

Bringing supercomputers and experiments together to accelerate discoveries

January 31 2024, by Jim Collins



Researchers are using the Polaris supercomputer to test and demonstrate capabilities for integrating experiments with high performance computing resources. Credit: Argonne National Laboratory.

The seamless integration of DOE's large-scale experimental facilities and powerful computing resources will help researchers keep pace with the ever-increasing influx of scientific data.

When the massive upgrade at the Advanced Photon Source (APS) at the U.S. Department of Energy's (DOE) Argonne National Laboratory is completed later this year, experiments at the powerful [X-ray light source](#) are expected to generate 100–200 petabytes, or 100-200 million gigabytes, of [scientific data](#) per year.

That's a substantial increase over the approximately five petabytes that

were being produced annually at the APS, a DOE Office of Science user facility at Argonne, before the upgrade. And if you consider the DOE's four other light sources, the facilities are projected to yield an exabyte, or 1 billion gigabytes, of data per year in the coming decade.

"An exabyte of data is equivalent to streaming 1.5 million movies every day for a year," said Nicholas Schwarz, Argonne computer scientist and lead for scientific software and data management at the APS.

"But we need to do a lot more than simply move a lot of data around. For the X-ray experiments carried out at the APS, we need to use advanced computational tools to look at every pixel of every frame, analyze the data in near real time, and use the results to make decisions about the next experiment."

"To process all this data quickly, we require a lot of computing capabilities, from big computers and data storage, to [analysis software](#), to the computational fabric that ties all of these resources together," he added.

The growing deluge of scientific data is not unique to light sources. Telescopes, particle accelerators, fusion research facilities, remote sensors and other scientific instruments also produce large amounts of data. And as their capabilities improve over time, the data generation rates will only continue to grow.

"The scientific community's ability to process, analyze, store and share these massive datasets is critical to gaining insights that will spark new discoveries," said Michael E. Papka, Argonne deputy associate laboratory director for Computing, Environment and Life Sciences.

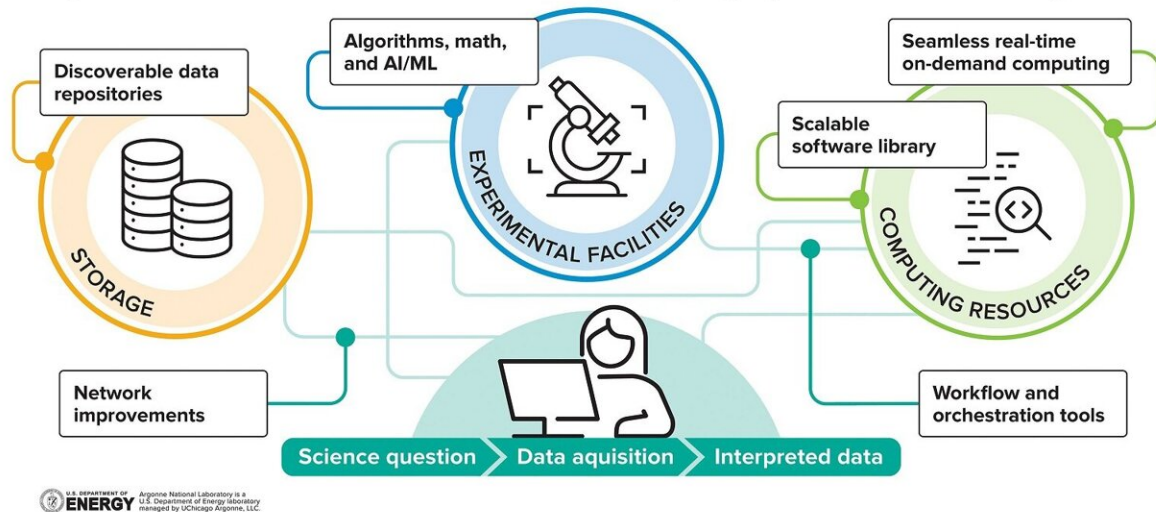
Papka also serves as director of the Argonne Leadership Computing Facility (ALCF), a DOE Office of Science user facility at Argonne, and

is a professor of computer science at the University of Illinois Chicago.

INTEGRATED RESEARCH INFRASTRUCTURE



Argonne's Nexus effort connects scientific facilities, supercomputing capabilities and data technologies.



Argonne's Nexus effort is working to develop an integrated research infrastructure that connects scientific facilities, supercomputing capabilities and data technologies. Credit: Argonne National Laboratory

Argonne's [Nexus effort](#) is playing a pivotal role in advancing DOE's vision to build an integrated research infrastructure (IRI). Developing an IRI would accelerate data-intensive research by seamlessly integrating DOE's cutting-edge experimental facilities with its world-class supercomputing, artificial intelligence (AI) and data resources.

