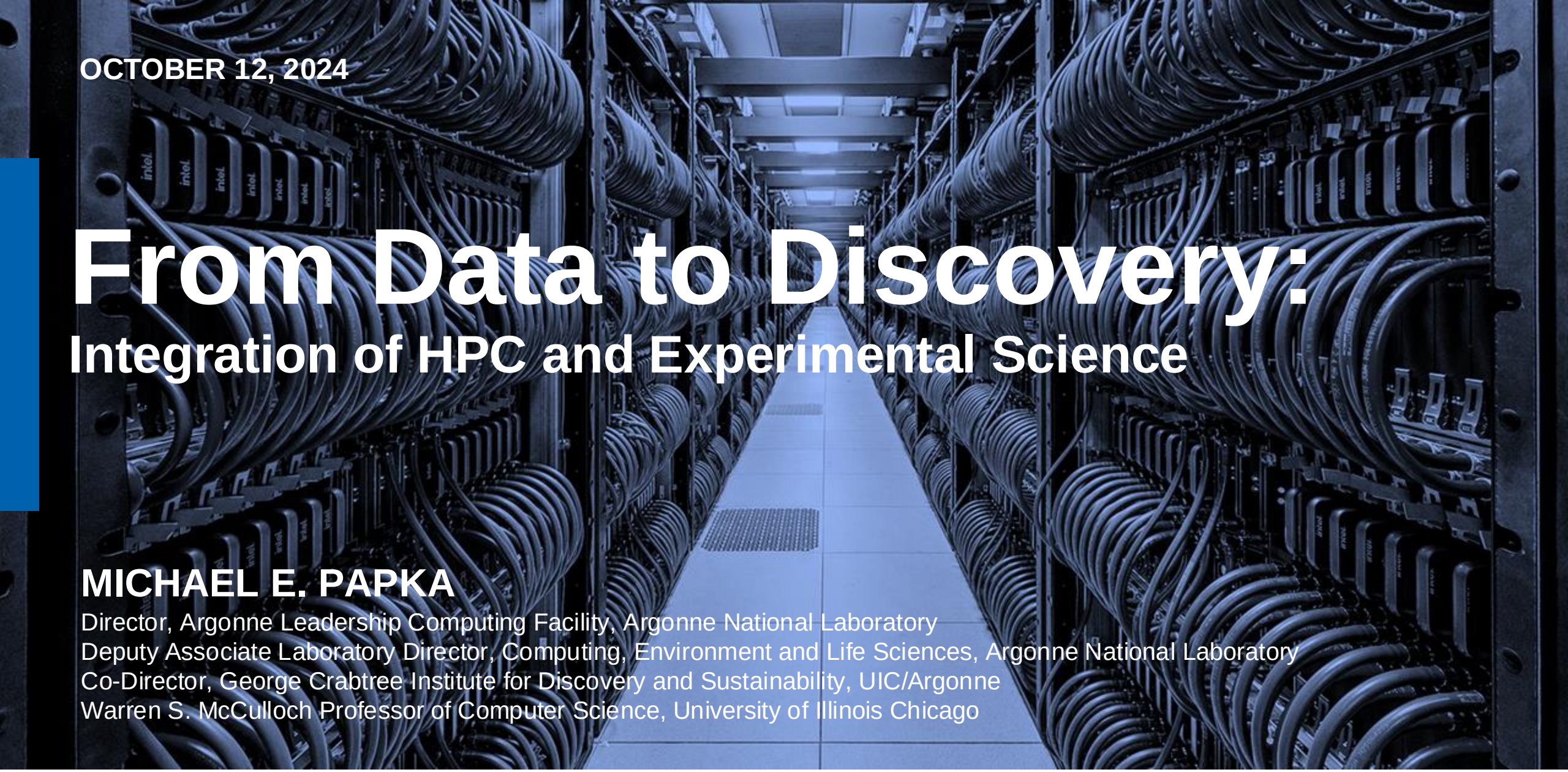


OCTOBER 12, 2024

From Data to Discovery:

Integration of HPC and Experimental Science

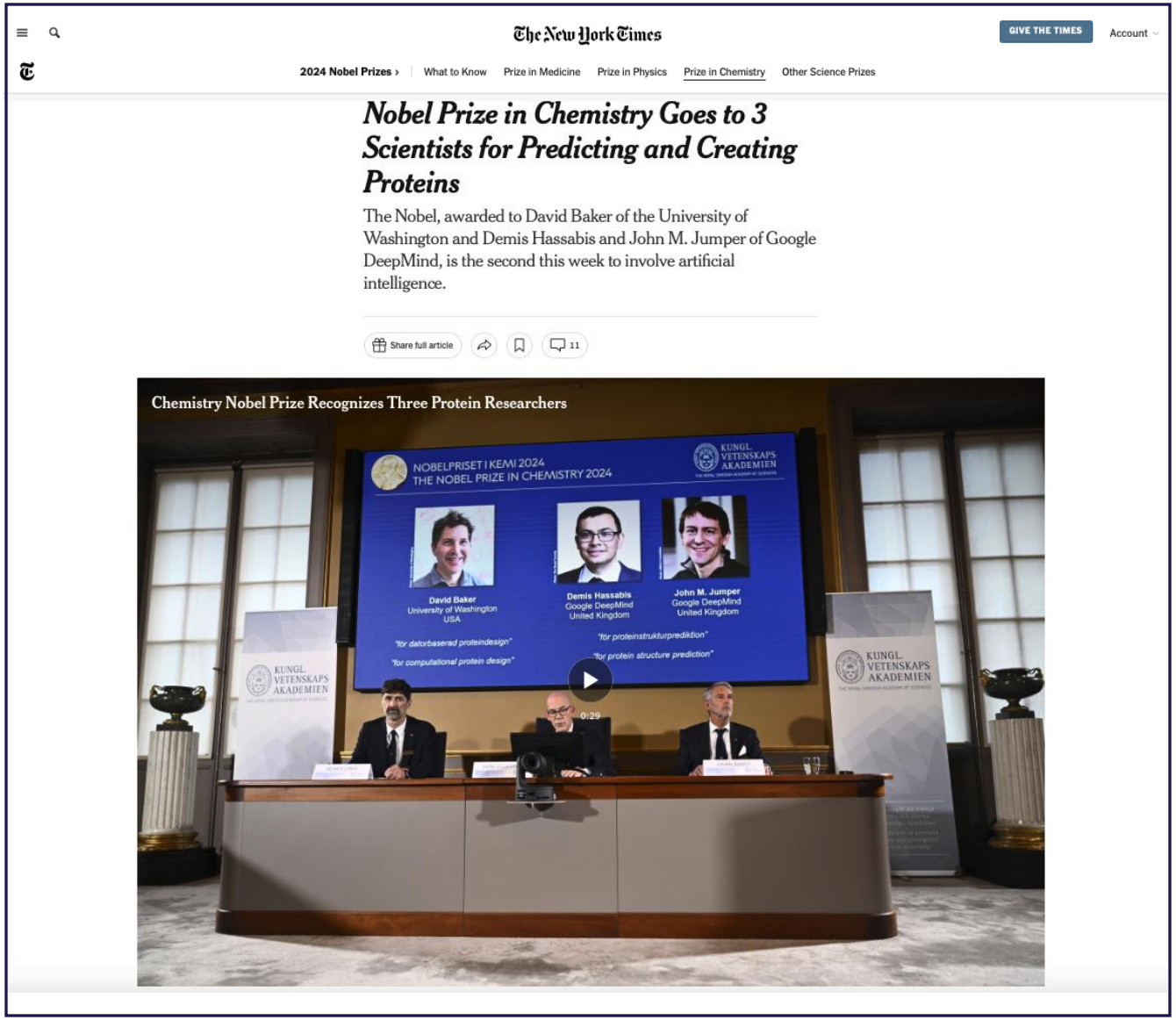
MICHAEL E. PAPKA

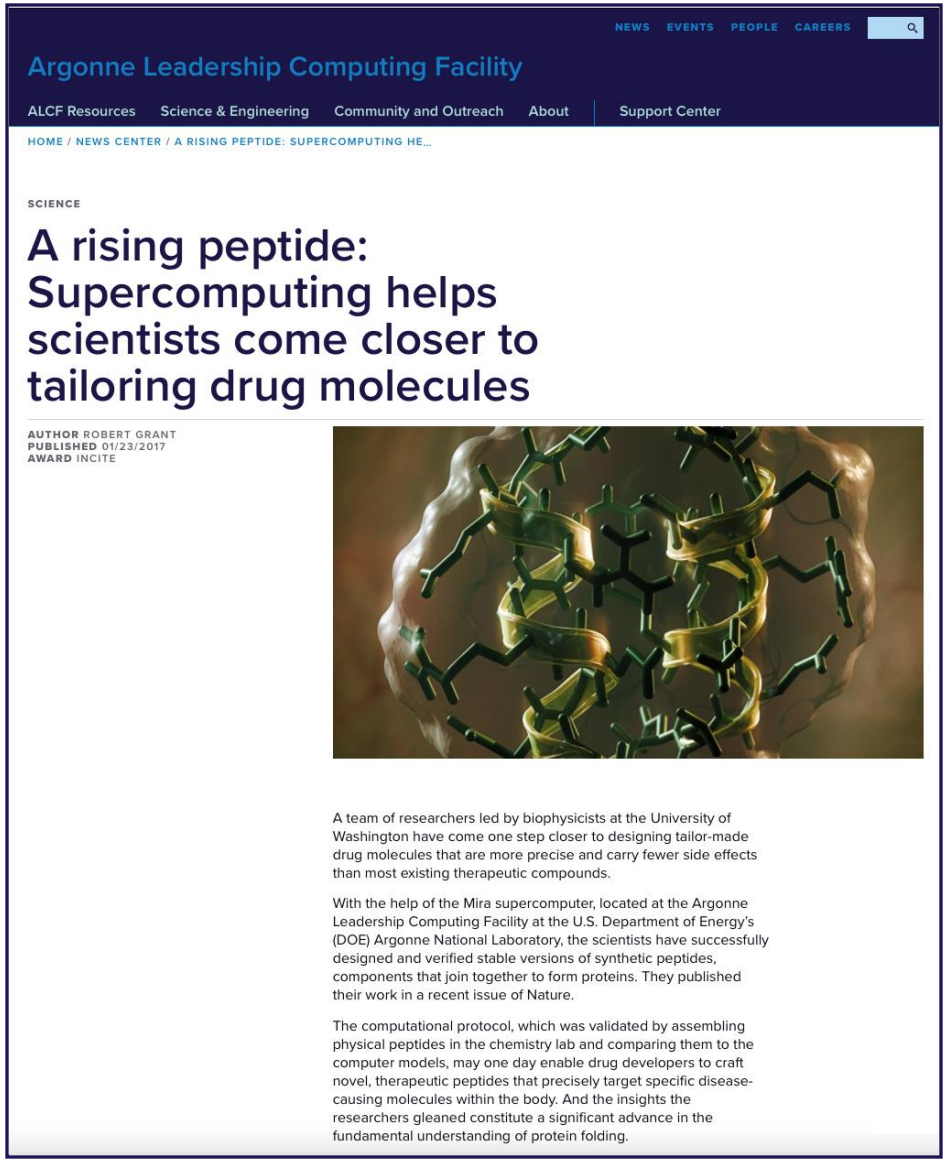
Director, Argonne Leadership Computing Facility, Argonne National Laboratory
Deputy Associate Laboratory Director, Computing, Environment and Life Sciences, Argonne National Laboratory
Co-Director, George Crabtree Institute for Discovery and Sustainability, UIC/Argonne
Warren S. McCulloch Professor of Computer Science, University of Illinois Chicago

