

OCTOBER 12, 2024

From Data to Discovery: Integration of HPC and Experimental Science

MICHAEL E. PAPKA

Director, Argonne Leadership Computing Facility, Argonne National Laboratory

Deputy Associate Laboratory Director, Computing, Environment and Life Sciences, Argonne National Laboratory

Co-Director, George Crabtree Institute for Discovery and Sustainability, UIC/Argonne

Warren S. McCulloch Professor of Computer Science, University of Illinois Chicago

Nobel Prize in Chemistry 2024 – David Baker

The New York Times

2024 Nobel Prizes | What to Know | Prize in Medicine | Prize in Physics | Prize in Chemistry | Other Science Prizes

Nobel Prize in Chemistry Goes to 3 Scientists for Predicting and Creating Proteins

The Nobel, awarded to David Baker of the University of Washington and Demis Hassabis and John M. Jumper of Google DeepMind, is the second this week to involve artificial intelligence.

Share full article | Print | Email | Comment 11

Chemistry Nobel Prize Recognizes Three Protein Researchers

Argonne Leadership Computing Facility

ALCF Resources | Science & Engineering | Community and Outreach | About | Support Center

HOME / NEWS CENTER / A RISING PEPTIDE: SUPERCOMPUTING HELPS SCIENTISTS COME CLOSER TO TAILORING DRUG MOLECULES

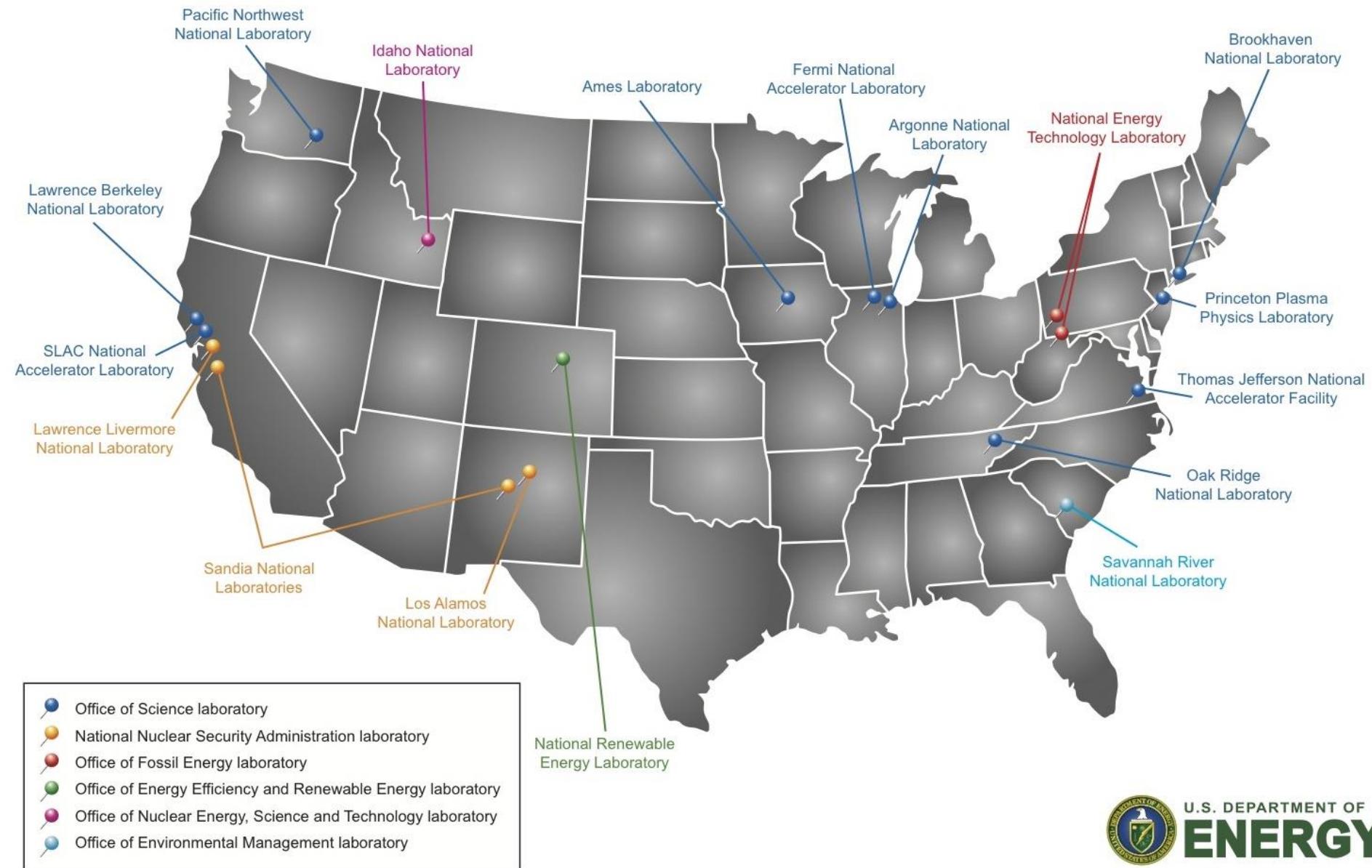
SCIENCE

A rising peptide: Supercomputing helps scientists come closer to tailoring drug molecules

AUTHOR ROBERT GRANT
PUBLISHED 01/23/2017
AWARD INCITE

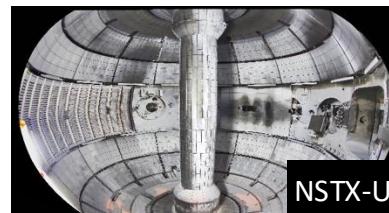
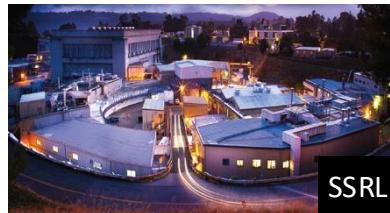
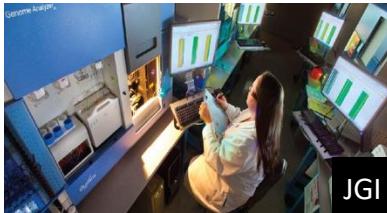
A team of researchers led by biophysicists at the University of Washington have come one step closer to designing tailor-made drug molecules that are more precise and carry fewer side effects than most existing therapeutic compounds. With the help of the Mira supercomputer, located at the Argonne Leadership Computing Facility at the U.S. Department of Energy's (DOE) Argonne National Laboratory, the scientists have successfully designed and verified stable versions of synthetic peptides, components that join together to form proteins. They published their work in a recent issue of Nature. The computational protocol, which was validated by assembling physical peptides in the chemistry lab and comparing them to the computer models, may one day enable drug developers to craft novel, therapeutic peptides that precisely target specific disease-causing molecules within the body. And the insights the researchers gleaned constitute a significant advance in the fundamental understanding of protein folding.

Department of Energy National Laboratories



Department of Energy User Facilities

FY 2024
28 scientific
user facilities
>39,500 users



U.S. DEPARTMENT OF
ENERGY

Office of
Science

Department of Energy 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

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**

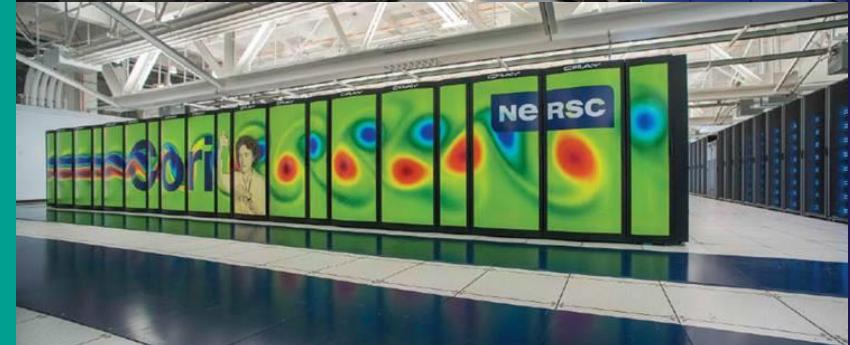
Argonne
Leadership
Computing
Facility
(ALCF)



Oak Ridge
Leadership
Computing
Facility
(OLCF)



National Energy
Research Scientific
Computing Center
(NERSC)



Energy Sciences
Network (ESnet)





MAN ACHIEVED HERE
THE FIRST SELF-SUSTAINING CHAIN REACTION
AND THEREBY INITIATED THE
CONTROLLED RELEASE OF NUCLEAR ENERGY

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.

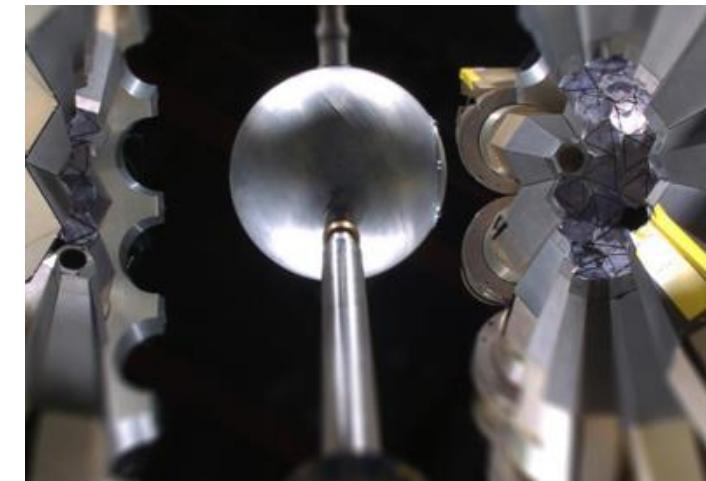
Argonne National Laboratory DOE User Facilities



Advanced Photon Source



Argonne Leadership Computing Facility

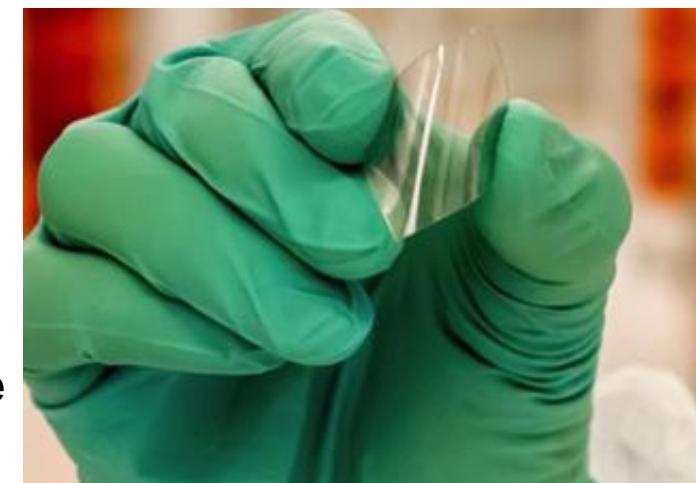


Argonne Tandem Linear Accelerator System

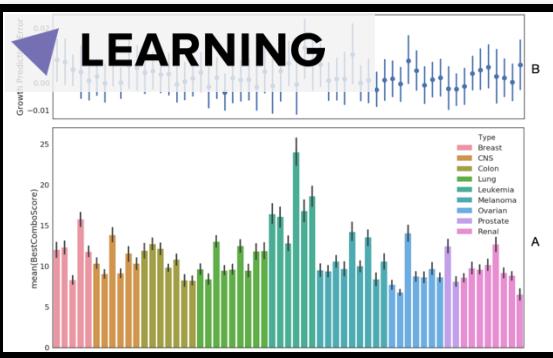
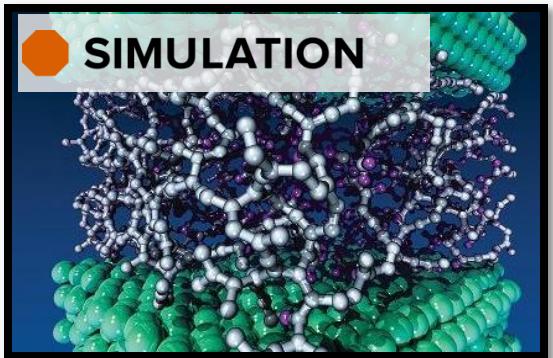


