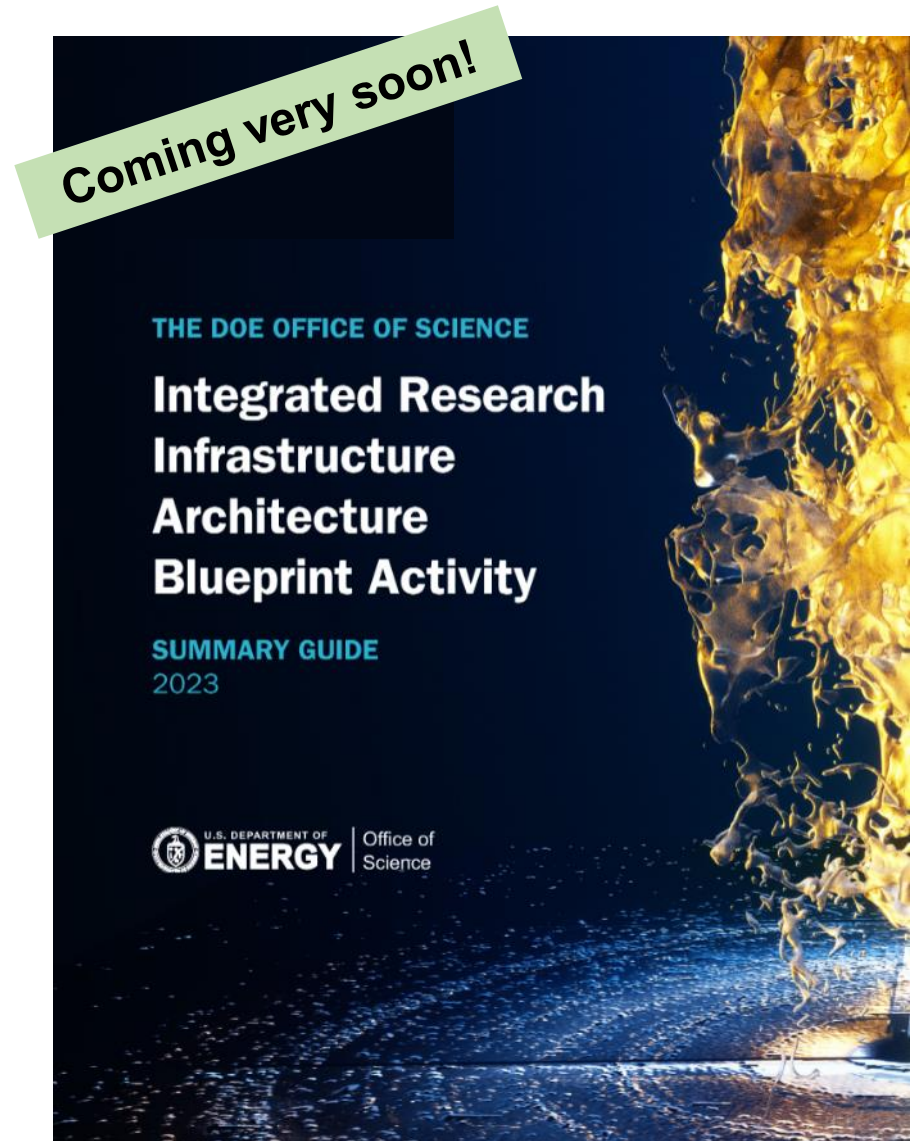
The IRI Framework: Science Patterns

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.

U.S. DEPARTMENT OF **ENERGY** | Office of Science

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).

U.S. DEPARTMENT OF **ENERGY** | Office of Science

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

U.S. DEPARTMENT OF ENERGY | Office of Science

IRI Look Ahead for 2023

- Release of **IRI Blueprint Activity** final report
- Release of **ESnet Requirements Reviews IRI meta-analysis**
- Release of the ASCR Facilities' **IRI Testbed** whitepaper
- **High Performance Data Facility project**
- **Convening event(s)**
  - IRI Testbed: describing what it is and how to engage
  - Early IRI partnerships: identifying and forging
  - Steering/governance of the IRI Program: growing into our shoes
  - Authentication/Authorization exploratory activity

Coming very soon!

THE DOE OFFICE OF SCIENCE

Integrated Research
Infrastructure
Architecture
Blueprint Activity

SUMMARY GUIDE
2023

U.S. DEPARTMENT OF ENERGY | Office of Science

U.S. DEPARTMENT OF ENERGY | Office of Science