Argonne NATIONAL LABORATORY | Argonne Leadership Computing Facility

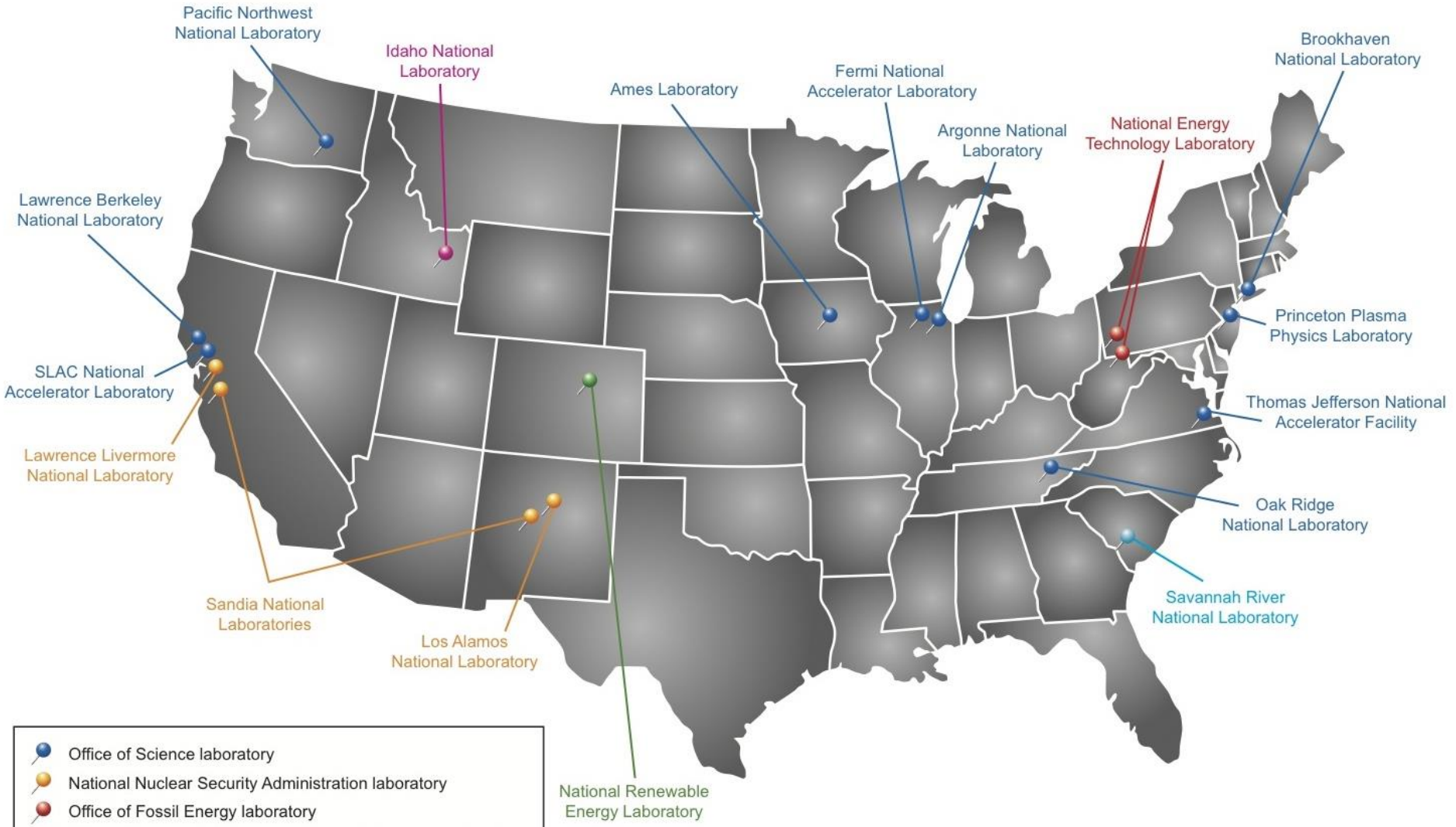
UIC ENGINEERING Computer Science

Nobel Prize in Chemistry 2024 – David Baker





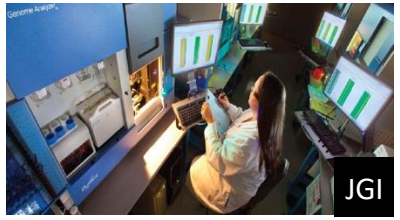
Department of Energy National Laboratories

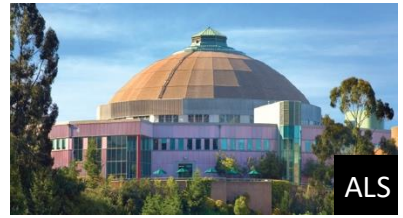


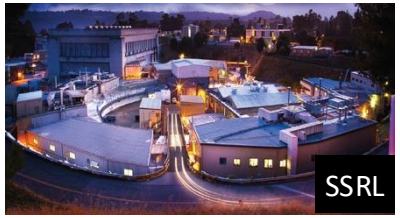
U.S. DEPARTMENT OF ENERGY

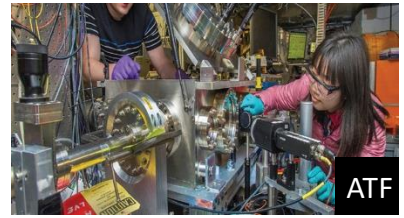
Department of Energy User Facilities

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

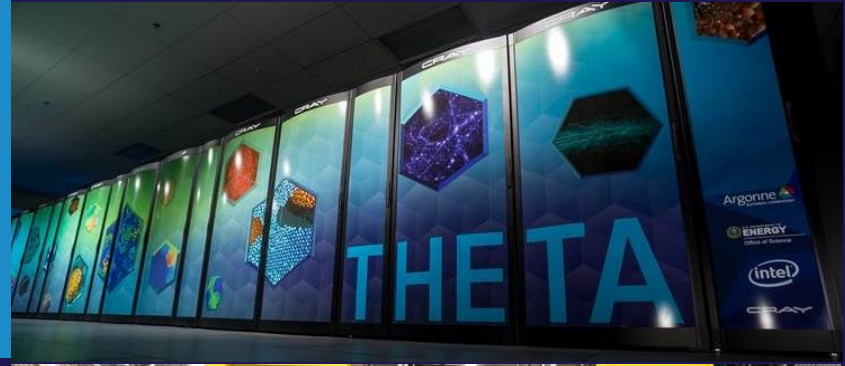
National priorities: Research in security, energy, and sustainability

DOE SC
Advanced Scientific Computing Research User Facilities

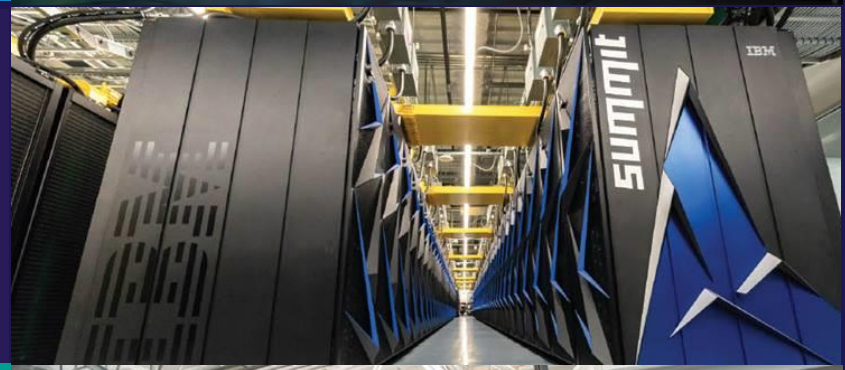
The Advanced Scientific Computing Research (ASCR) program leads the nation and the world in supercomputing, high-end computational science, and advanced networking for science.

ALCF and OLCF make up the DOE Leadership Computing Facility

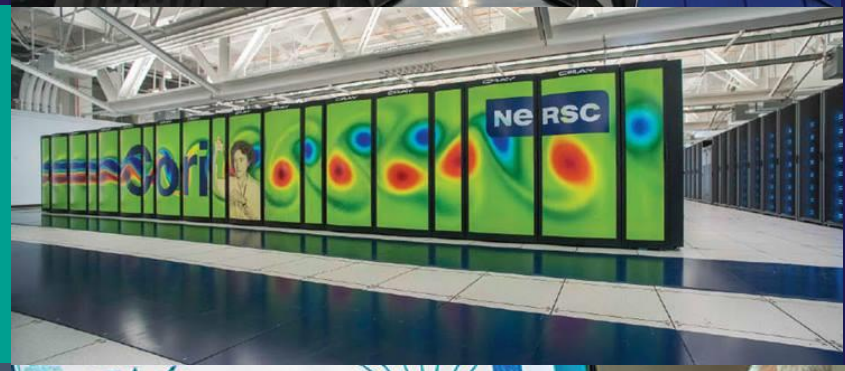
Argonne Leadership Computing Facility (ALCF)



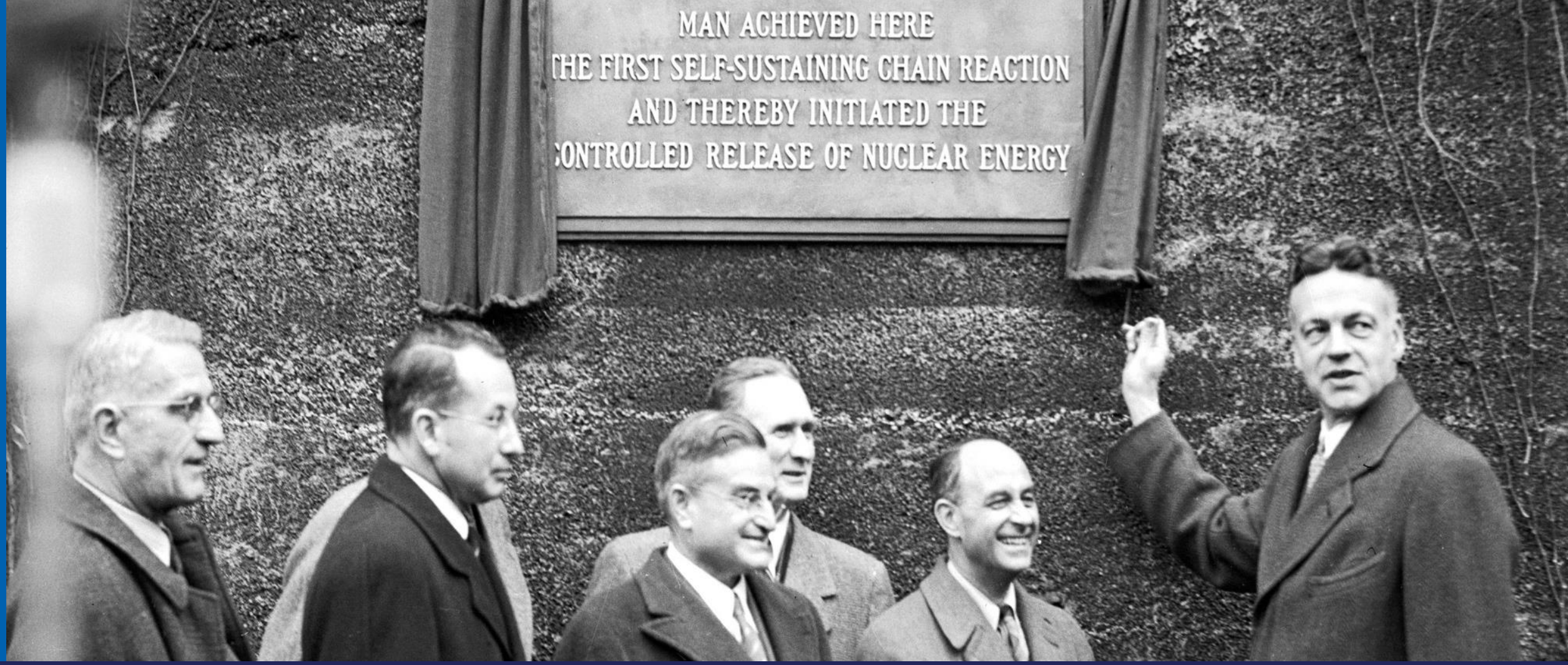
Oak Ridge Leadership Computing Facility (OLCF)



National Energy Research Scientific Computing Center (NERSC)



Energy Sciences Network (ESnet)



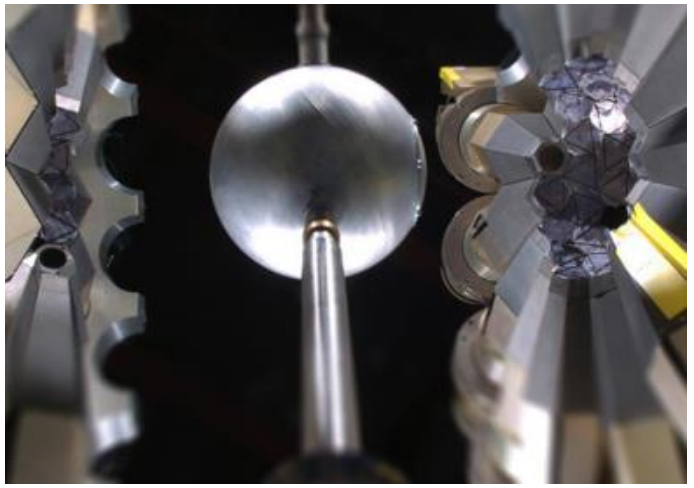
Argonne National Laboratory

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

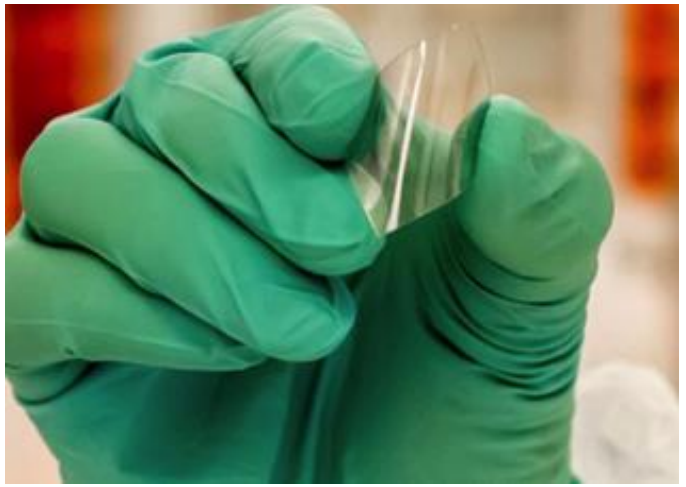
Argonne Leadership Computing Facility



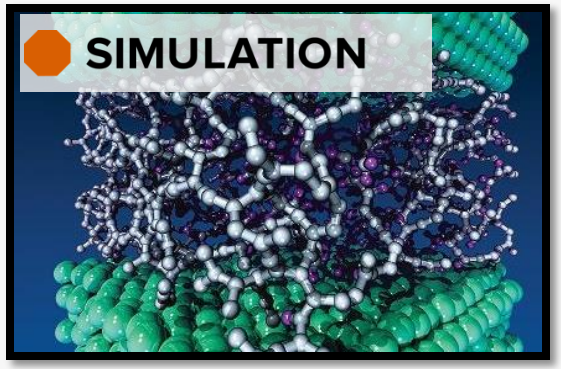
Argonne Tandem Linear Accelerator System