Atmospheric Radiation Measurement – The Southern Great Plains

Center for Nanoscale Materials



Argonne Leadership Computing Facility



ALCF offers different pipelines based on your computational readiness



2024 ALCF Annual Report

The ALCF provides world-class computing resources to the scientific community

- Users pursue scientific challenges
- In-house experts to help maximize results
- Resources fully dedicated to open science

ALCF AI Testbeds



ALCF at a Glance in 2023

- Users pursue scientific challenges
- In-house experts to help maximize results
- Resources fully dedicated to open science

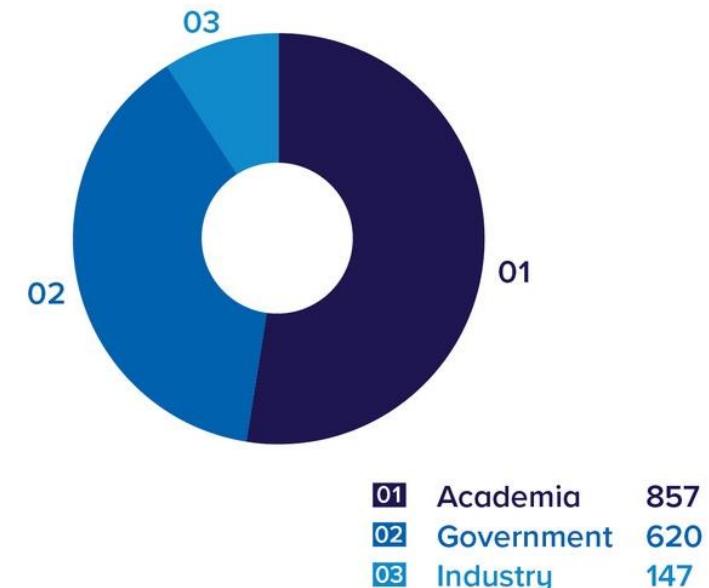
35.7M node-hours of compute time

417 active projects

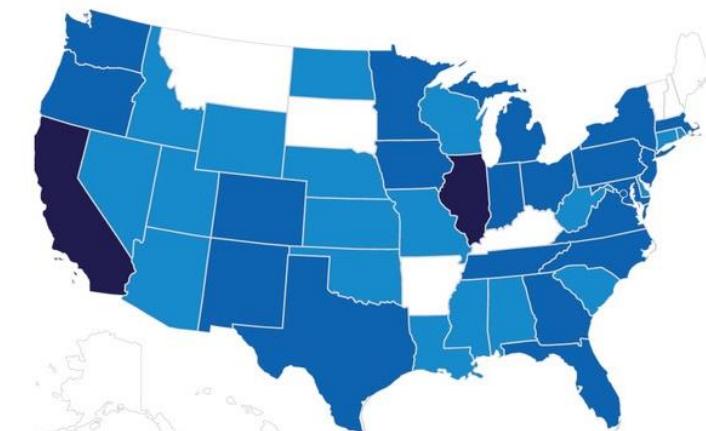
1,624 facility users

230+ publications

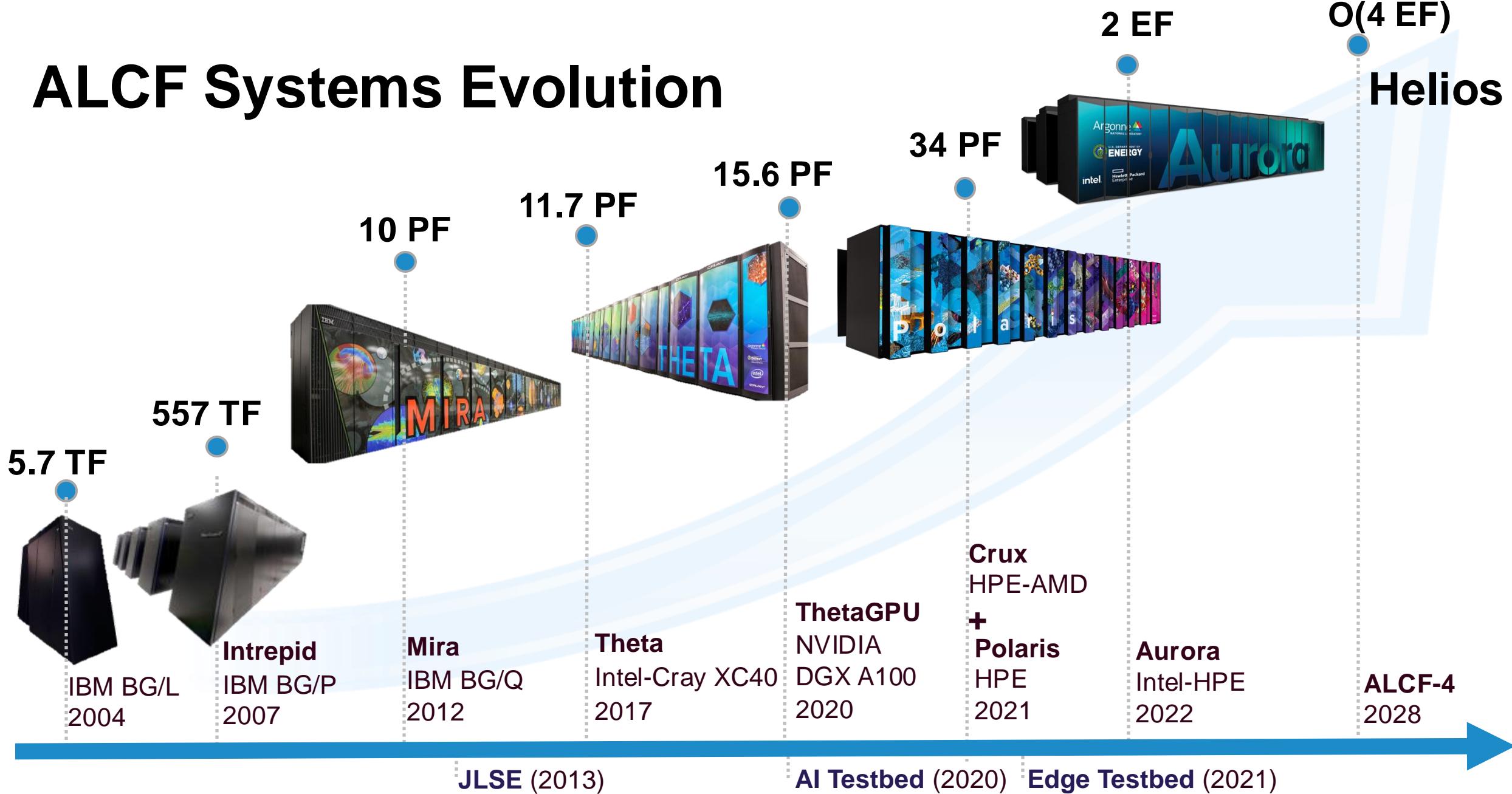
2023 ALCF Users by Affiliation



2023 U.S. ALCF Users by State



ALCF Systems Evolution



Aurora Specifications

Compute

21,248
CPUs

63,744
GPUs