For over a decade, Argonne has been working to develop tools and methods to connect its powerful computing resources with large-scale experiments. [Merging ALCF supercomputers with the APS](#) has been a

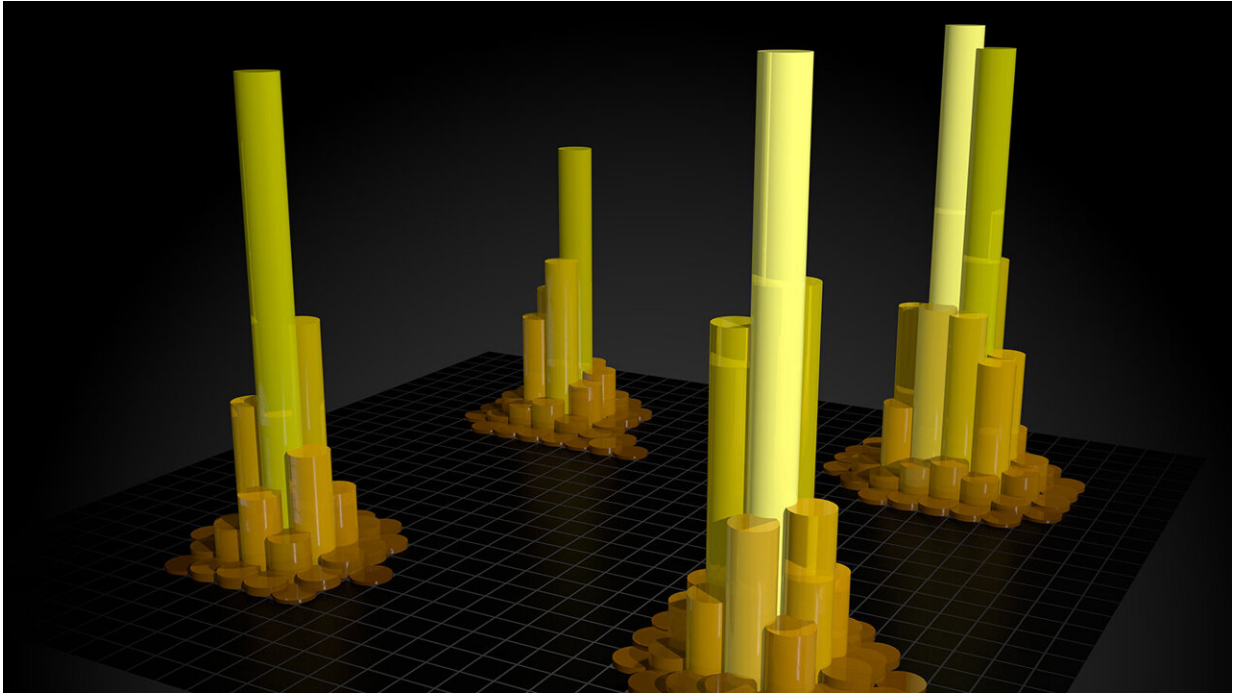
significant focus of the lab's IRI-related research, but the work has also included collaborations with the DIII-D National Fusion Facility in California and CERN's Large Hadron Collider in Switzerland. DIII-D is a DOE Office of Science user facility hosted by General Atomics.

"We've been partnering with experimental facilities for several years now to help them use our supercomputing resources to process huge amounts of data more quickly," Papka said. "With the launch of Nexus, we have a vehicle to coordinate all of our research and collaborations in this space to align with DOE's broader efforts to lead the new era of integrated science."

Argonne's ongoing work has led to the creation of tools for managing computational workflows and the development of new capabilities for on-demand computing, giving the lab valuable experience to support the DOE IRI initiative. [Globus](#) and the [ALCF Community Data Co-Op \(ACDC\)](#) are critical resources in enabling the IRI vision.

Globus, a research automation platform created by researchers at Argonne and the University of Chicago, is used to manage high-speed data transfers, computing workflows, data collection and other tasks for experiments. ACDC provides large-scale data storage capabilities, offering a portal that makes it easy to share data with external collaborators across the globe.

The ALCF's upcoming Aurora exascale supercomputer will also bolster the lab's IRI efforts, providing a significant boost in computing power and advanced capabilities for AI and data analysis.



Scientific visualization of Bragg diffraction peaks in a 15x15 pixel patch of an undeformed bi-crystal gold sample. The height denotes photon counts. This data was acquired at the APS and processed on ALCF supercomputers. Credit: Argonne National Laboratory

Streamlining science

The IRI will not only enable experiments to analyze vast amounts of data, but it will also allow them to process large datasets quickly for rapid results. This is crucial as experiment-time analysis often plays a key role in shaping subsequent experiments.

For the Argonne-DIII-D collaboration, researchers demonstrated how the close integration of ALCF supercomputers could benefit a fast-paced experimental setup. Their work centered on a fusion experiment that used a series of plasma pulses, or shots, to study the behavior of plasmas

under controlled conditions.

The shots were occurring every 20 minutes, but the data analysis required more than 20 minutes using their local computing resources, so the results were not available in time to inform the ensuing shot. DIII-D researchers teamed up with the ALCF to explore how they could leverage supercomputers to speed up the analysis process.