Atmospheric Radiation Measurement – The Southern Great Plains

Center for Nanoscale Materials



Argonne Leadership Computing Facility



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 offers different pipelines based on your computational readiness

2024 ALCF Annual Report

Argonne NATIONAL LABORATORY

ALCF at a Glance in 2023

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

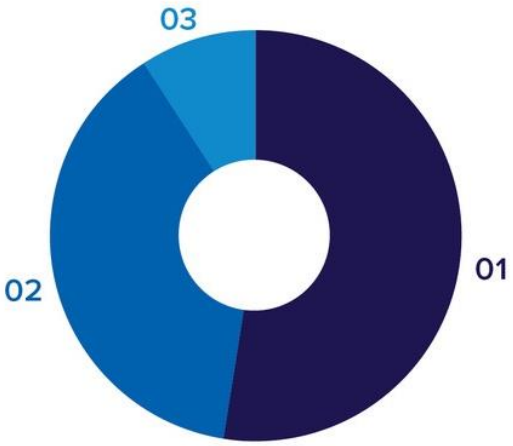
35.7M node-hours of compute time

417 active projects

1,624 facility users

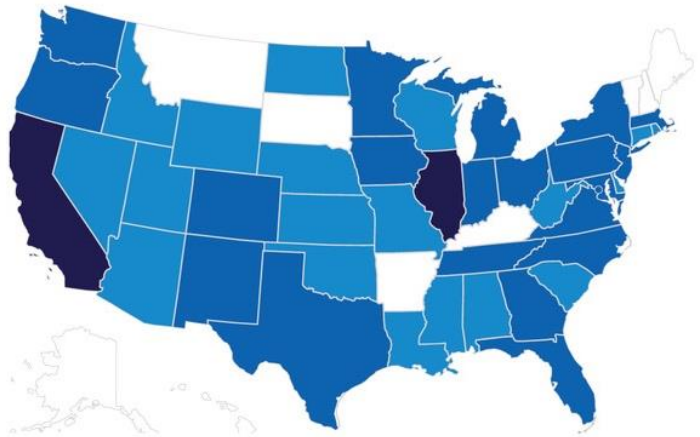
230+ publications

2023 ALCF Users by Affilliation

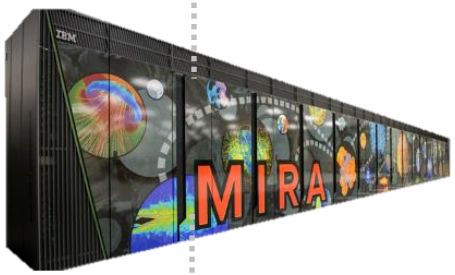


01	Academia	857
02	Government	620
03	Industry	147

2023 U.S. ALCF Users by State



ALCF Systems Evolution



5.7 TF
IBM BG/L
2004

557 TF
Intrepid
IBM BG/P
2007

10 PF
Mira
IBM BG/Q
2012

11.7 PF
Theta
Intel-Cray XC40
2017

15.6 PF
ThetaGPU
NVIDIA
DGX A100
2020

34 PF
Crux
HPE-AMD
+
Polaris
HPE
2021

2 EF
Aurora
Intel-HPE
2022

O(4 EF)
ALCF-4
2028

Helios

JLSE (2013)

AI Testbed (2020)

Edge Testbed (2021)

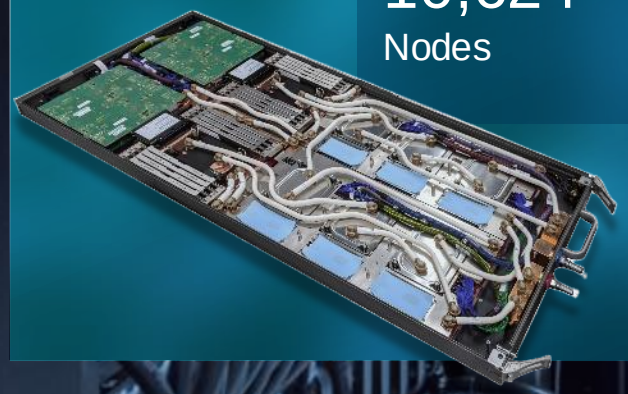
Aurora Specifications

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
- 10.6EF/s
- ~89% of the system

Top500

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

HPCG

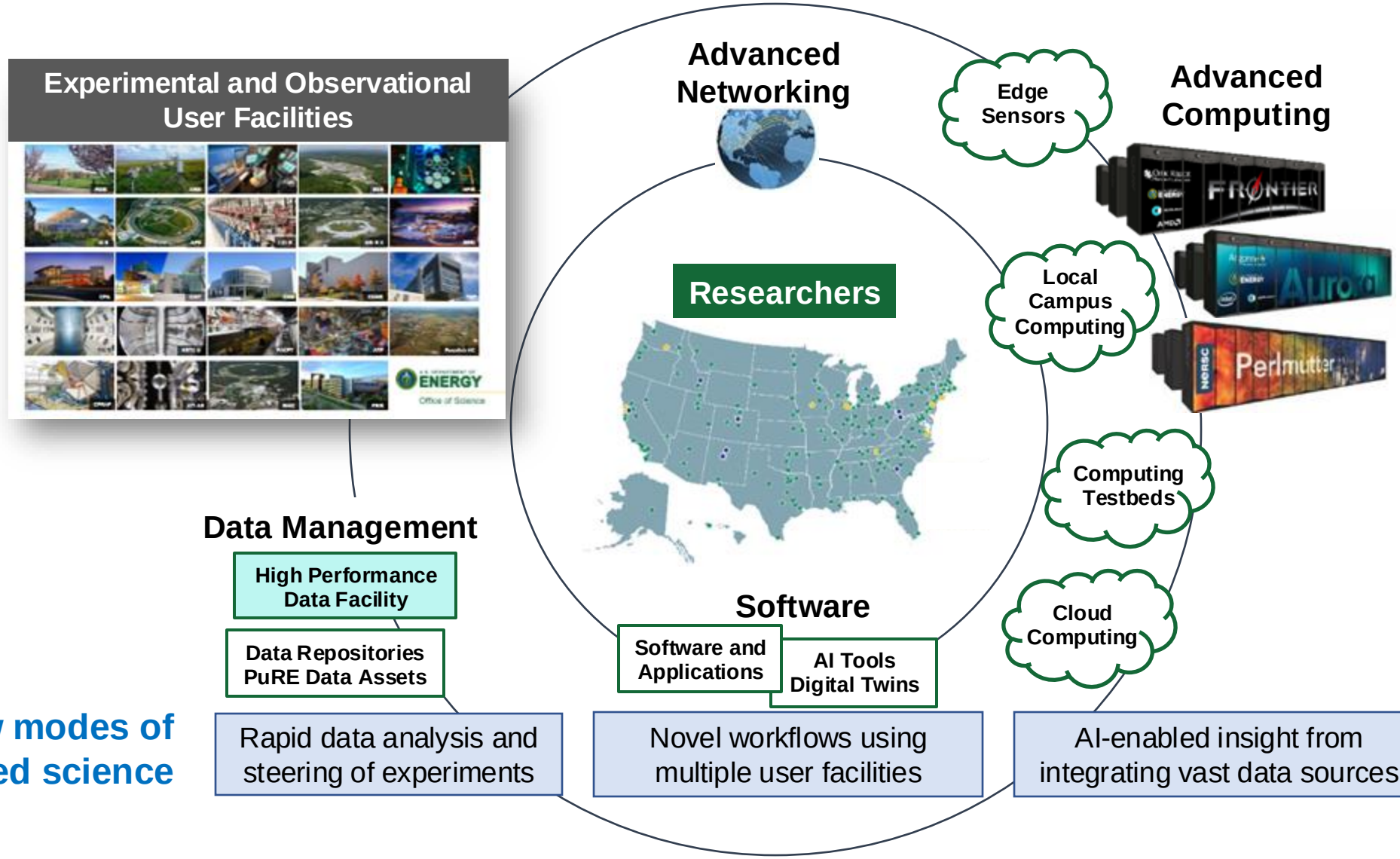
- #3 Ranking
- 5,612. TF/s
- ~38.5% of the system

Graph500

- #5 Ranking
- 24,250 GTEPS

DOE's Integrated Research Infrastructure (IRI) Vision:

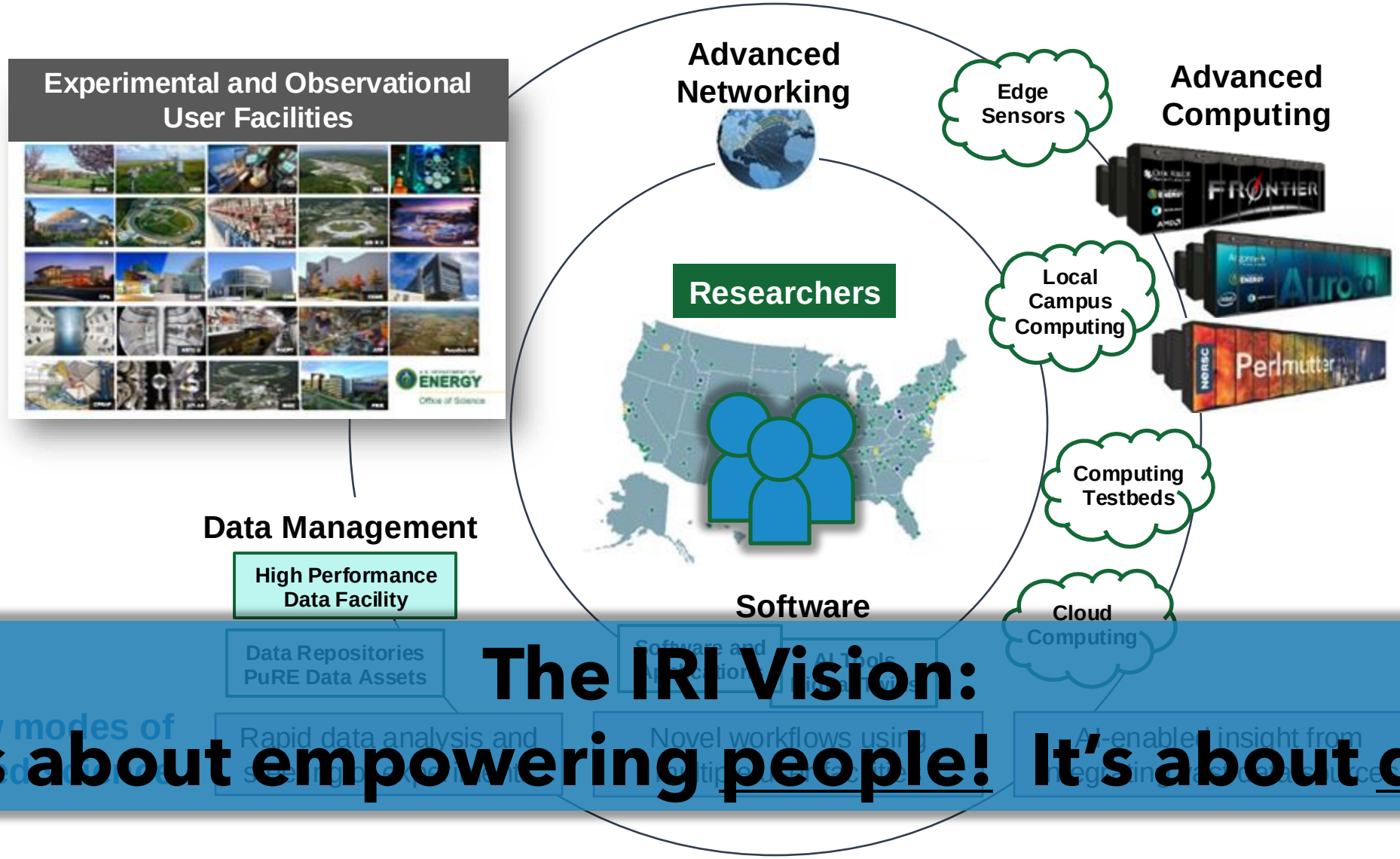
To empower researchers to meld DOE's world-class research tools, infrastructure, and user facilities seamlessly and securely in novel ways to radically accelerate discovery and innovation



Slide Credit: Ben Brown (DOE/SC/ASCR)