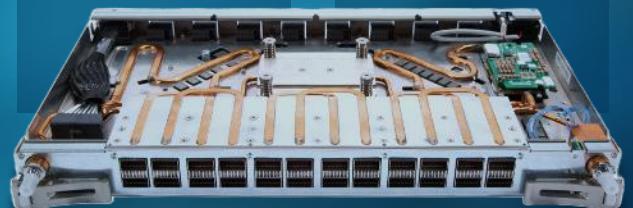
10,624
Nodes



Fabric

Peak
Injection
Bandwidth
2.12
PB/s

Peak
Bisection
Bandwidth
0.69
PB/s



Dragonfly Topology

Memory

10.9PB
DDR Capacity

1.36PB
HBM CPU Capacity

8.16PB
HBM GPU Capacity

5.95PB/s
Peak DDR BW

30.5PB/s
Peak HBM BW CPU

208.9PB/s
Peak HBM BW GPU

Storage

230PB
DAOS Capacity

31TB/s
DAOS Bandwidth

1024
DAOS Node #

Aurora Performance Summary



HPL-MxP

- #1 Ranking - fastest AI supercomputer
- 10.6 EF/s
- ~89% of the system



Top500

- #2 Ranking on Top500
- 1.012 EF/s
- ~87% of the system



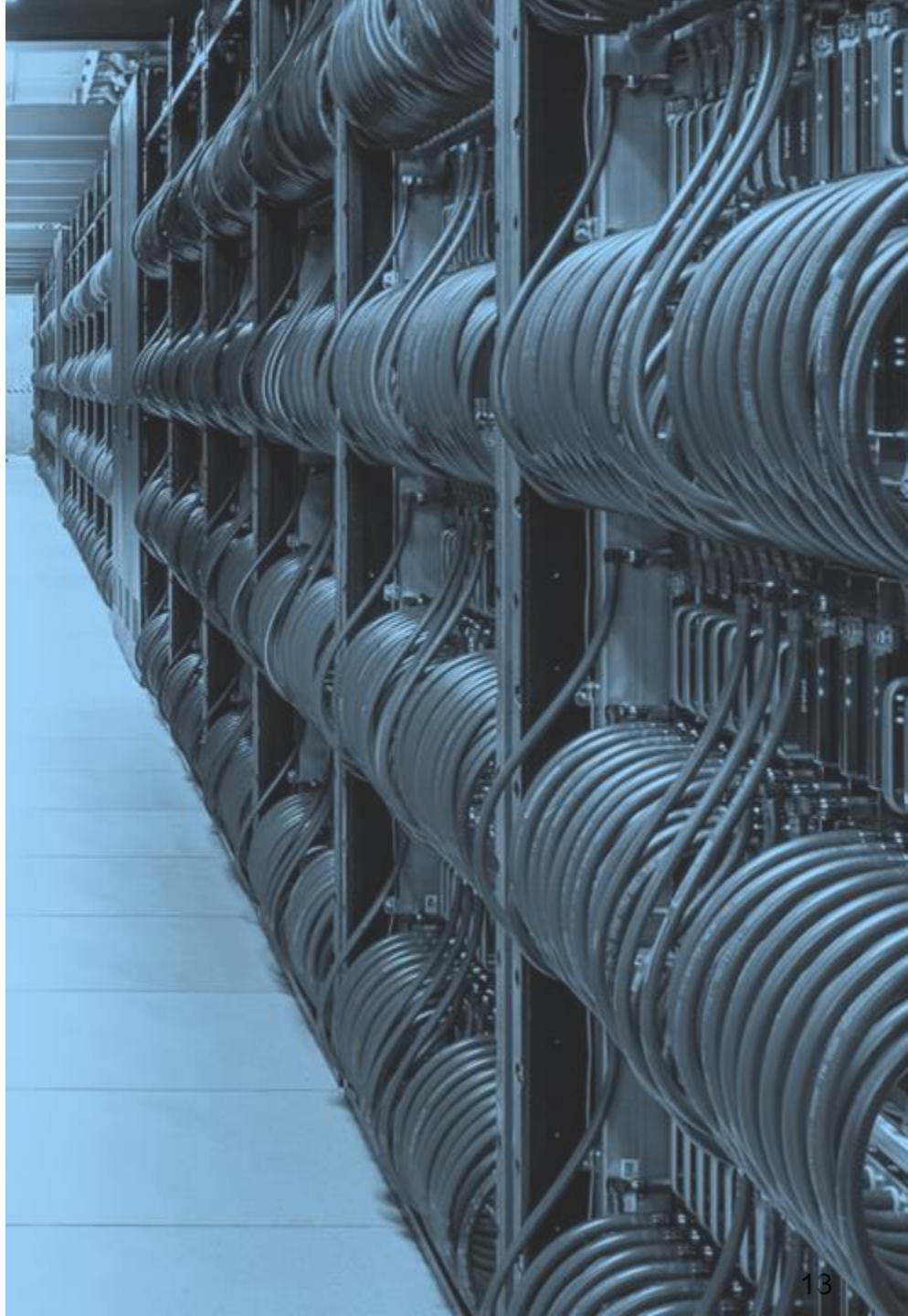
HPCG

- #3 Ranking
- 5,612. TF/s
- ~38.5% of the system



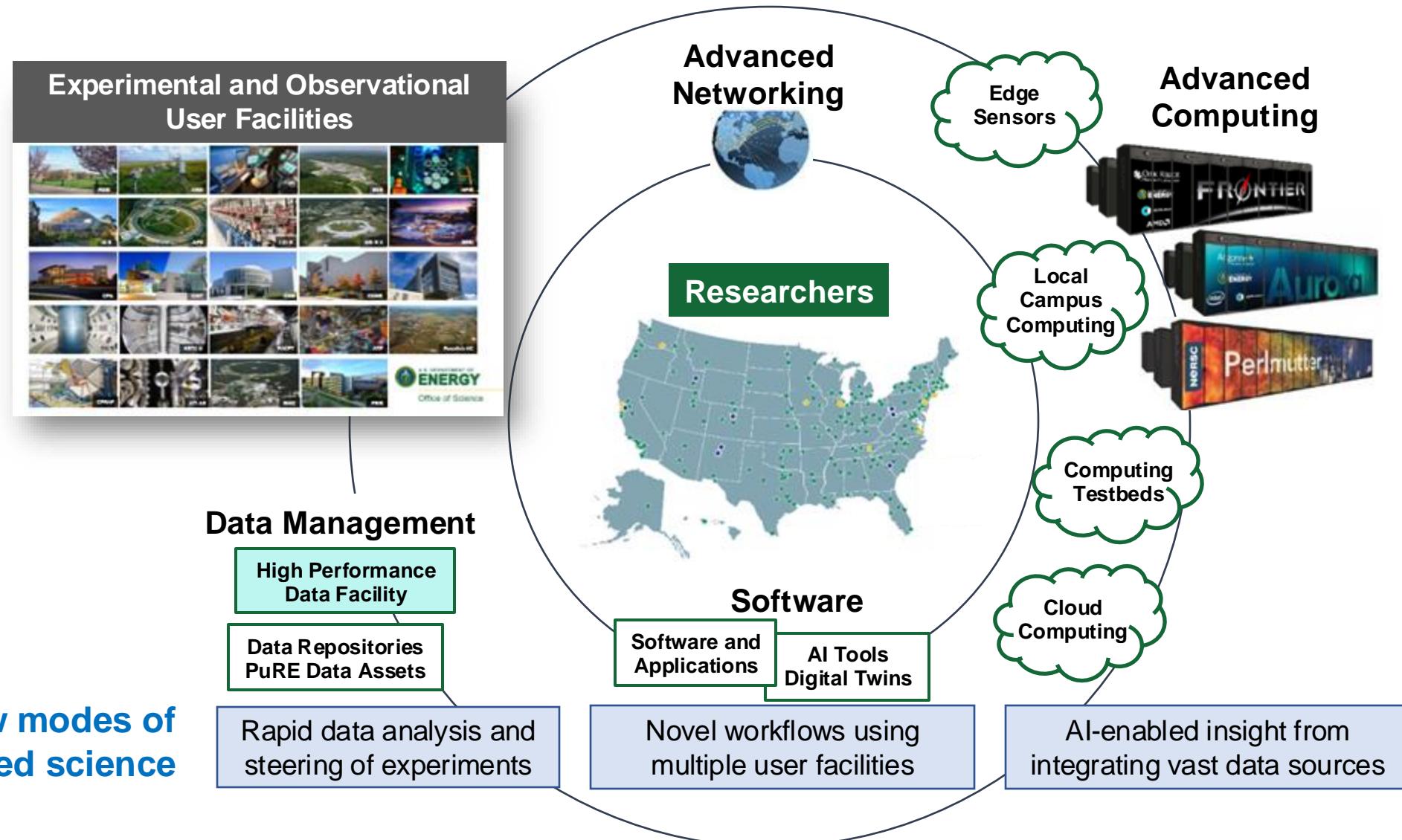
Graph500

- #5 Ranking
- 24,250 GTEPS



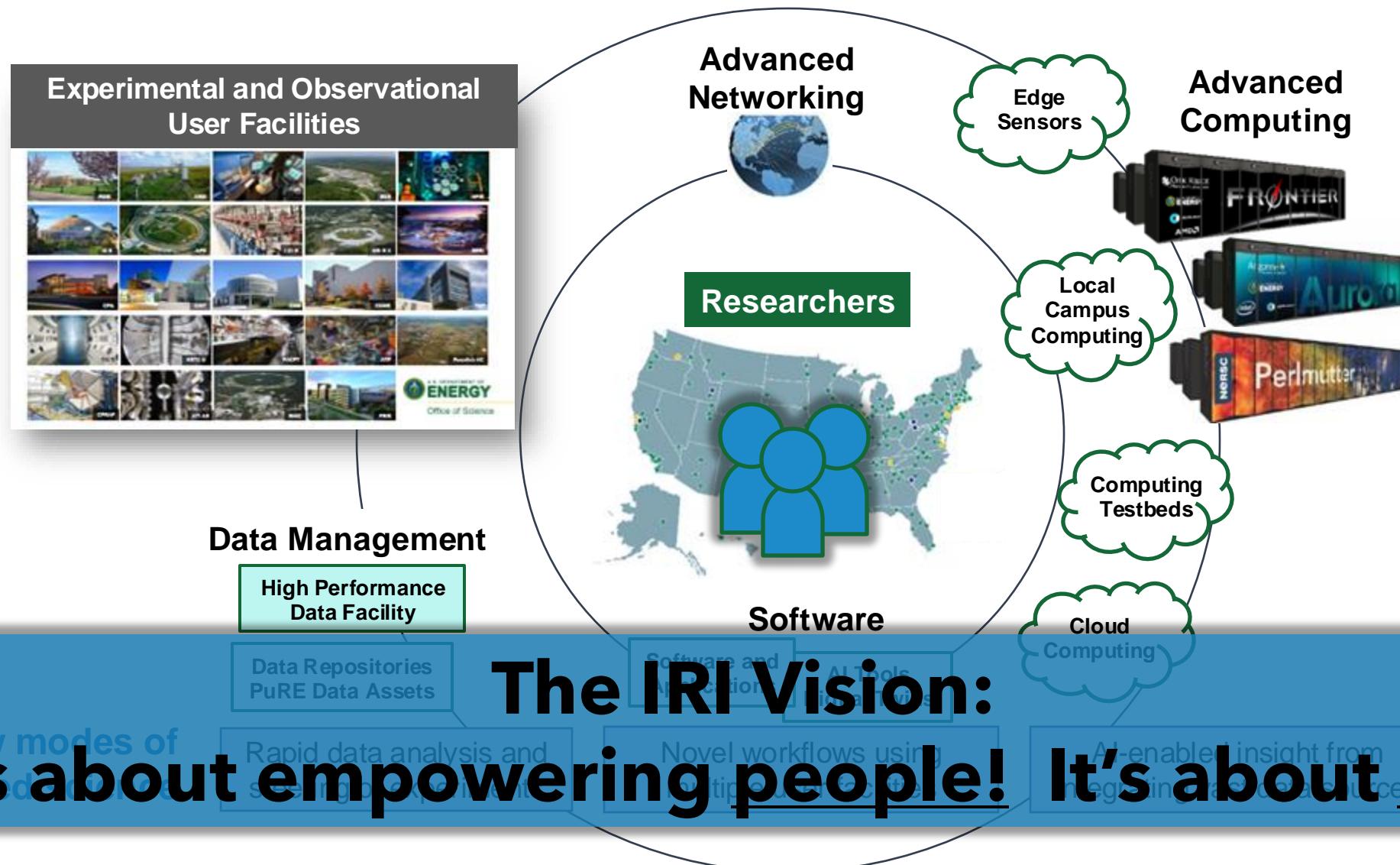
DOE's Integrated Research Infrastructure (IRI) Vision:

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

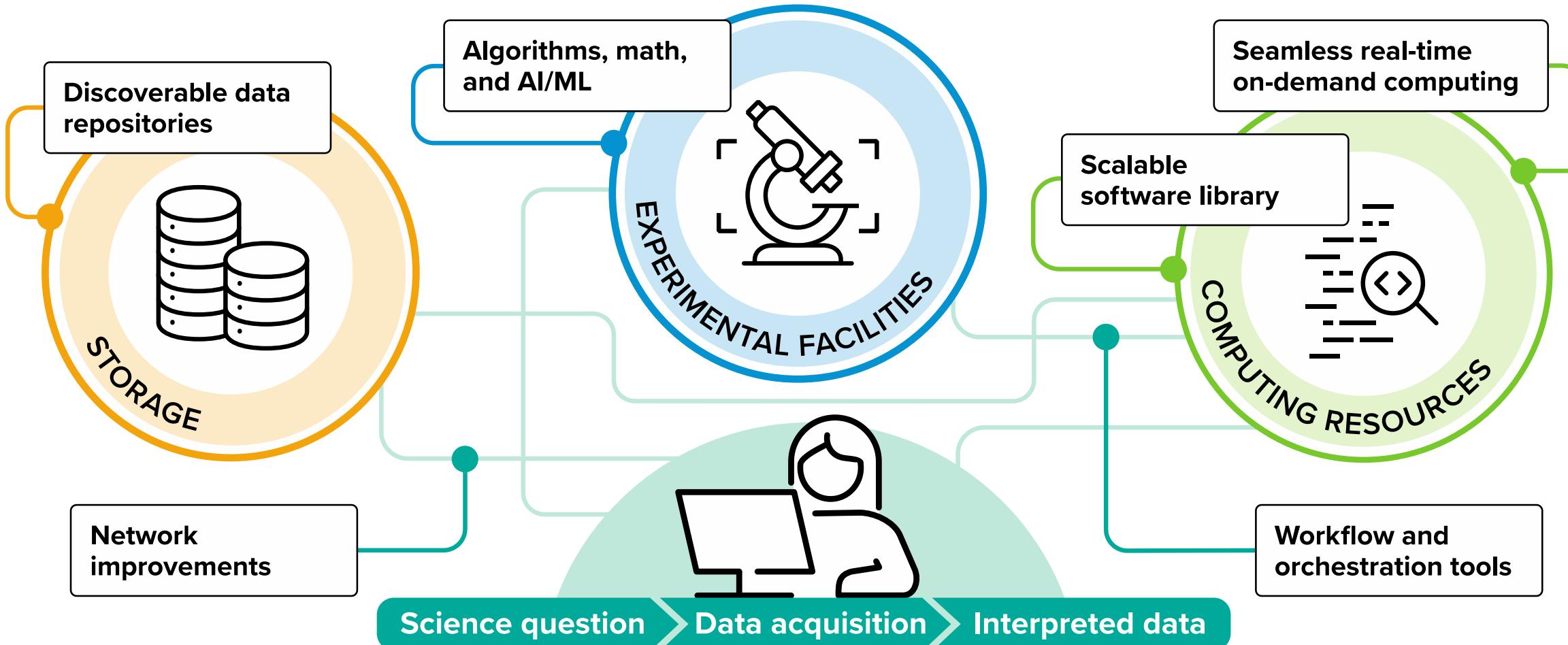


DOE's Integrated Research Infrastructure (IRI) Vision:

