OCTOBER 21, 2025

TOWARD A PRODUCTIVE SOFTWARE ECOSYSTEM FOR SCIENTIFIC USER FACILITIES

HANNAH PARRAGA, RYAN CHARD, IAN FOSTER, NICHOLAS SCHWARZ

U.S. DEPARTMENT OF ENERGY
Argonne National Laboratory is a U.S. Department of Energy laboratory managed by UChicago Argonne, LLC.

Argonne NATIONAL LABORATORY

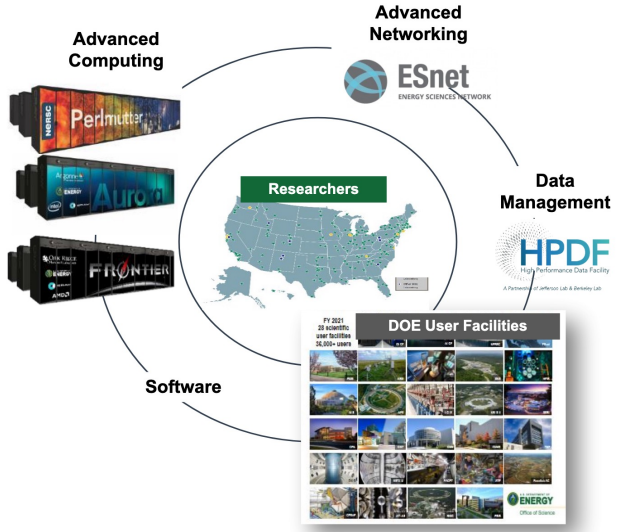
NATIONAL INITIATIVES DRIVING A CONNECTED SCIENCE ECOSYSTEM

Pairing advanced computational power across multiple sites with a flexible software ecosystem will enable researchers to accelerate innovation and unlock insights at scale.

Programs like the DOE Integrated Research Infrastructure (IRI), the American Science Cloud, and the Transformational AI Models Consortium offer new opportunities for seamless multi-site operations.

AmSC
AMERICAN SCIENCE CLOUD

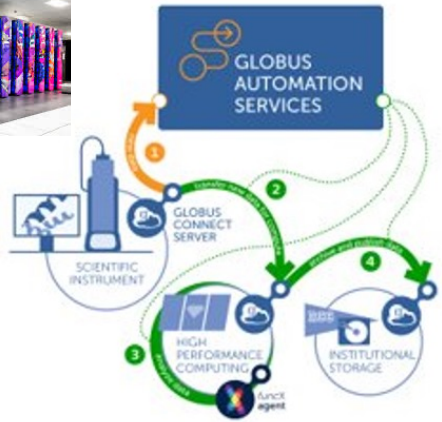
MODCoN



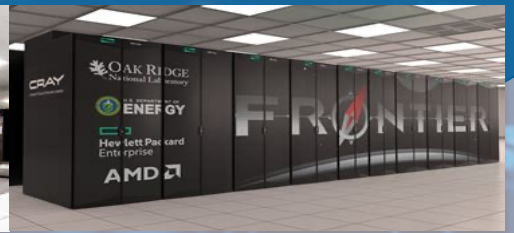
FOUNDATIONAL DEVELOPMENTS

- Secure data access and sharing via Globus and APS Data Management System
- Integrated data life cycle management reduces manual effort and frees beamtime for science
- APS collaborations with ALCF and NERSC connect experimental data to compute through:
  - Demand queue
  - Dedicated allocations
  - Service accounts
  - Globus Compute endpoints

DM
APS DATA MANAGEMENT SYSTEM



To enable experiment steering and streamline workloads that integrate distributed resources, we must close the loop between measurement and insight.



FORWARD LOOKING OPPORTUNITIES

CROSS FACILITY AWARENESS

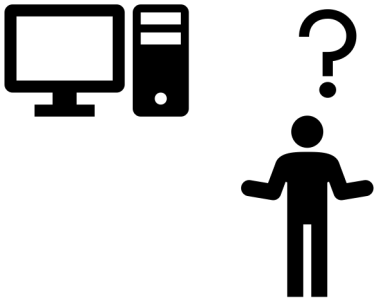
The Challenge

- Data can be processed at APS, ALCF, or NERSC
- Most users lack ALCF/OLCF/NERSC accounts or HPC expertise
- System abstracts HPC complexity from APS beamline staff and users
- Workflow selects machine by name (e.g. polaris, perlmutter)
- No visibility for APS users to estimate queue depth at each site

CROSS FACILITY AWARENESS

What's Needed

- Automate site selection based on estimated wait time
- Users shouldn’t need to choose compute machine manually
- System should select best site automatically or suggest to the user
- Requires common API from facilities with queue wait-time estimates
- APS DM system must query API and route job to optimal site
- Shared scheduler across facilities = ideal but unlikely