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

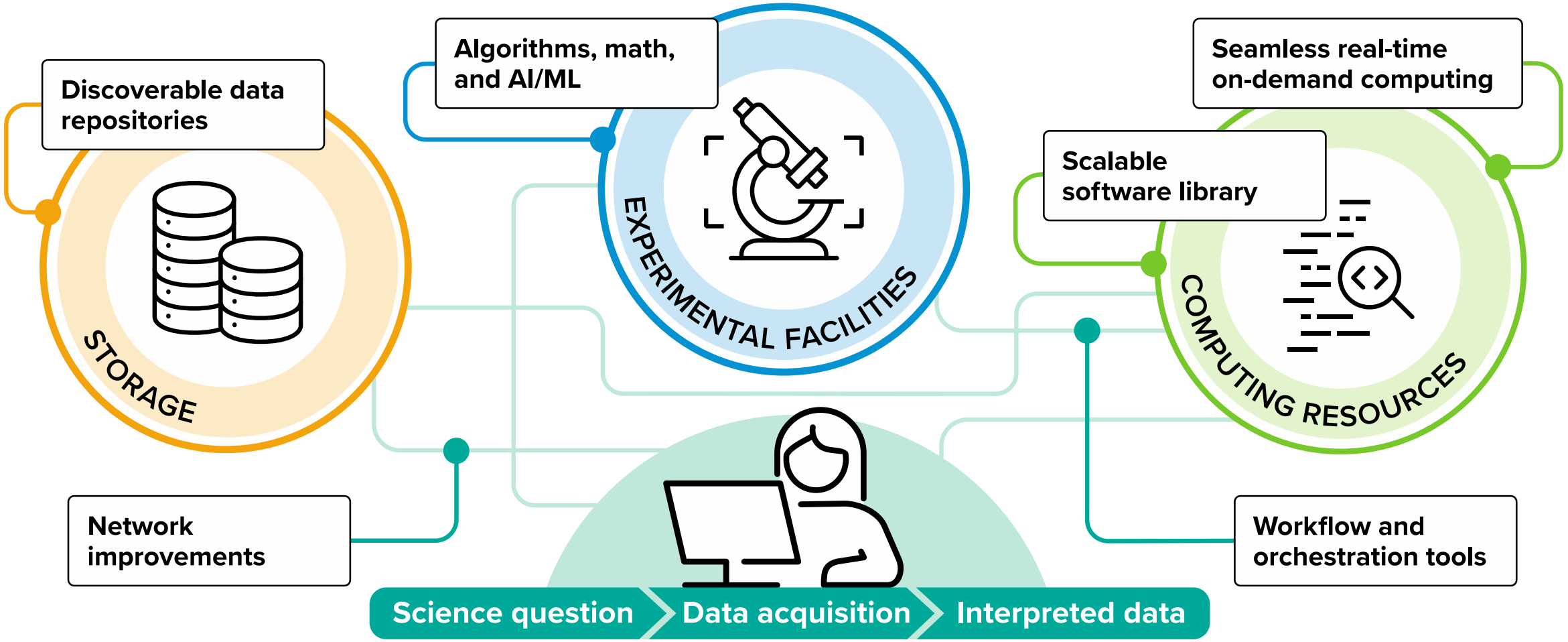


The IRI Vision:

It's about empowering people! It's about data!

Slide Credit: Ben Brown (DOE/SC/ASCR)

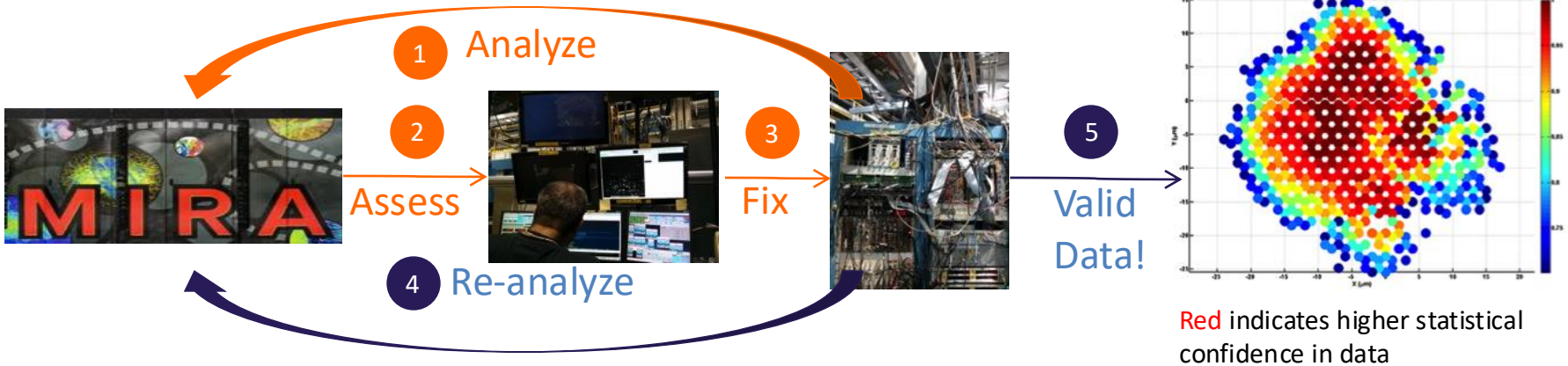
Integrated Research Infrastructure



Argonne National Laboratory's **Nexus** Effort (https://www.anl.gov/nexus-connect)

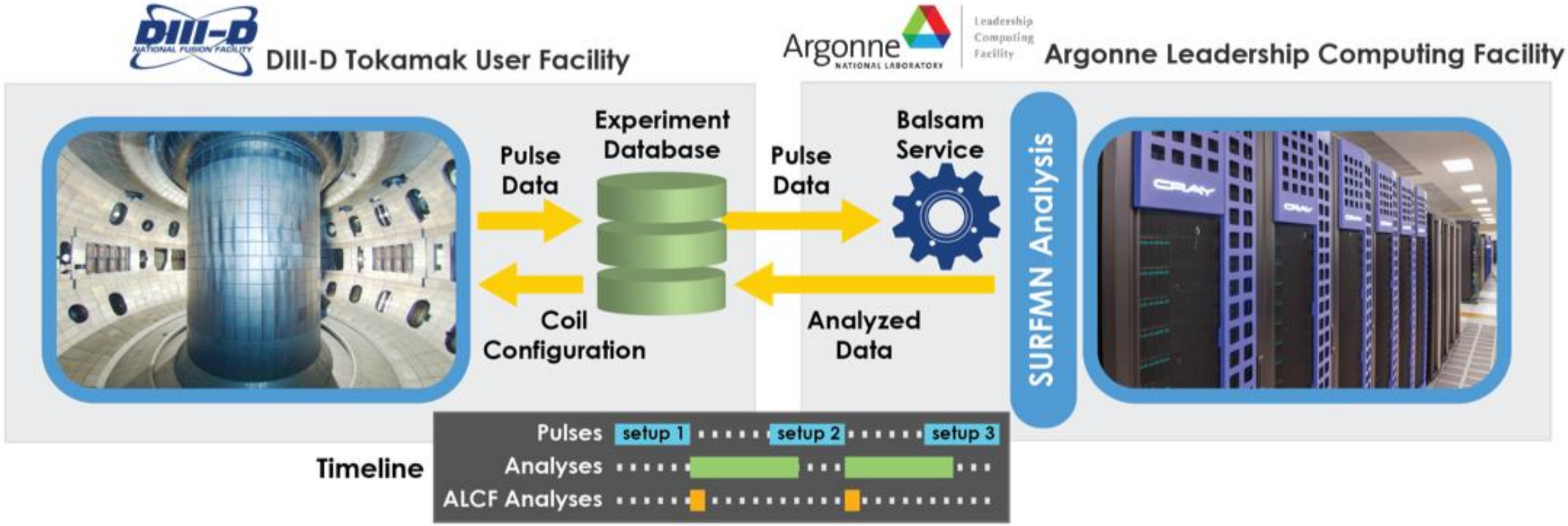
Experiments Integrating Research Infrastructure

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



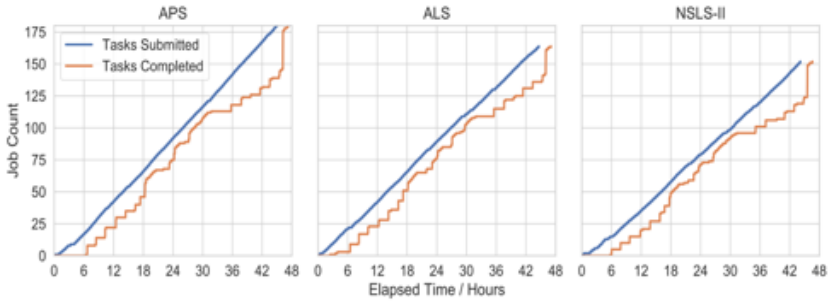
M. Kostuk, T. D. Uram, T. Evans, D. M. Orlov, M. E. Papka and D. Schissel, *Automatic Between-Pulse Analysis of DIII-D Experimental Data Performed Remotely on a Supercomputer at Argonne Leadership Computing Facility*, **Fusion Science and Technology**, 74(2), 2018.

Experiments Integrating Research Infrastructure

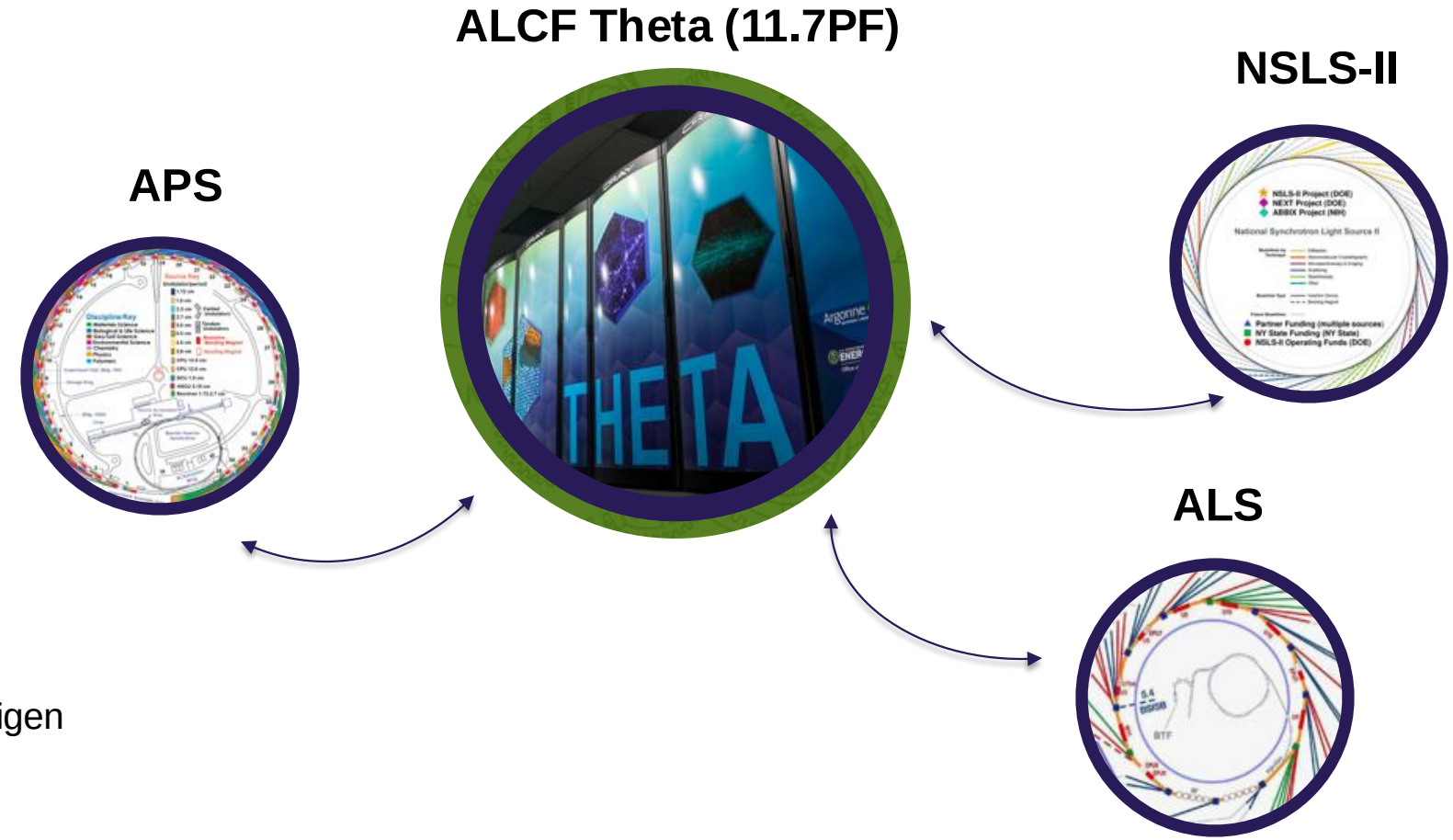
Utilizing leadership computing facility for continuous near real-time XPCS data processing