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

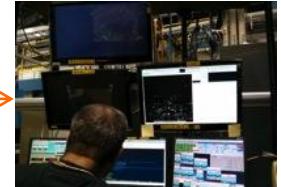
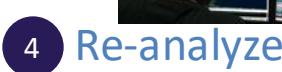


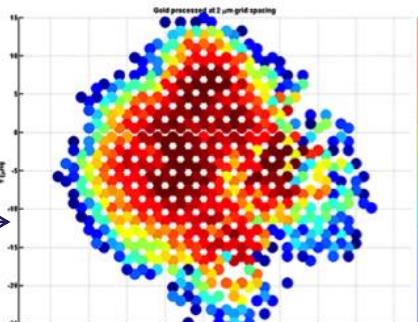
Integrated Research Infrastructure



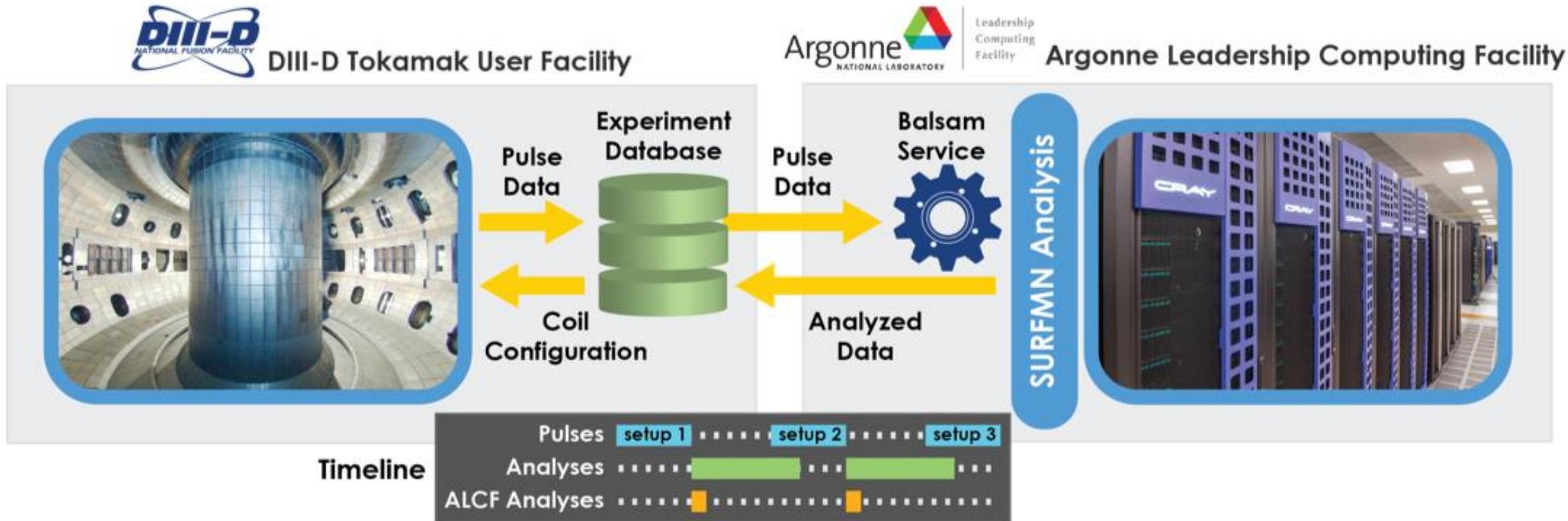
Experiments Integrating Research Infrastructure

Impact	Accomplishments	Status
<ul style="list-style-type: none"> APS scientists use Mira to process data from live HEDM experiments, providing real-time feedback to correct or improve in-progress experiments 	<ul style="list-style-type: none"> 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. 	<ul style="list-style-type: none"> Workflow is established Augmenting real-time scheduling 

 → **Analyze**
 → **Assess**
 → **Fix**
Re-analyze ← 


 Red indicates higher statistical confidence in data

Experiments Integrating Research Infrastructure

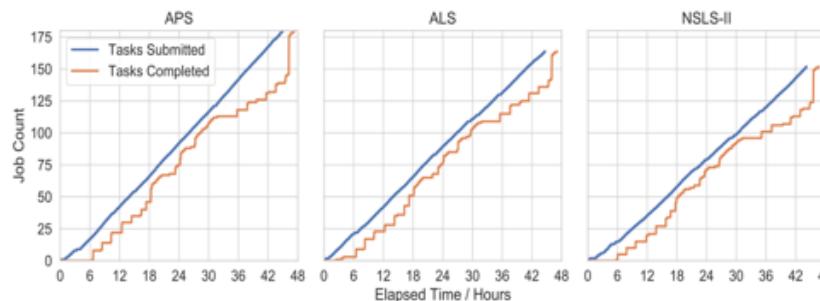


Experiments Integrating Research Infrastructure

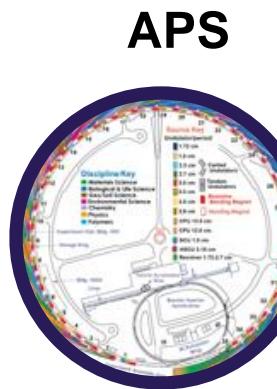
Utilizing leadership computing facility for continuous near real-time XPCS data processing

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



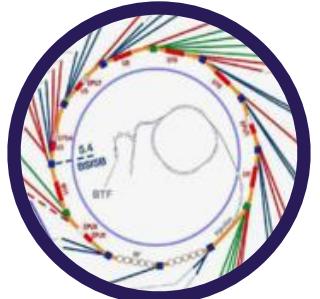
ALCF Theta (11.7PF)