"Every time they took a shot, we started a job at the ALCF. It fetched the data from DIII-D, ran the analysis, and pushed the results back to them in time to calibrate the next shot," said Thomas Uram, Argonne computer scientist and the IRI lead at the ALCF.

"Because we had more computing power than DIII-D had available locally, we could analyze their data faster and at a resolution 16 times greater than their in-house systems. Not only did they get the results in advance of the next shot, they also got significantly higher resolution analyses to improve the accuracy of their configuration."

Many experiments at the APS will also benefit from near-real-time data analysis, including battery research, the exploration of materials failure and drug development.

"By getting analysis results in seconds or less instead of hours, days or even weeks, scientists can gain real-time insight into their experiments as they occur," Schwarz said. "Researchers will be able to use this feedback to steer an experiment and zoom in on a particular area to see critical processes, like the molecular changes that occur during a battery's charge and discharge cycles, as they are happening."

A fully realized IRI would also impact the people conducting the research. Scientists must often devote considerable time and effort to managing data when running an experiment. This includes tasks like

storing, transferring, validating and sharing data before it can be used to gain new insights.

"The IRI vision is to automate many of these tedious [data management](#) tasks so researchers can focus more on the science," Uram said. "This would substantially streamline the [scientific process](#), freeing up scientists so they have more time to form hypotheses while experiments are being carried out."

Supercomputing on demand

Getting instant access to DOE supercomputers for data analysis requires a shift in how the computing facilities operate. Each facility has established policies and processes for gaining access to machines, setting up user accounts, managing data and other tasks.

"If a researcher is set up at one computing facility but needs to use supercomputers at the other facilities, they would have to go through a similar set of steps again for each site," Uram said. "And that takes time. It takes time away from doing actual science."

Once a project is set up, researchers submit their "job" to a queue, where they wait their turn to run on the supercomputer. While the traditional queuing system helps optimize supercomputer usage at the facilities, it doesn't support the rapid turnaround times needed for the IRI.

To make things easy for the end users, the IRI will require implementing a uniform way for experimental teams to gain quick access to the DOE supercomputing resources.

To that end, Argonne has developed and demonstrated methods for overcoming both the user account and job scheduling challenges. The co-location of the APS and the ALCF on the Argonne campus has offered

an ideal environment for testing and demonstrating such capabilities.

When the ALCF launched the Polaris supercomputer in 2022, four of the system's racks were dedicated to advancing the integration efforts with experimental facilities.

In the case of user accounts, the existing process can get unwieldy for experiments involving several team members who need to use the computing facilities for data processing. The Argonne team has piloted the idea of employing "service accounts" that provide secure access to a particular experiment instead of requiring each team member to have an active account.

"This is important because many experiments have a team of people collecting data and running analysis jobs over the course of a few days or a week," Uram said. "We need a way to support the experiment independent of who is operating the instruments that day."

To address the job scheduling issue, the Argonne team has set aside a portion of Polaris nodes to run with "on-demand" and "preemptable" queues. This approach allows time-sensitive jobs to run on the dedicated nodes immediately.

The team has completed successful test runs of the service accounts and on-demand and preemptable queues on Polaris using data generated during an APS experiment. The runs were fully automated with no humans in the loop.

"This capability is truly exciting for the experimental integration efforts here at Argonne, but there is much work ahead to develop workable solutions that can be used across all DOE experimental and computing facilities," Papka said.

Bringing it all together

While Argonne and its fellow national labs have been working on projects to demonstrate the promise of an integrated research paradigm for the past several years, DOE's Advanced Scientific Computing Research (ASCR) program made it a more formal initiative in 2020 with the launch of the IRI Task Force.

Comprised of members from several national labs, including Argonne's Schwarz, Uram, Jini Ramprakash and Corey Adams, the task force identified the opportunities, risks and challenges posed by such an integration.

In 2022, ASCR launched the IRI Blueprint Activity to create a framework for implementing the IRI. The blueprint team, which included Schwarz and Ramprakash, released a [report](#) that describes a path forward from the lab's individual partnerships and demonstrations to a broader long-term strategy that will work across the DOE ecosystem.

Over the past year, the blueprint activities have started to formalize with the introduction of IRI testbed resources and environments. Now in place at each of the DOE computing facilities, the testbeds facilitate research to explore and refine IRI ideas in collaboration with teams from DOE experimental facilities.

"With the launch of the Nexus effort here at Argonne, we will continue to leverage our collective knowledge, expertise and resources to help DOE and the larger scientific community enable and scale this new paradigm across a diverse range of research areas, scientific instruments and user facilities," Uram said.

More information: William Miller et al, Integrated Research

Infrastructure Architecture Blueprint Activity (Final Report 2023). [DOI: 10.2172/1984466](https://doi.org/10.2172/1984466)

Provided by Argonne National Laboratory

Citation: Bringing supercomputers and experiments together to accelerate discoveries (2024, January 31) retrieved 27 April 2024 from <https://techxplore.com/news/2024-01-supercomputers-discoveries.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.