Balsam enables transparent access to remote resources

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

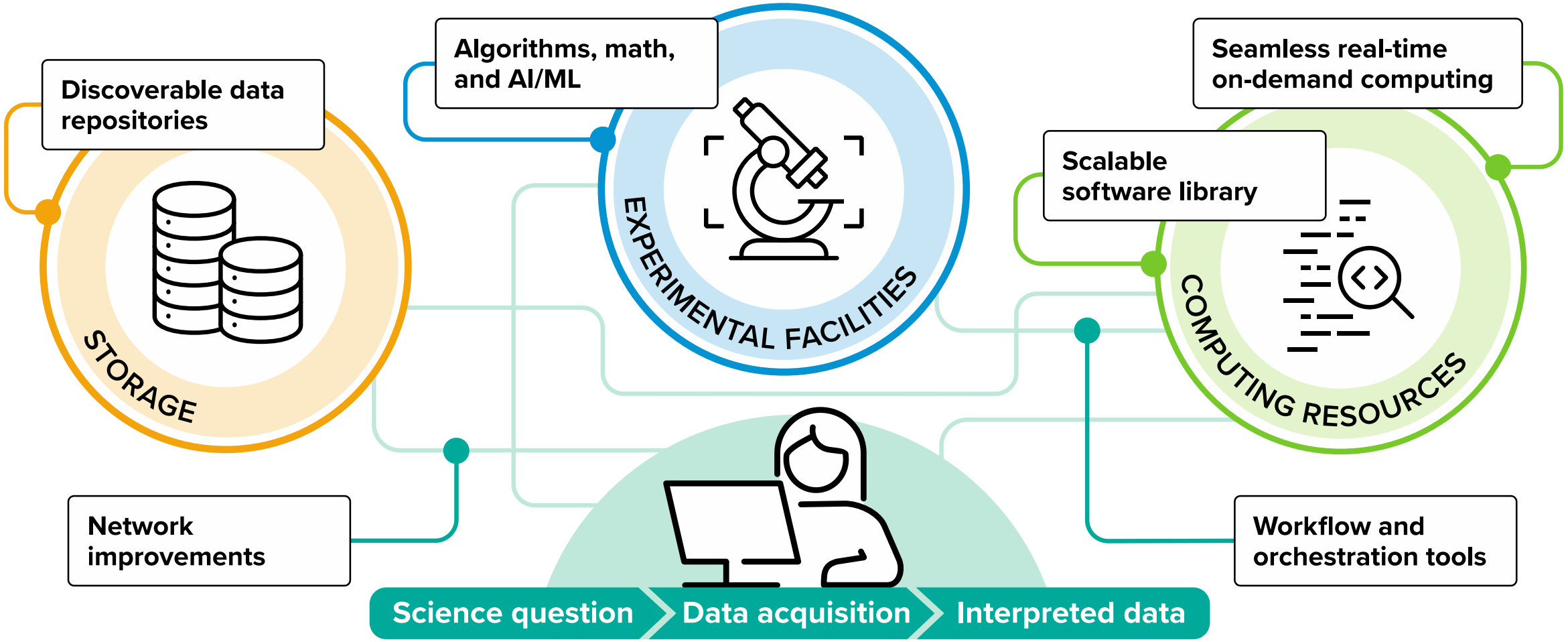


- Continuously executed for 48+ hours
- Process data in near real-time with XPCS-Eigen using special queue on Theta



M. Salim, T. Uram, J. T. Childers, V. Vishwanath, M. 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.

Integrated Research Infrastructure



Argonne National Laboratory's **Nexus** Effort (https://www.anl.gov/nexus-connect)

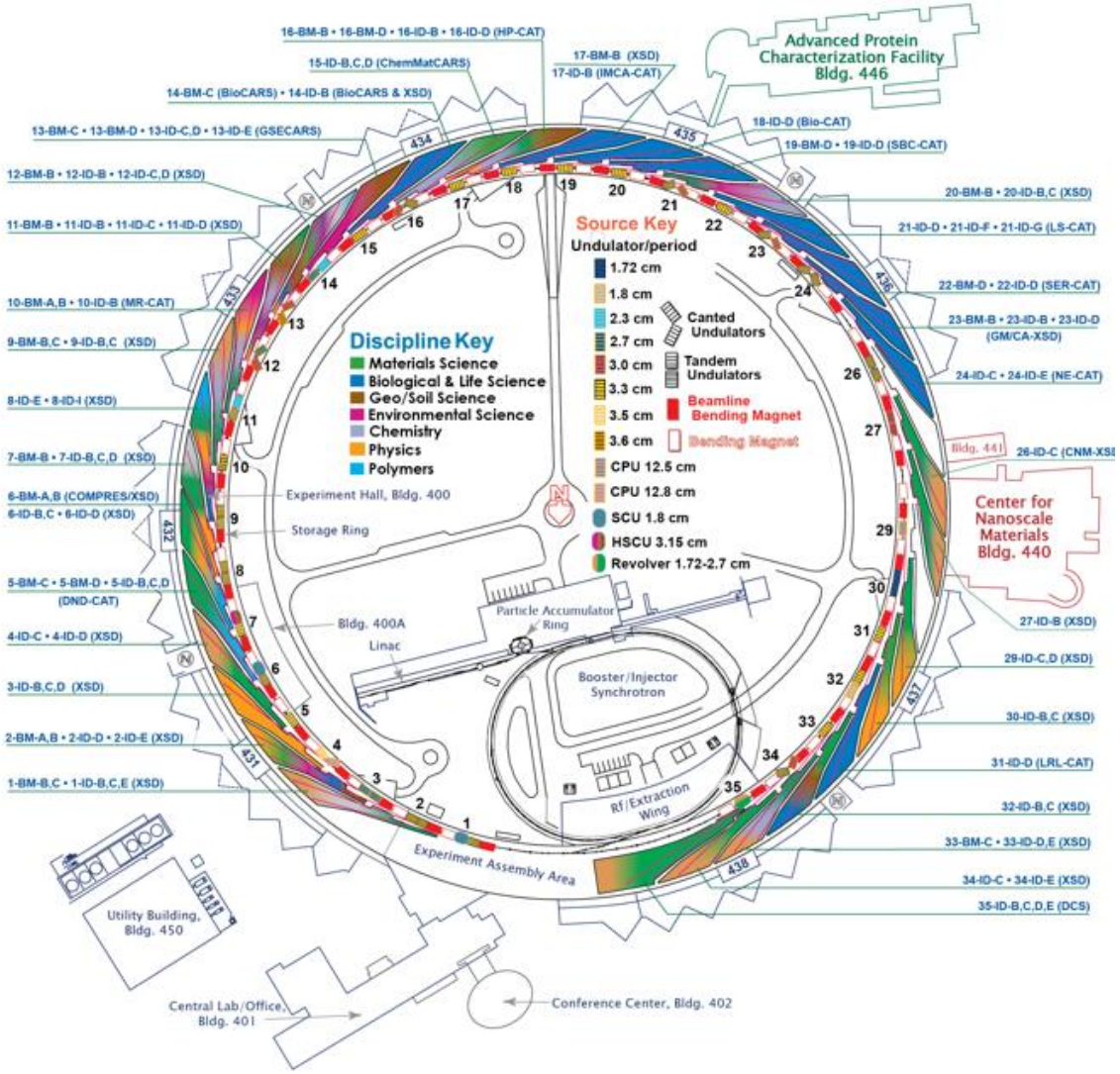
Advanced Photon Source

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)

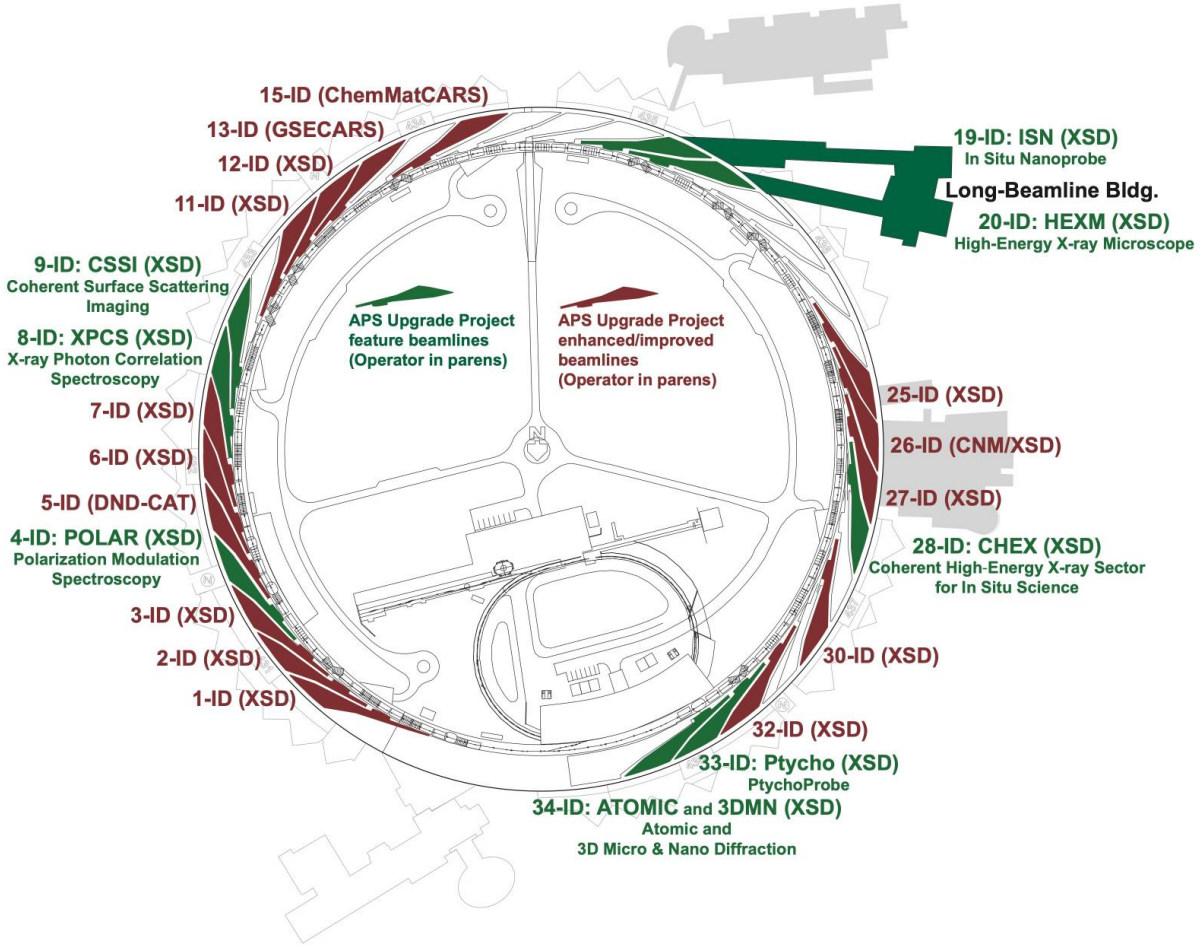


Slide Credit: Nicholas Schwartz (ANL/APS)

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)

Slide Credit: Nicholas Schwartz (ANL/APS)

Scale of the Challenge

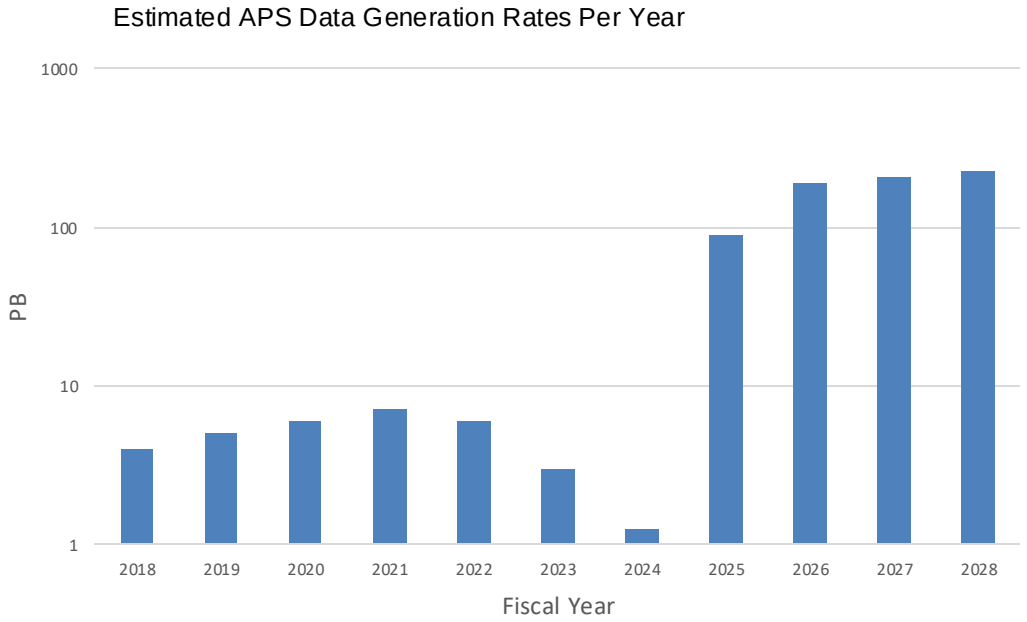
Multiple order-of-magnitude increase in demand for computing resources over next decade

APS-U Era

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

Over the next decade the APS will

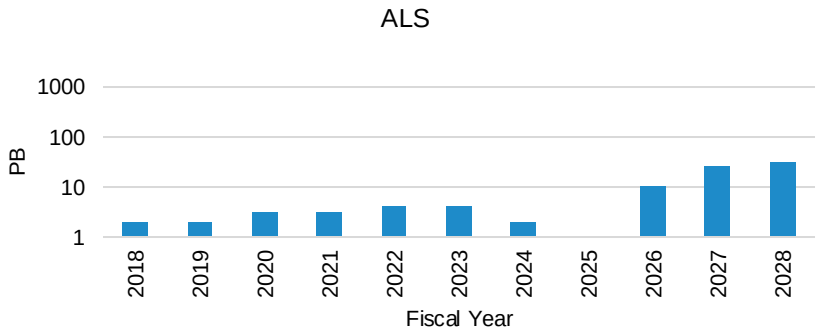
- Generate 100s of petabytes (PBs) of raw data per year
- Require 10s of PFLOP/s of on-demand computing power for first pass data processing and reduction