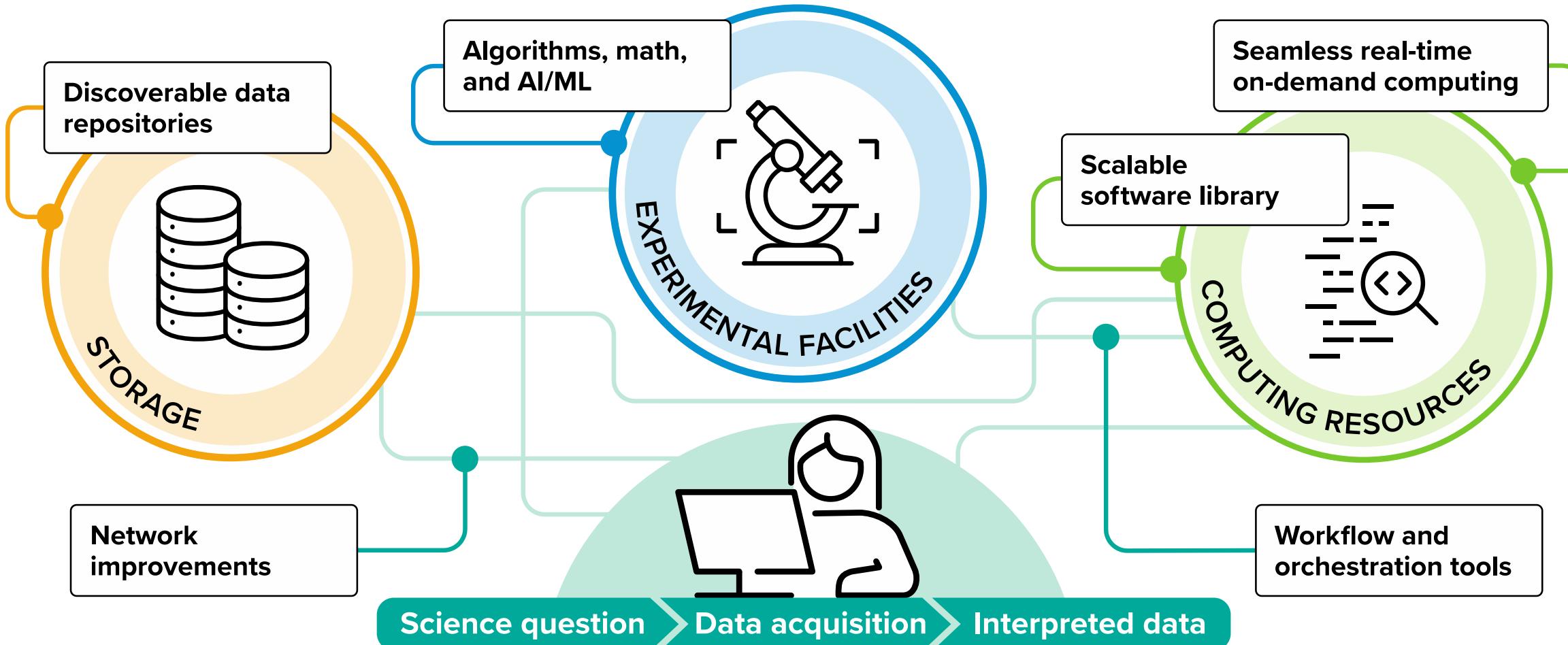
NSLS-II



ALS



Integrated Research Infrastructure



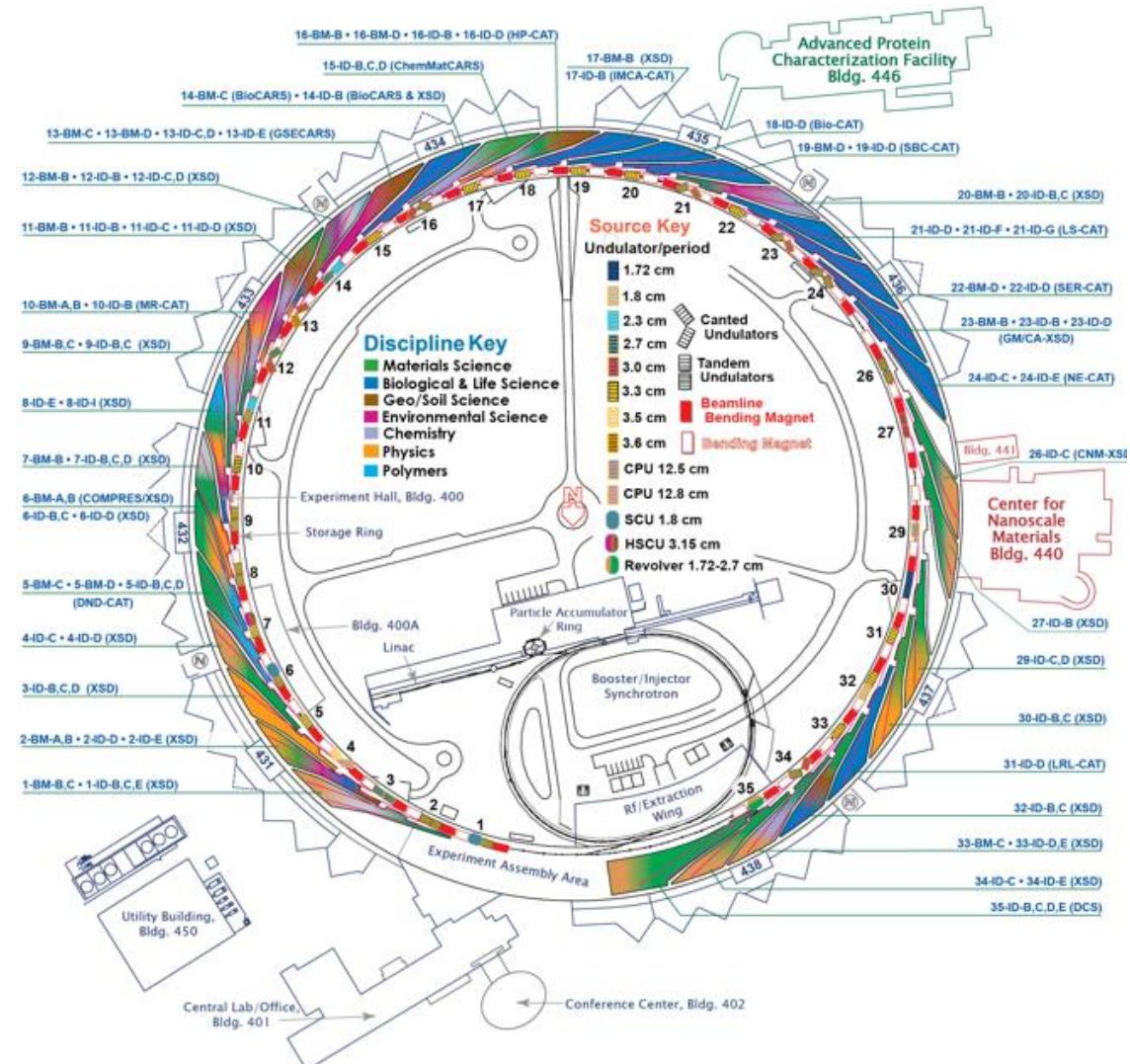
Advanced Photon Source

67 beamlines capable of independent operation; all unique

~5,700 researchers per year from academia, industry, and government

Measurements performed at the APS:

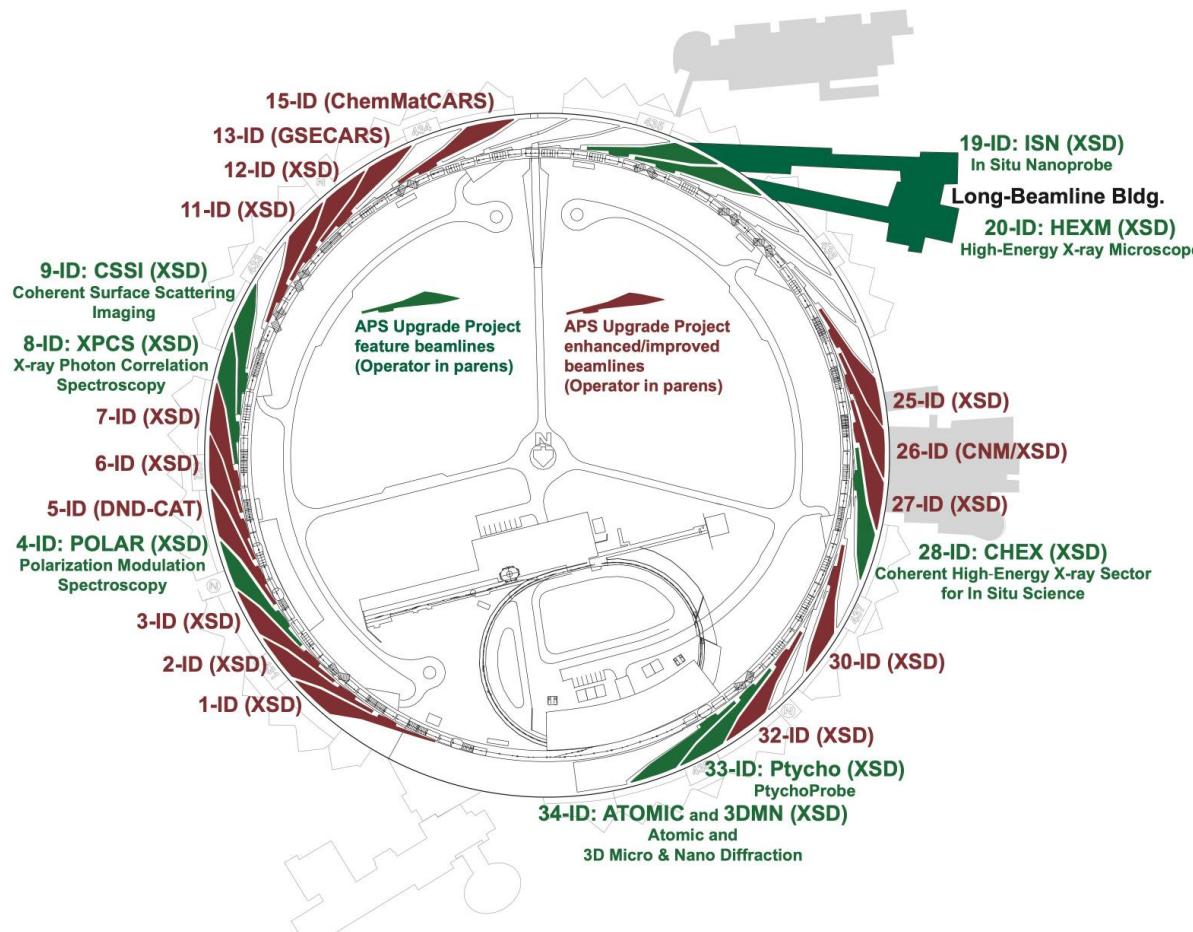
- Imaging (tomography, radiography)
- Scanned probe microscopy (fluorescence mapping)
- Coherent scattering (XPCS, Ptychography)
- Diffraction (MX, powder, PDF, HEDM, stress/strain, SAXS, GISAXS)
- Spectroscopy (IXS, nuclear resonant scattering, XMCD, XAFS)



Advanced Photon Source Upgrade (APS-U)

Scope

- \$815 M project to update and renew the facility
- Re-uses \$1.5 B in existing infrastructure



- Completely new storage ring, 42 pm emittance @ 6 GeV, 200 mA
- New and updated insertion devices
- Combined result in brightness increases of up to 500x
- 9 new feature beamlines (green)
- 15 beamline enhancements (red)

Scale of the Challenge

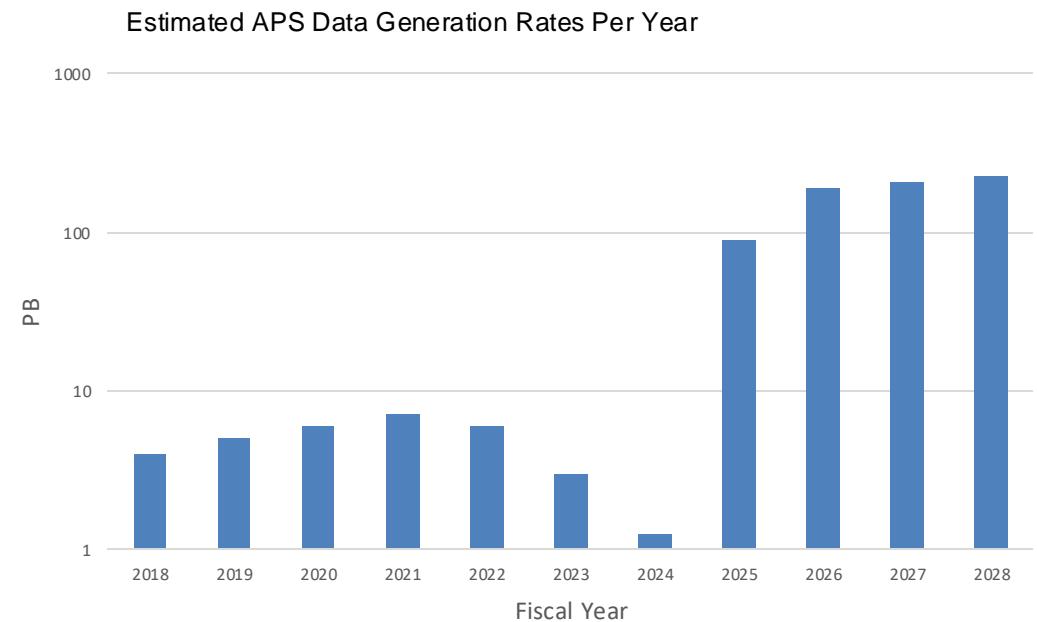
Multiple order-of-magnitude increase in demand for computing resources over next decade