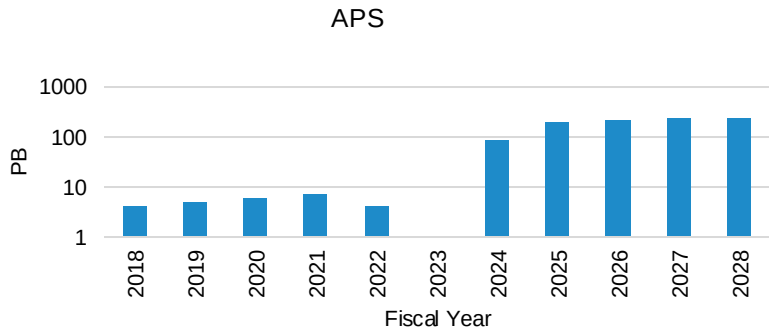


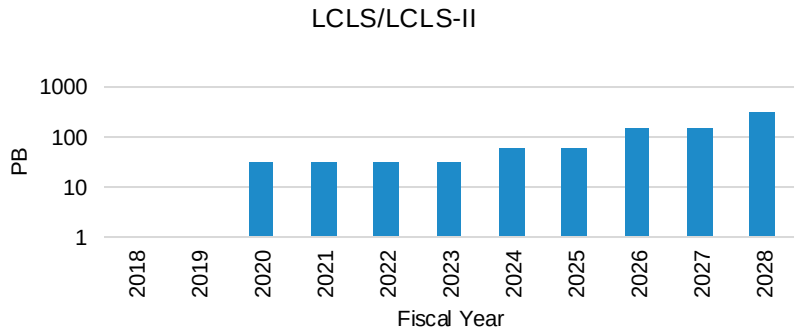
Log scale: Anticipated aggregate APS X-ray Science Division data generation per year. Data generation during FY23 is estimated to be at 50% of the peak due to the beginning of the storage ring replacement period. Data generation during FY24 is considerably lower due to the storage ring replacement period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

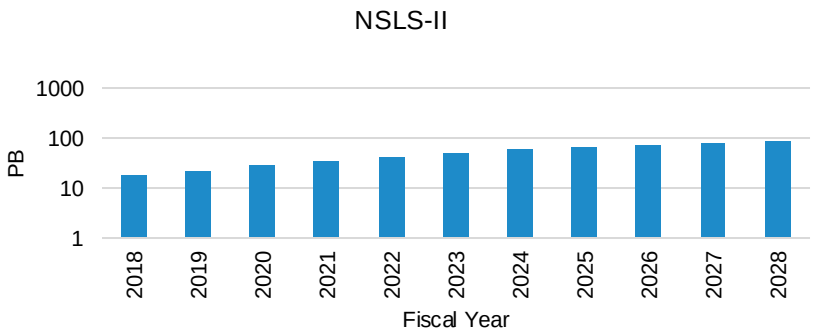
Slide Credit: Nicholas Schwartz (ANL/APS)

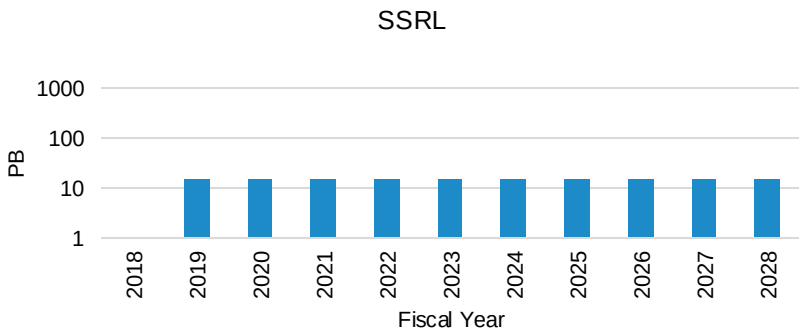
US Light Source Data Generation Estimates

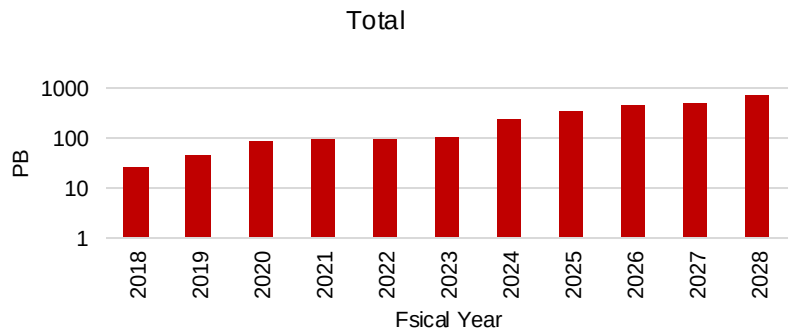








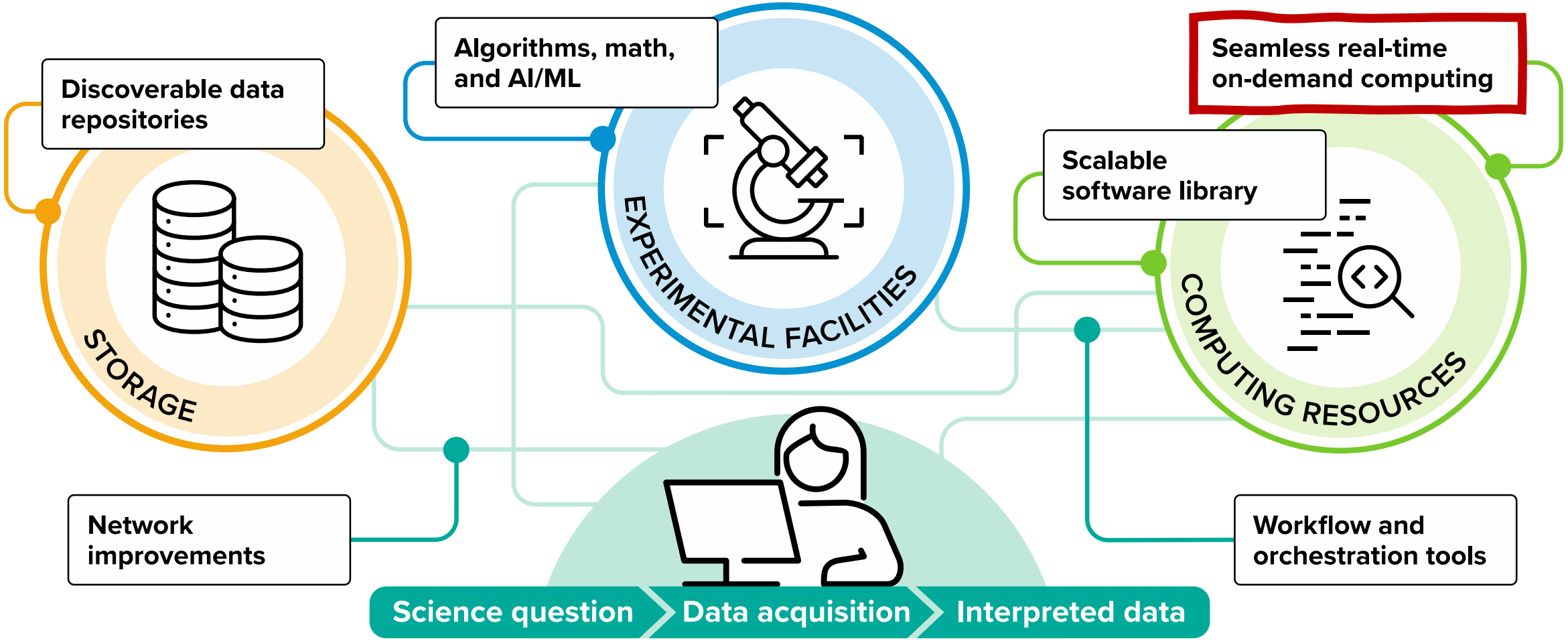




Log Scale: Estimated data generation rates per year at the US Department of Energy, Basic Energy Sciences, light sources. At the ALS and APS, data generation will stop during 2025 and 2023, respectively, due to installations of new storage rings. Aggregate data generation across the light sources will 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.

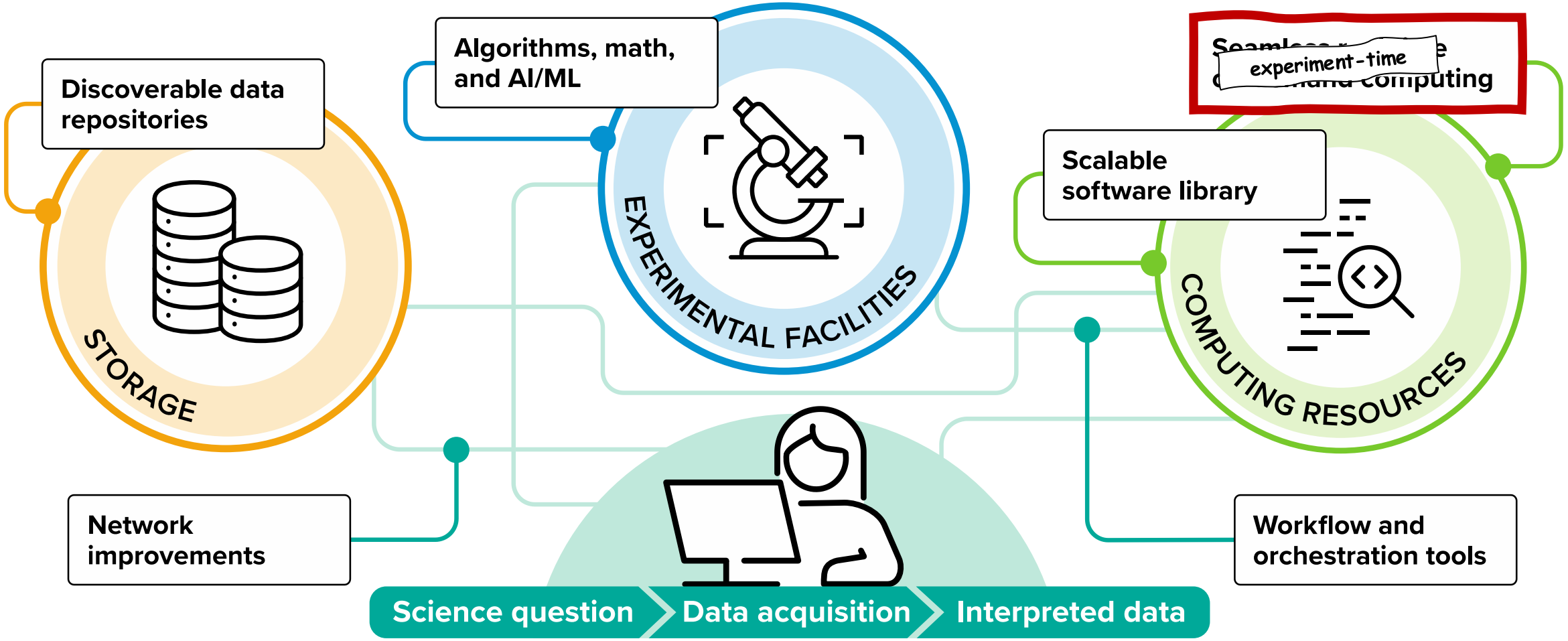
Slide Credit: Nicholas Schwartz (ANL/APS)

Integrated Research Infrastructure



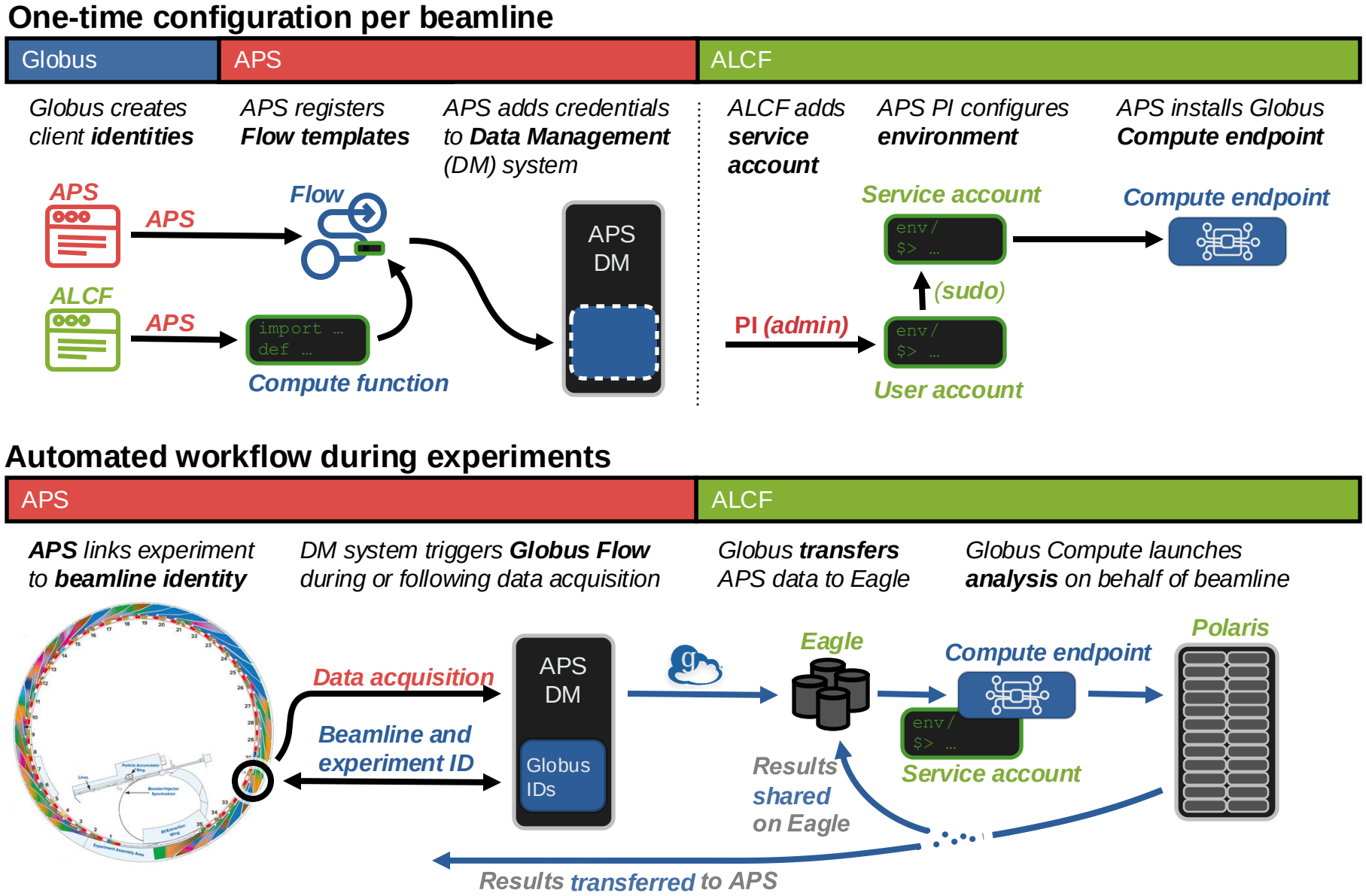
Argonne National Laboratory's **Nexus** Effort (https://www.anl.gov/nexus-connect)

Integrated Research Infrastructure



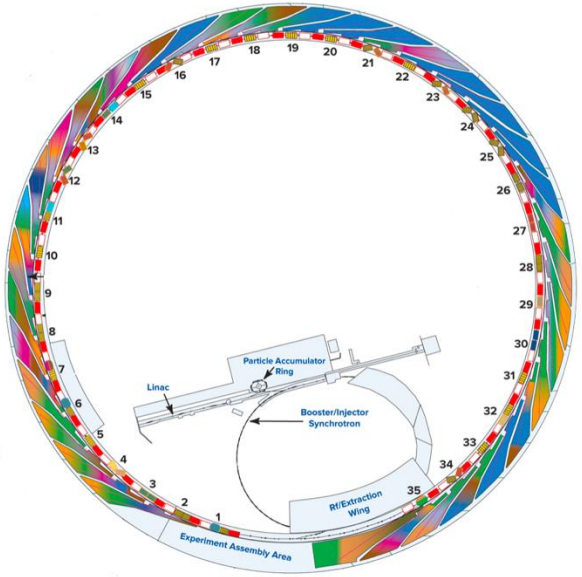
Argonne National Laboratory's **Nexus** Effort (https://www.anl.gov/nexus-connect)

Putting the Research Infrastructure Together (Nexus)

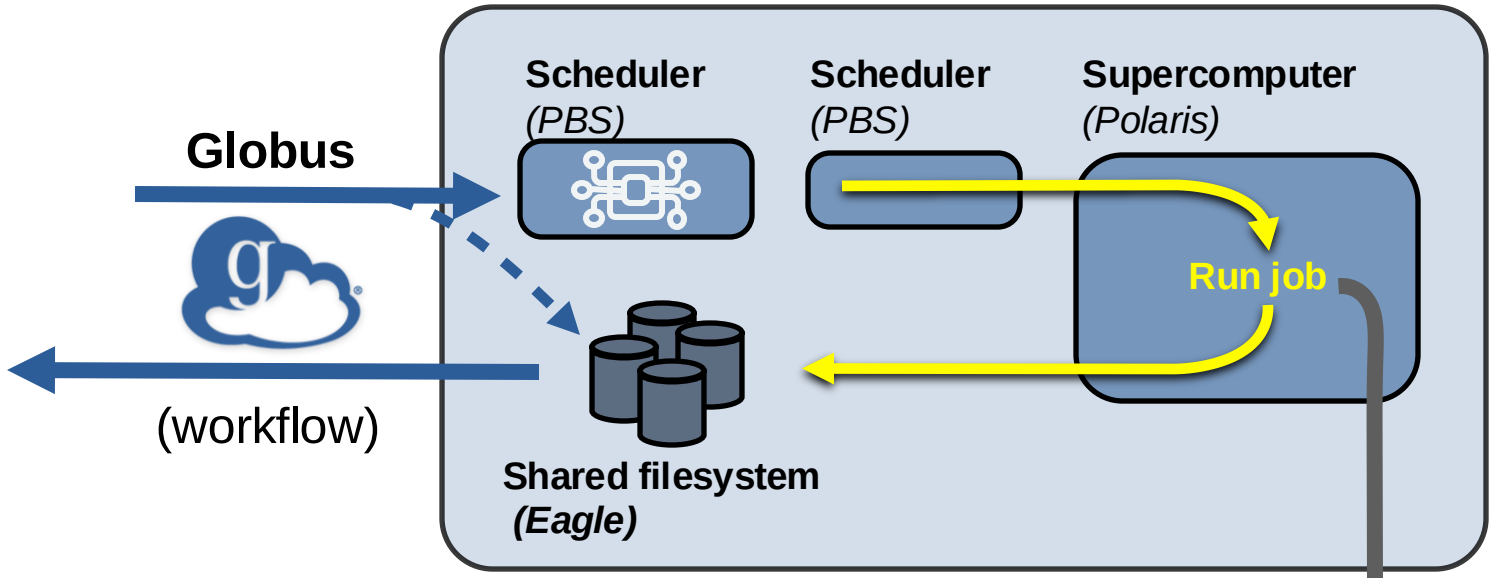


No Human in the Loop Experiment

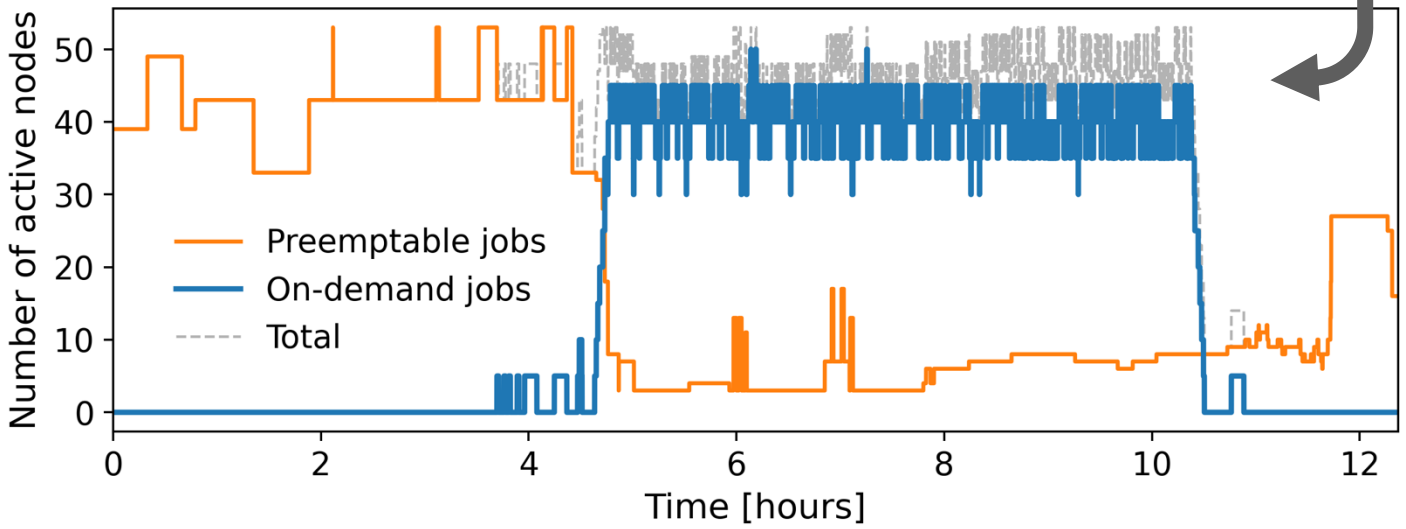
APS (data acquisition)



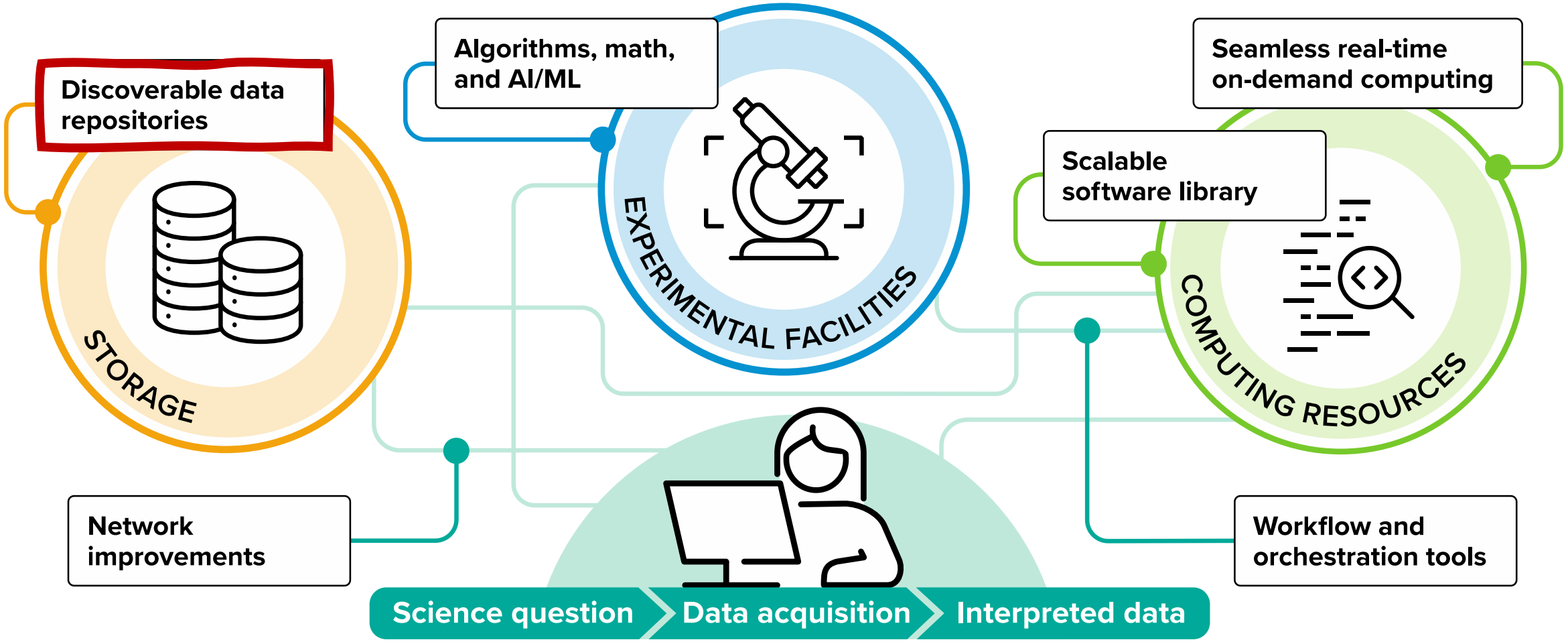
ALCF (on-demand data analysis)