APS-U Era

- ~68 beamlines
- 9 feature beamlines and many enhanced beamlines

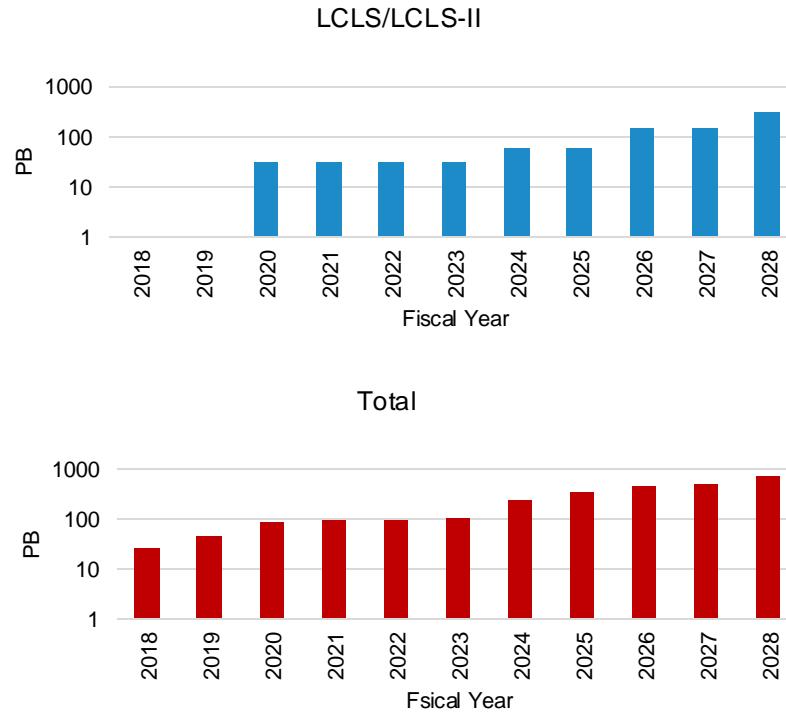
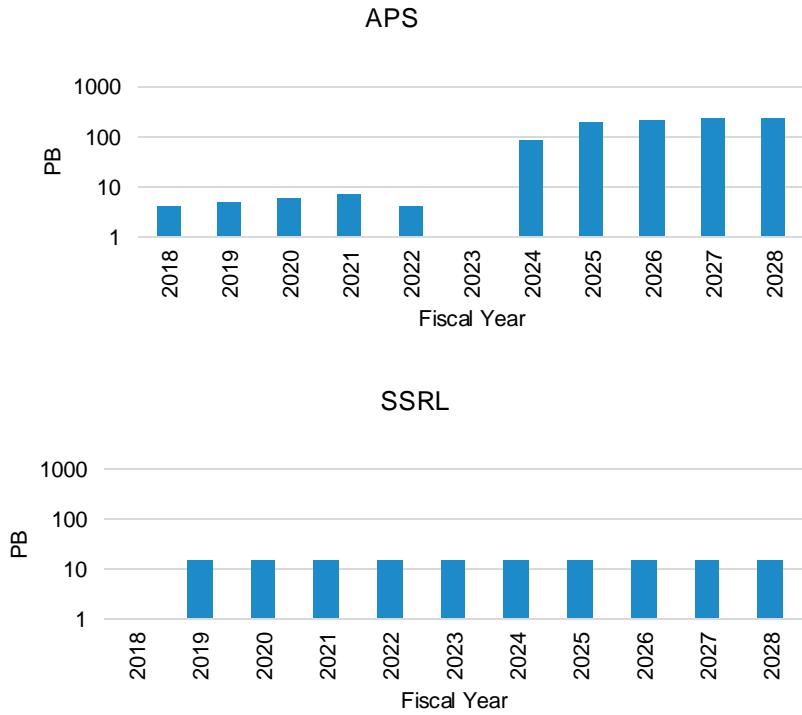
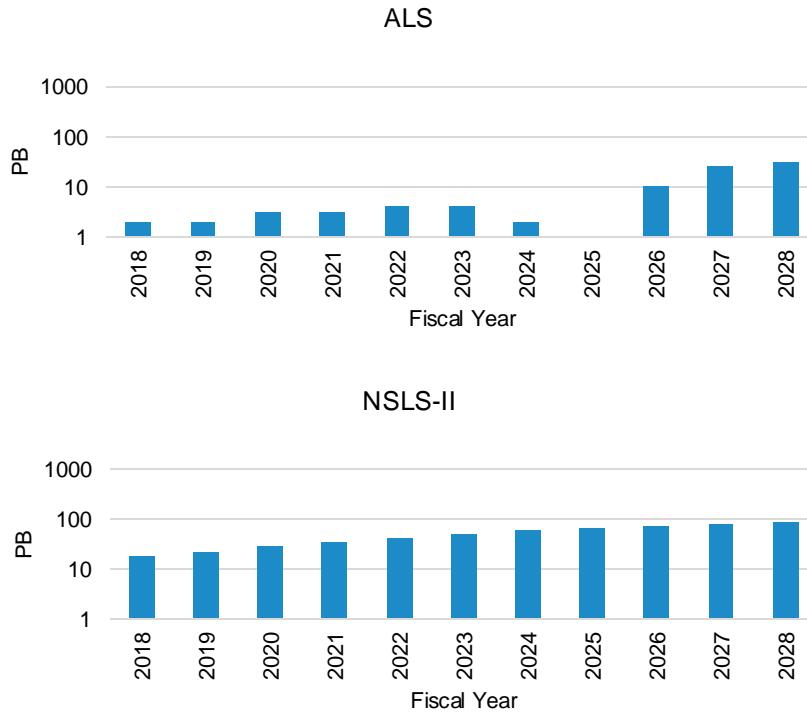
Over the next decade the APS will

- Generate 100s of petabytes (PBs) of raw data per year
- Require 10s of PFLOP/s of on-demand computing power for first pass data processing and reduction



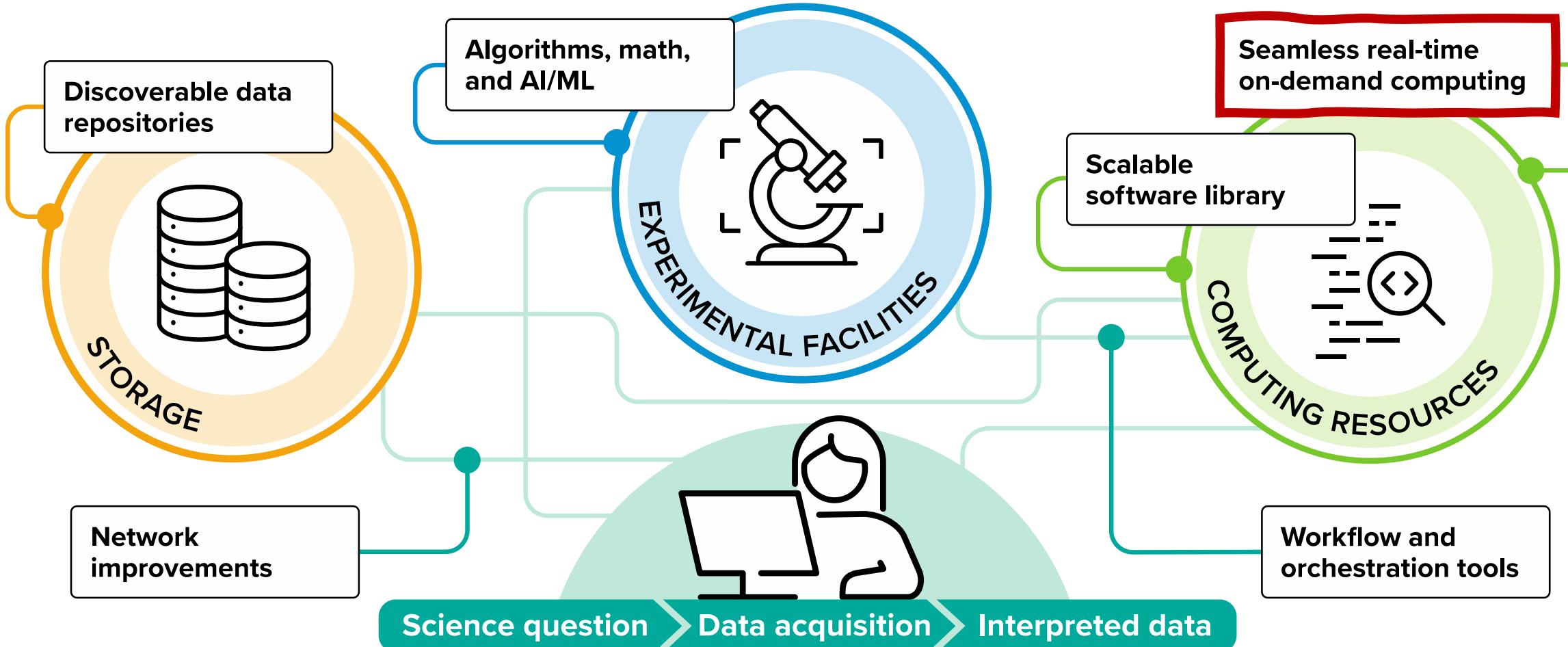
Log scale: Anticipated aggregate APS X-ray Science Division data generation per year. Data generation during FY23 is estimated to be at 50% of the peak due to the beginning of the storage ring replacement period. Data generation during FY24 is considerably lower due to the storage ring replacement period and beginning of beamline commissioning.

US Light Source Data Generation Estimates

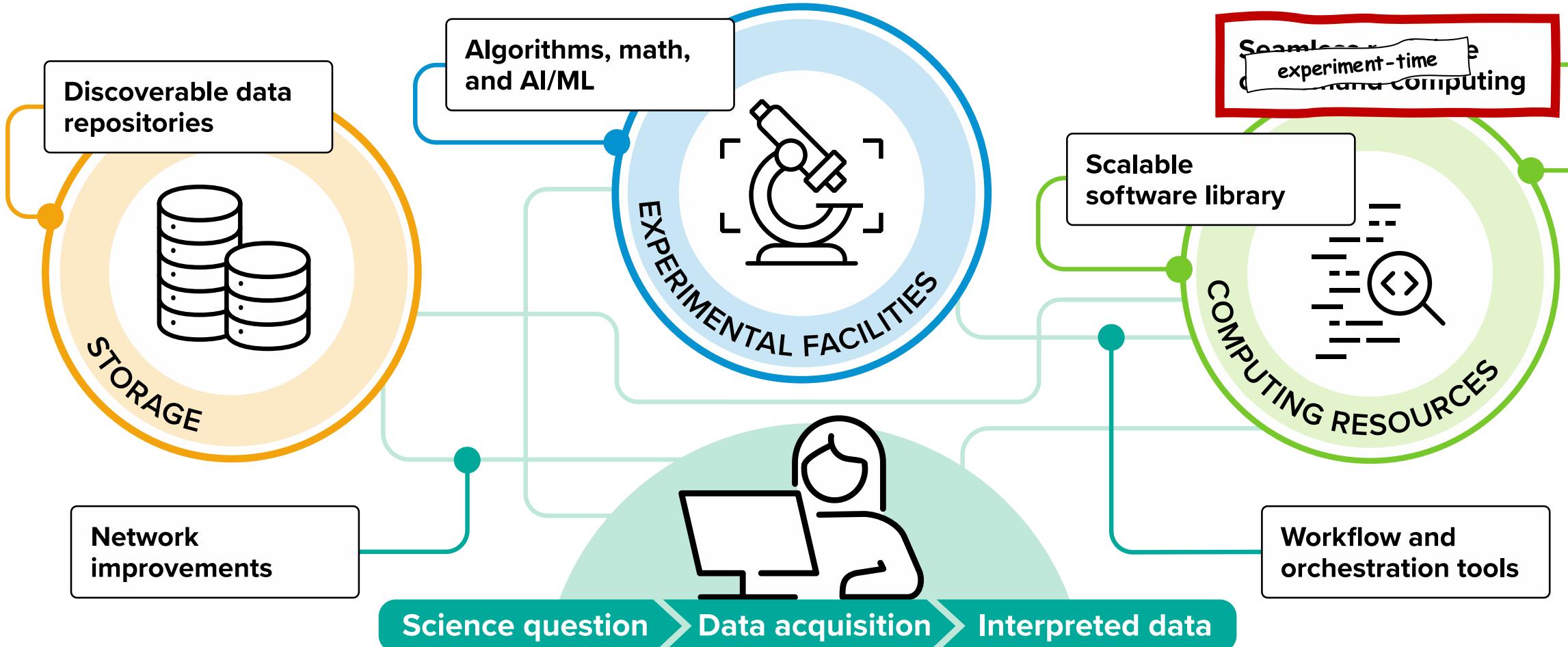


Log Scale: Estimated data generation rates per year at the US Department of Energy, Basic Energy Sciences, light sources. At the ALS and APS, data generation will stop during 2025 and 2023, respectively, due to installations of new storage rings. Aggregate data generation across the light sources will approach the exabyte (EB) range per year by 2028. The differences in data generation rates across the facilities depend on the number, rate and resolution of the detectors at each instrument which in turn depend on factors like the brightness of the source and the actual requirements of the experimental technique specific to that instrument.