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

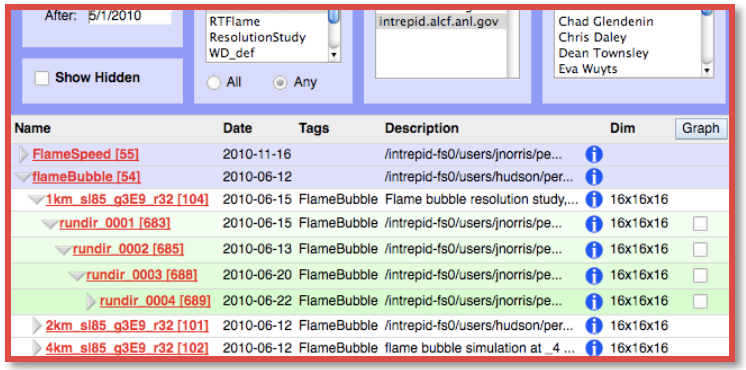


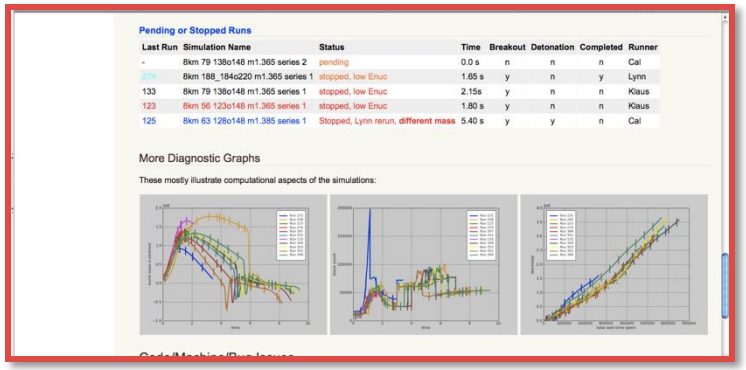
Integrated Research Infrastructure

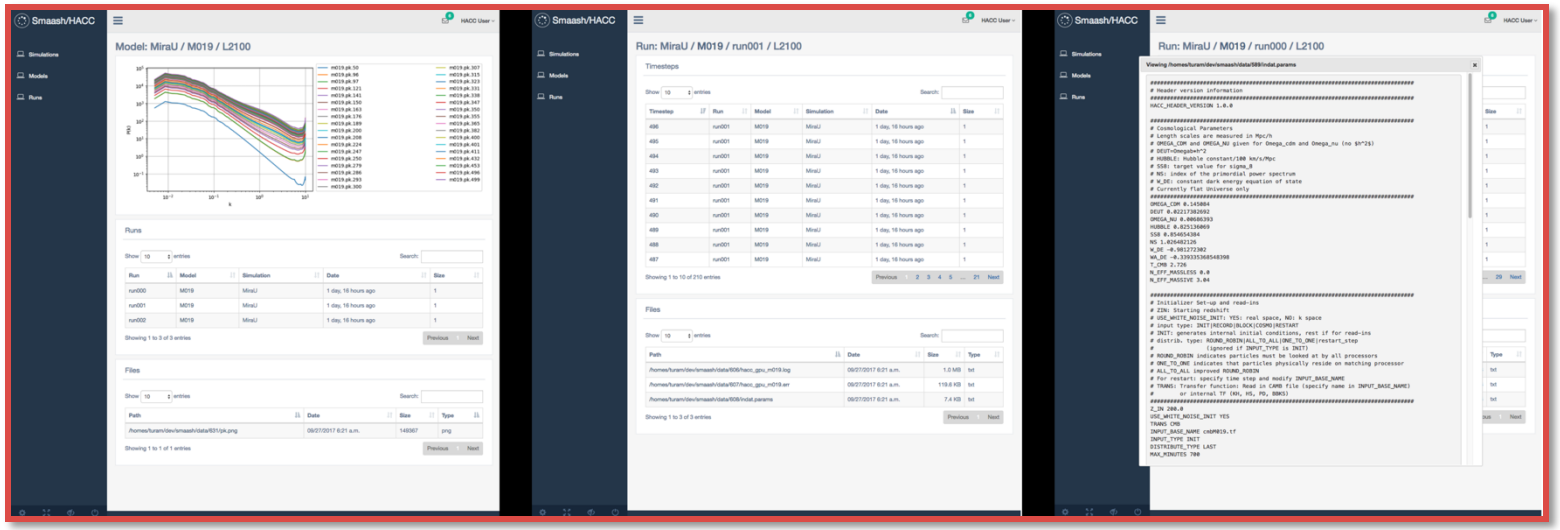


Argonne National Laboratory's **Nexus** Effort (https://www.anl.gov/nexus-connect)

Data Sharing and Community Resources

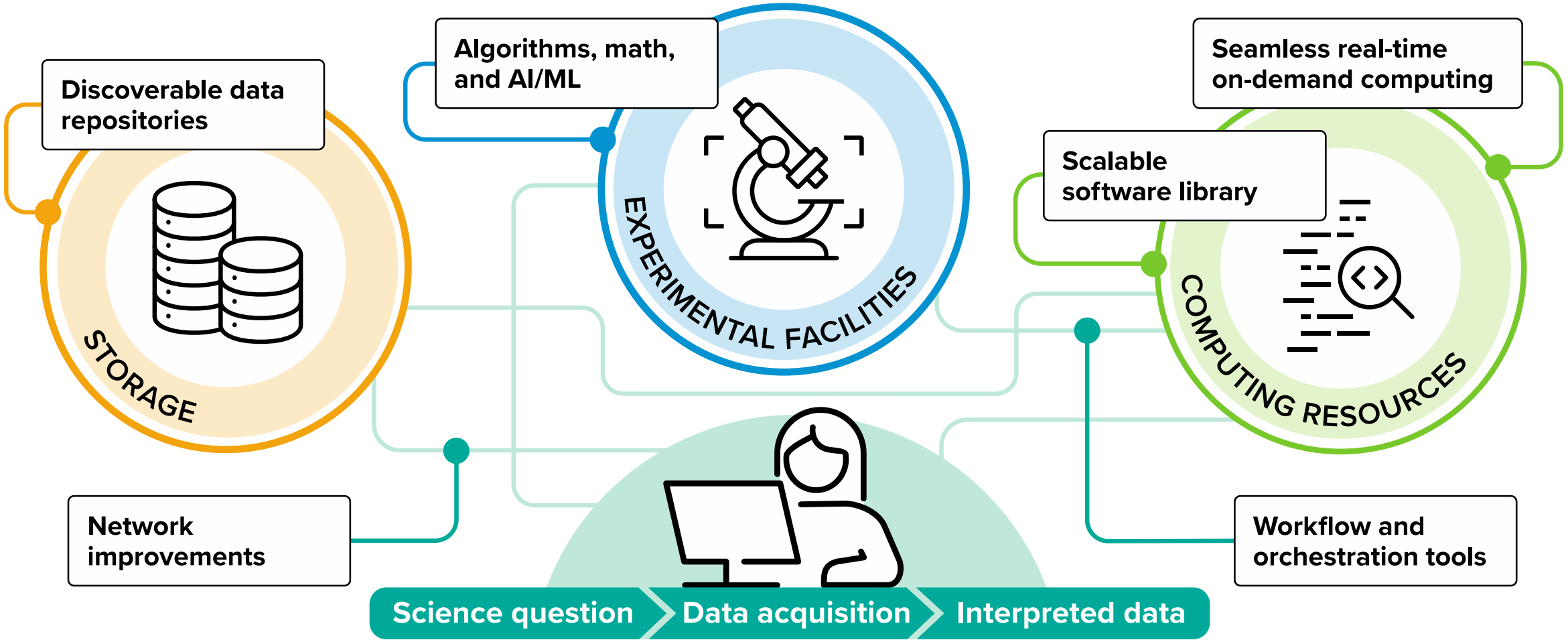






Katrin Heitmann, Thomas D. Uram, Hal Finkel, Nicholas Frontiere, Salman Habib, Adrian Pope, Esteban Rangel, Joseph Hollowed, Danila Korytov, Patricia Larsen, Benjamin S. Allen, Kyle Chard, and Ian Foster, *HACC Cosmological Simulations: First Data Release*, **The Astrophysical Journal Supplement Series**,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

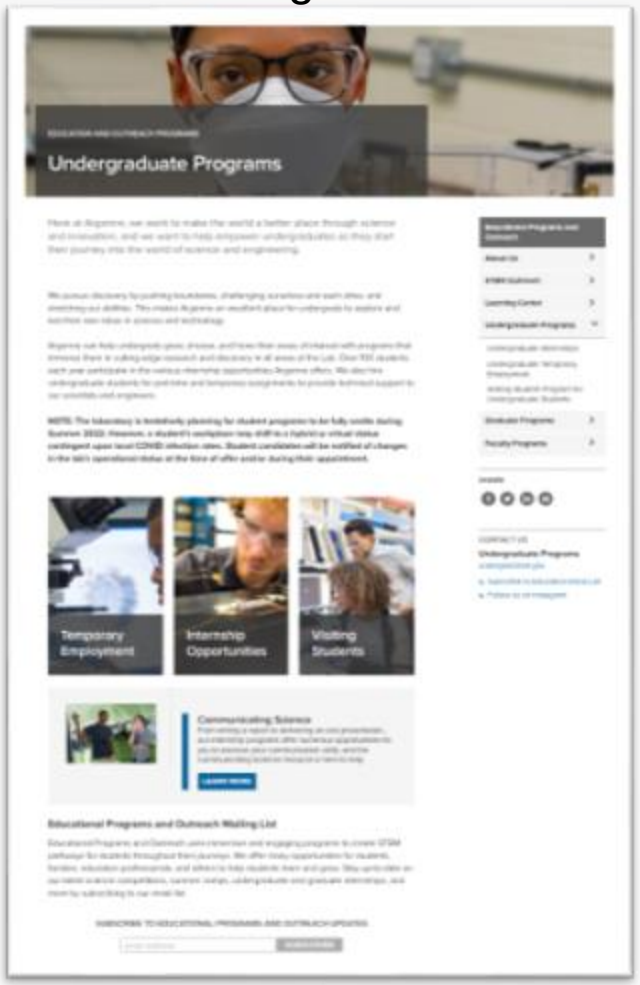
Integrated Research Infrastructure



Argonne National Laboratory's **Nexus** Effort (https://www.anl.gov/nexus-connect)

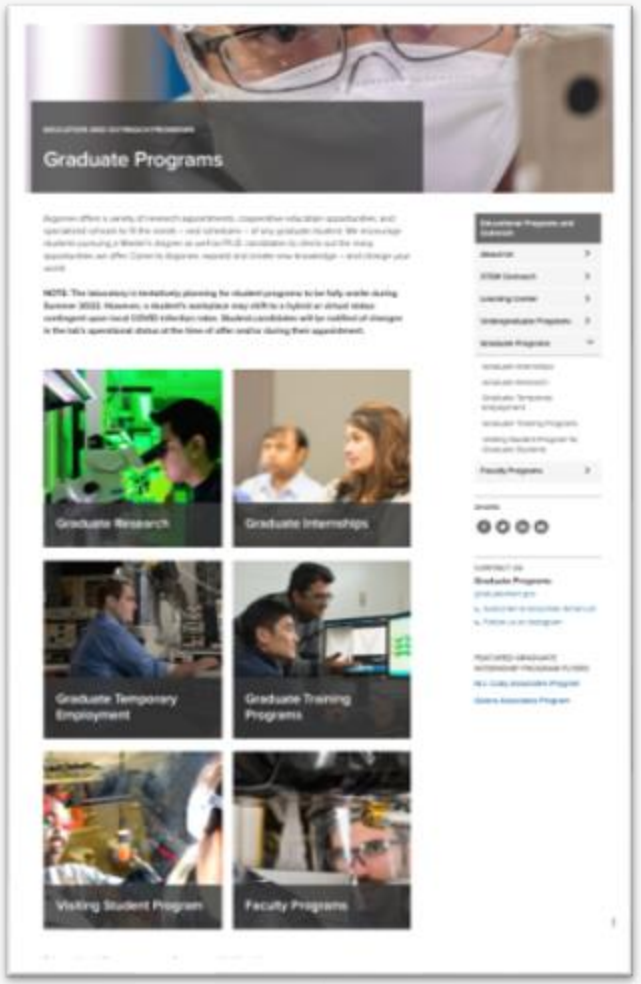
Opportunities at Argonne

Undergraduate



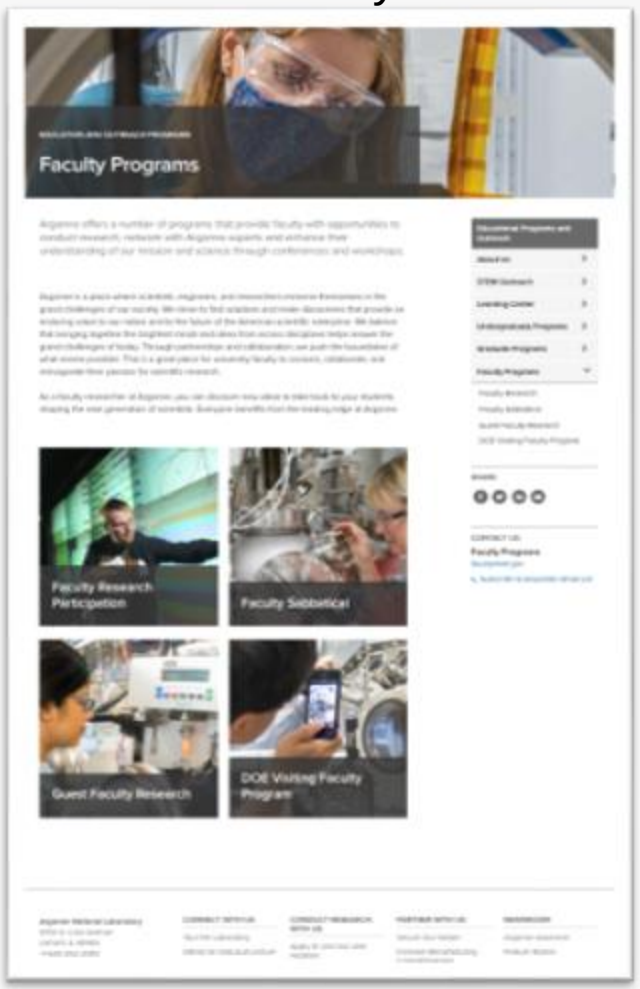
https://www.anl.gov/education/undergraduate-programs

Graduate



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

Faculty



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

Argonne NATIONAL LABORATORY


55
50
EXIT

UIC ENGINEERING
Computer Science

Argonne NATIONAL LABORATORY

Thank You