Integrated Research Infrastructure

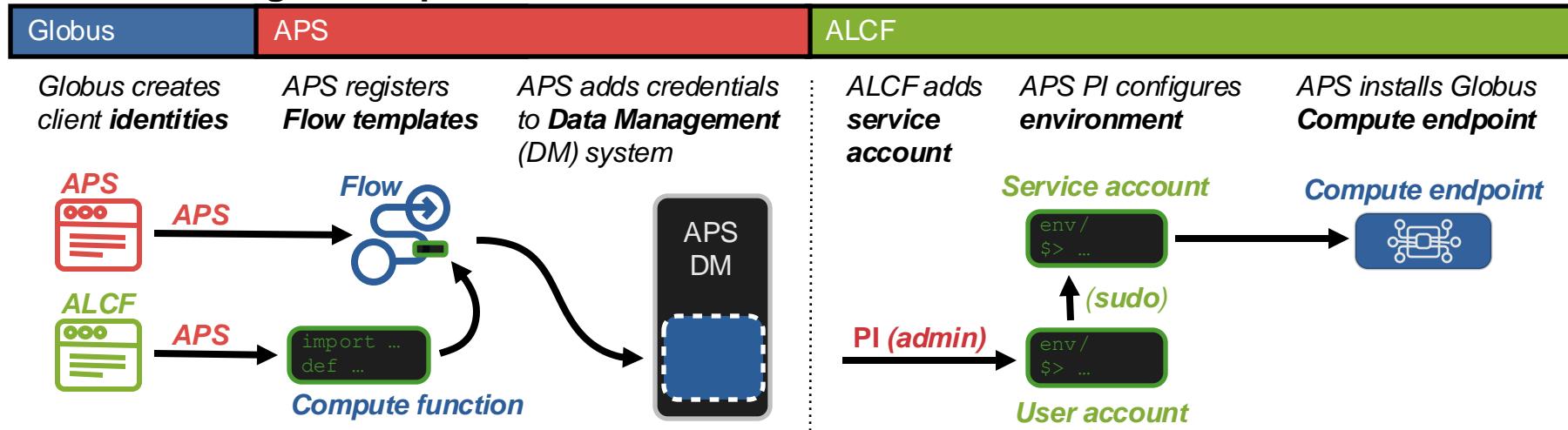


Integrated Research Infrastructure

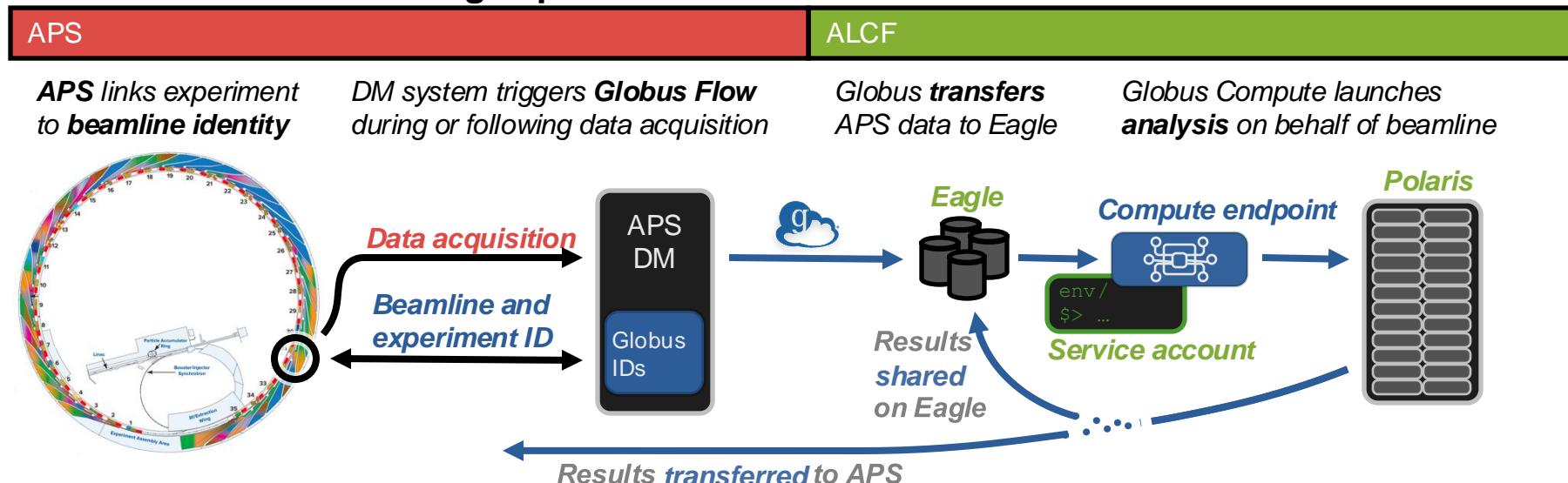


Putting the Research Infrastructure Together (Nexus)

One-time configuration per beamline

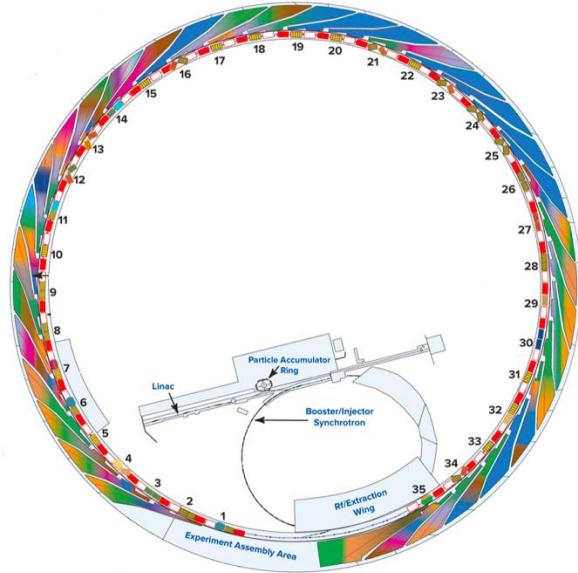


Automated workflow during experiments

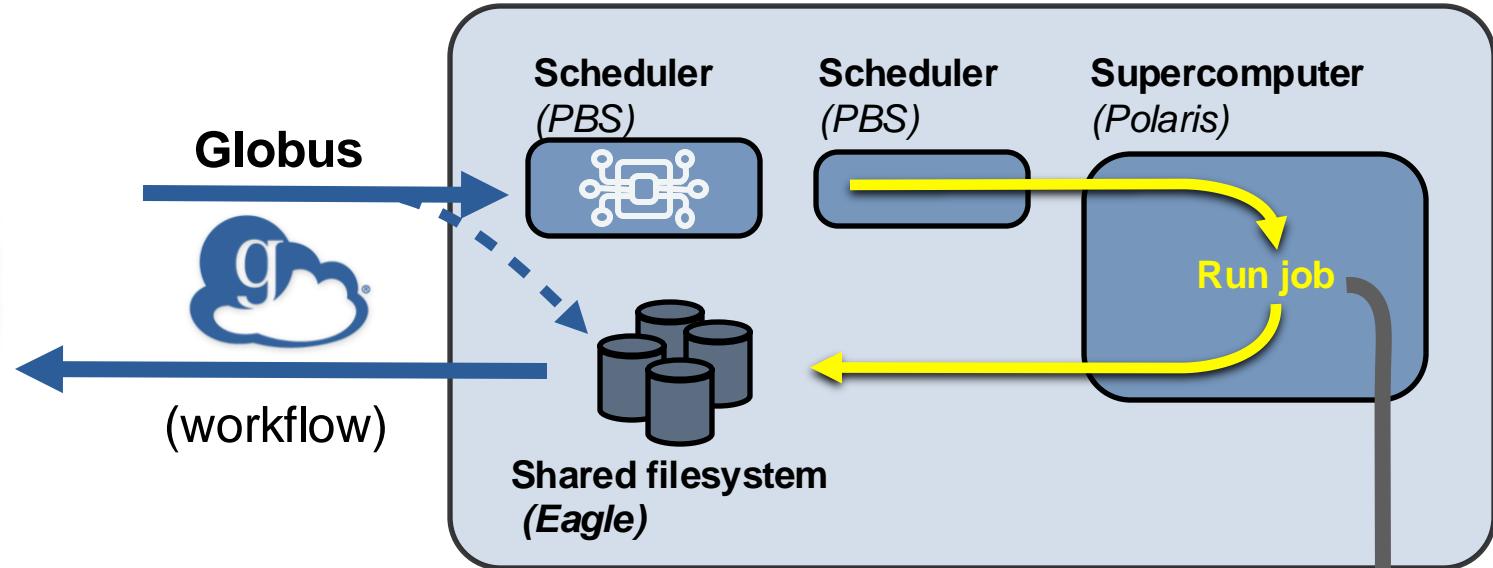


No Human in the Loop Experiment

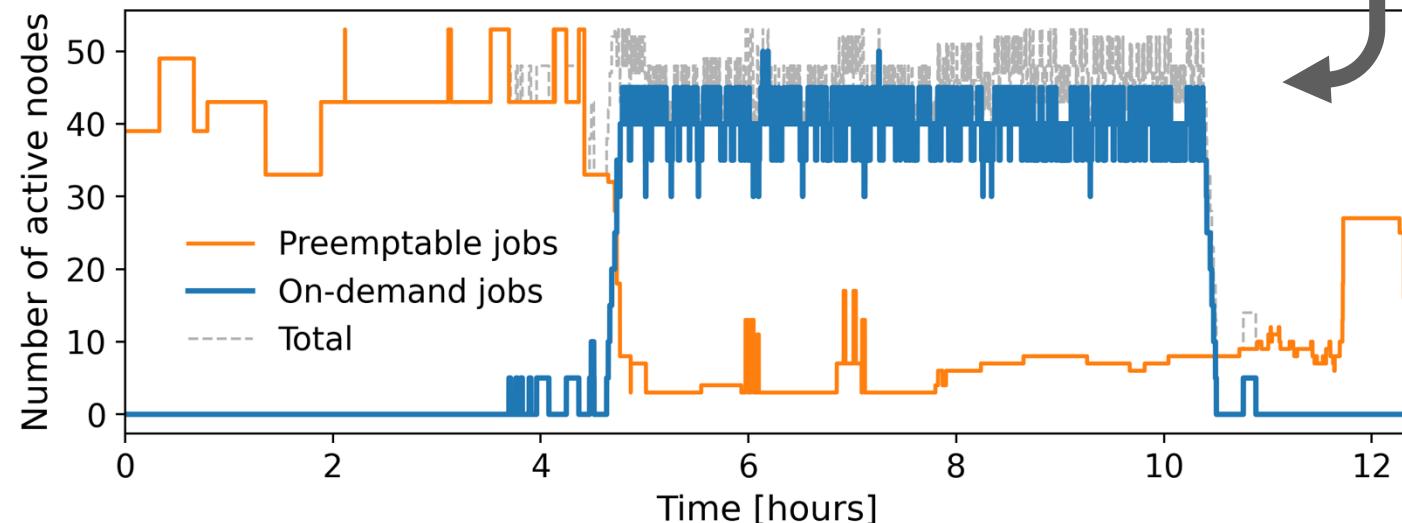
APS (data acquisition)



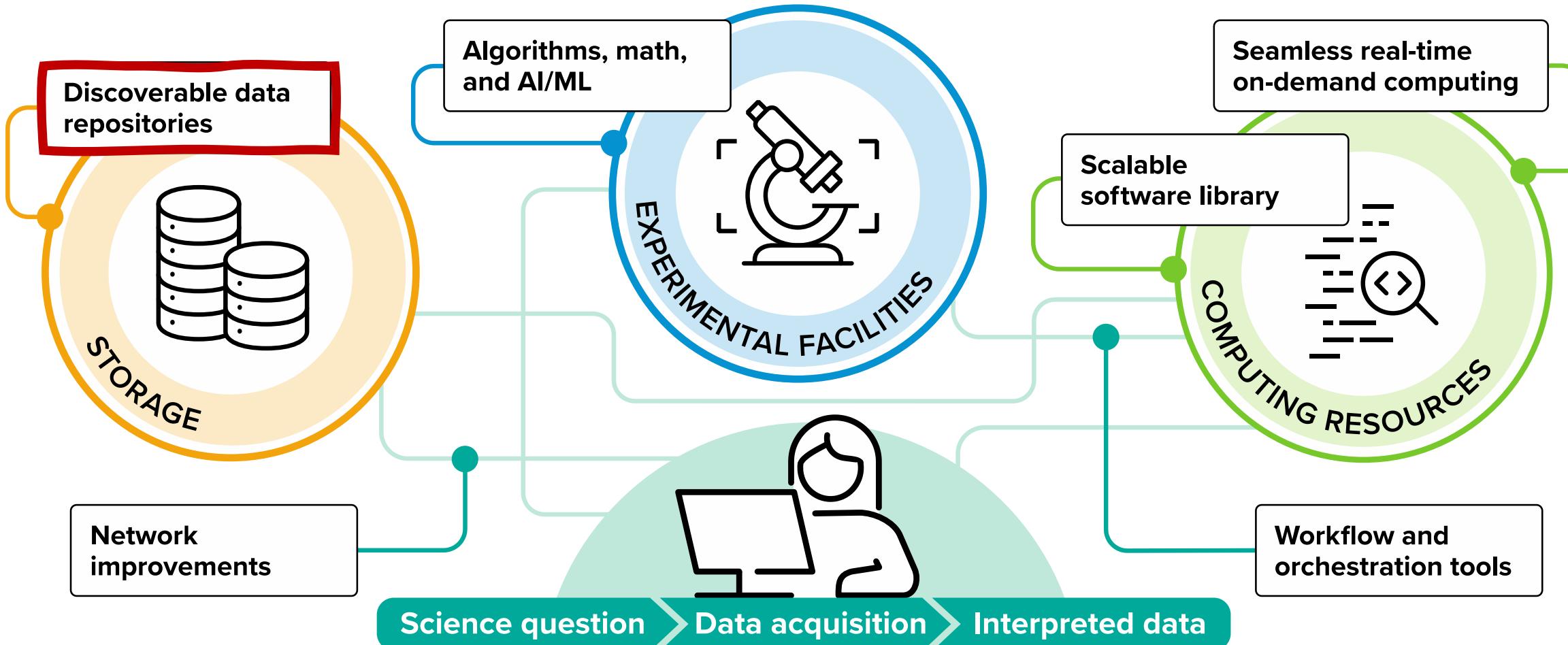
ALCF (on-demand data analysis)



Computation example
(Part of Laue experiment)
April 1st, 2023



Integrated Research Infrastructure



Data Sharing and Community Resources

This site relies on ACDC, a resource of the Argonne Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC02-06CH13573.

Argonne NATIONAL LABORATORY

globe icon

g cloud icon

ALCF Documentation Login

ALCF Community Data Co-Op

Providing Researchers Data Sharing Capabilities

APS 2ID XRF
X-Ray fluorescence imaging for a variety of applications
Private [Search This Project](#)

APS HEDM
High-Energy Diffraction Microscopy (HEDM) is a 3D technique which resolves the structures and strains of individual grains within bulk polycrystalline materials.
Private [Search This Project](#)

APS Ptychography
A computational method of microscopic imaging by processing many coherent interference patterns
Private [Search This Project](#)

APS XPCS
X-ray Photon Correlation Spectroscopy (XPCS) is a technique to study dynamics in materials at nanoscale by identifying correlations in time series of area detector images
Private [Search This Project](#)

DLHub
A multi-tenant system that provides both model repository and serving capabilities with a focus on science applications
Public Data [Search This Project](#)

GMCA Index
X-ray Photon Correlation Spectroscopy (XPCS) is a technique to study dynamics in materials at nanoscale by identifying correlations in time series of area detector images
Private [Search This Project](#)

Kanzus Beamline Chip Analysis
Synchrotron serial crystallography for the study of protein and enzyme dynamic processes
Private [Search This Project](#)

Neurocartography
Large volume reconstructions of the fine structure of the nervous system
Search Only [Search This Project](#)

RPL Index
RPL portal
Private [Search This Project](#)

SDL BIO Index
Self Driving Laboratories @ BIO
Private [Search This Project](#)

The Materials Data Facility
A simple way to publish, discover, and access materials datasets
Public Data [Search This Project](#)

X-ray Science Division Imaging
The IMG group designs, supports, and operates state-of-the-art x-ray instrumentation and scientific software for 2-D and 3-D full-field imaging.
Private [Search This Project](#)

About ALCF Community Data Co-Op (ACDC)

Data Portal for Discovery and Access

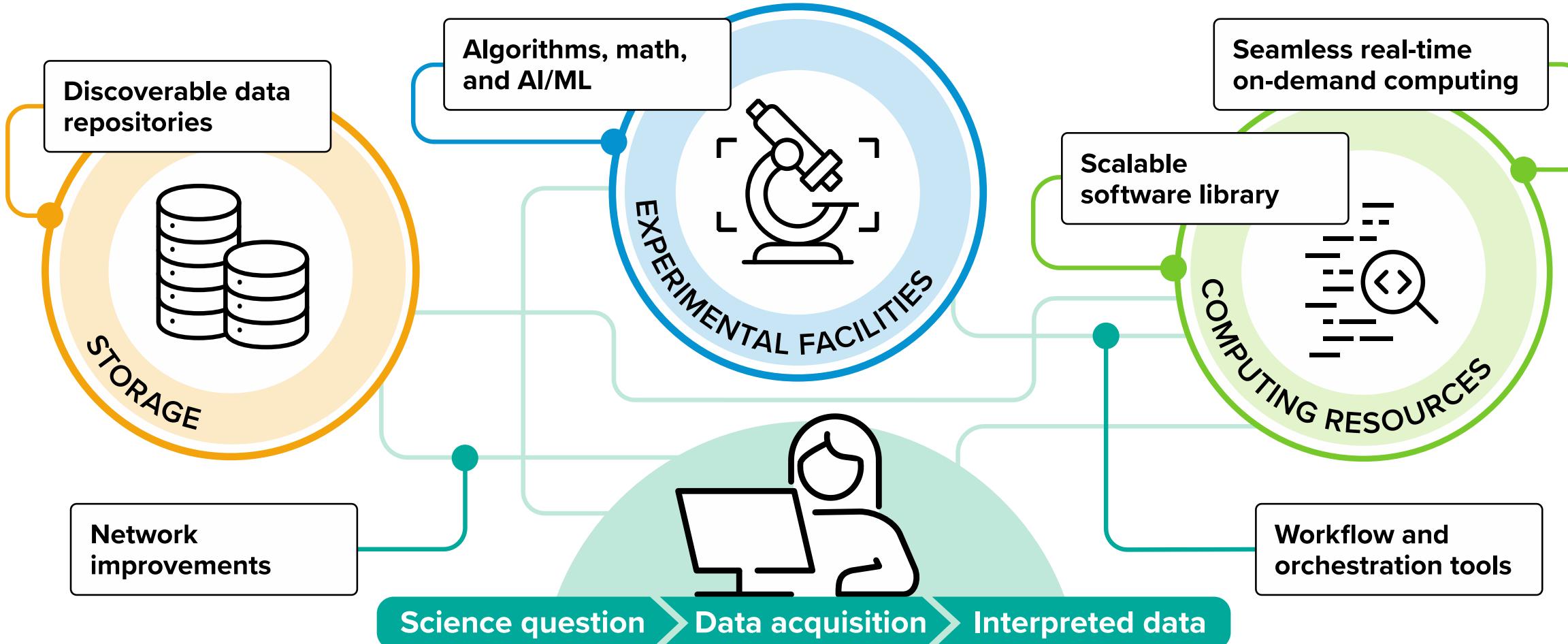
ACDC's fully supported production environment is the next step in the expansion of edge services that blur the boundaries between experimental laboratories and computing facilities. The use and prominence of such services at the ALCF are only expected to increase as they become more integral to the facility's ability to deliver data-driven scientific discoveries.

ACDC includes several project-specific data portals that enable search and discovery of the data hosted on Eagle. The portals allow users to craft queries and filters to find specific sets of data that match their criteria and use faceted search for the discovery of data. Portals also provide the framework for other interfaces including data processing capabilities, all secured with authentication and configured authorization policy.

The ACDC portal is a deployment of Django Globus Portal Framework customized for a variety of different projects. For most of these projects, the search metadata links directly to data on Eagle, with browser-based download, preview, and rendering of files, and bulk data access.

Katrin Heitmann, Thomas D. Uram, Hal Finkel, Nicholas Frontiere, Salman Habib, Adrian Pope, Esteban Rangel, Joseph Hollowed, Danila Korytov, Patricia Larsen, Benjamin S. Allen, Kyle Chard, and Ian Foster, *HACC Cosmological Simulations: First Data Release*, **The Astrophysical Journal Supplement Series**, 244(1) 2019.

Integrated Research Infrastructure



Opportunities at Argonne

Undergraduate

<https://www.anl.gov/education/undergraduate-programs>

Graduate

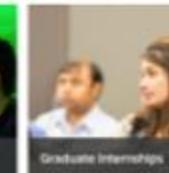
Graduate Programs

Programs offer a variety of research appointments, cooperative education opportunities, and specialized training to fit the needs – real scholars – of any graduate student. We encourage students pursuing a Master's degree as part-time M.S. candidates to apply and be fully supported in our offer. Doctoral programs reward and create new knowledge – and change your world.

The University of Manitoba is dedicated to providing the student experience to be fully work-ready starting Summer 2021. Therefore, it is the University's preference to employ students in a hybrid or virtual status contingent upon local COVID-19 infection rates. Student candidates will be notified of changes in the fall's operational status at the time of offer and/or during their appointment.



Graduate Research



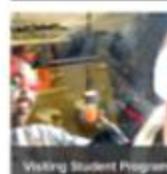
Graduate Internships



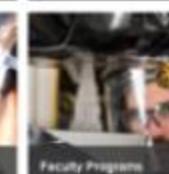
Graduate Temporary Employment



Graduate Training Programs



Visiting Student Program



Faculty Programs

<div style="position: absolute; left: 0; top: 0; width: 10

<https://www.anl.gov/education/graduate-programs>

Faculty

<https://www.anl.gov/education/faculty-programs>





Thank You