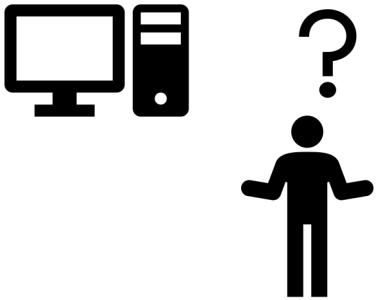


CROSS FACILITY AWARENESS

Current Work

- IRI Interfaces Pathfinder collaboration between ALCF, OLCF, and NERSC
- Developed facility status APIs for sites
  - https://api.alcf.anl.gov/docs
  - https://api.iri.nersc.gov/nersc/api/current/

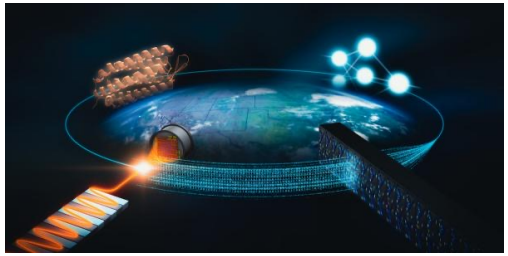
- Initial endpoints report machine availability or maintenance downtime
- Enables workflows to avoid submitting jobs to unavailable systems
- First step toward full cross-facility awareness



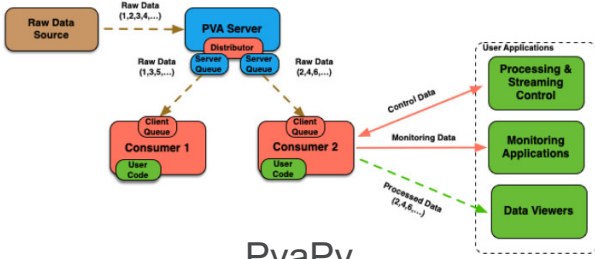
IMPROVING DATA STREAMING INFRASTRUCTURE

The Challenge

- Streaming to HPC demonstrated, but not used in live experiments
- Connections to HPC nodes are cumbersome
- Must reserve nodes and set up stream consumers manually
- Compute nodes lack open network connections and require special agreements
- No standard interfaces across facilities
- APS staff must configure streams and acquire HPC resources



LSLStream



PvaPy

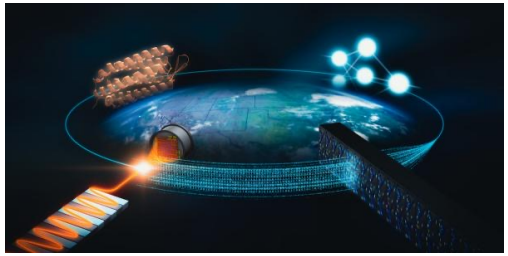
EJFAT

ESnet-Jefferson Lab FPGA Accelerated Transport

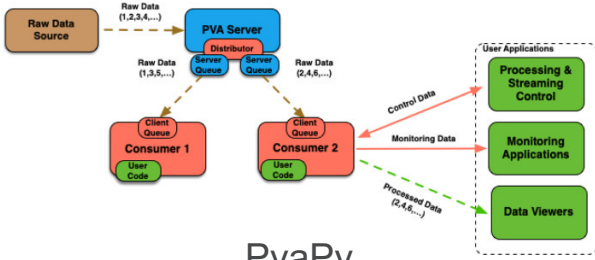
IMPROVING DATA STREAMING INFRASTRUCTURE

What's Needed

- Dynamically provisioning network connections
- Balancing loads across compute nodes
- Fault-tolerant transmission
- Consistent access between sites
- Visualization tools that ingest detector data streams from HPC centers and deliver live feedback to scientists



LSLStream



PvaPy

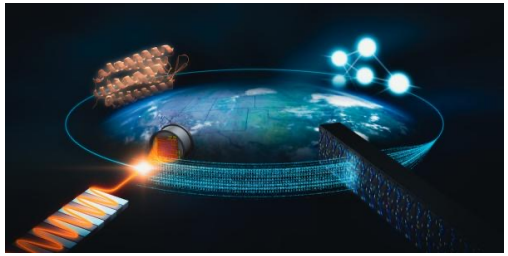
EJFAT

ESnet-Jefferson Lab FPGA Accelerated Transport

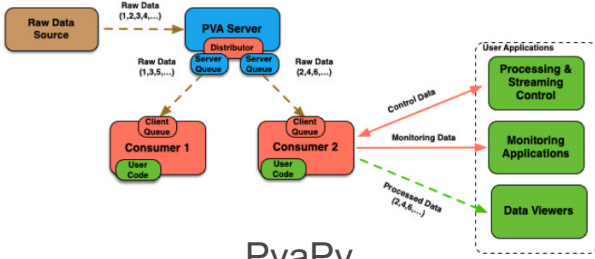
IMPROVING DATA STREAMING INFRASTRUCTURE

Current Work

- Rogers et al. (2025) demonstrated the LCLStream Ecosystem can submit to Slurm or start a Kubernetes pod on Frontier
- ESnet JLab FPGA Accelerated Transport (EJFAT) project aims to provide infrastructure for streaming from instruments to HPC facilities
- Veseli et al. (2025) demonstrated the PvaPy Streaming Framework can be used to stream EPICS data from X-ray instruments
- Skluzacek et al. (2025) developed the Secure Scientific Service Mesh (S3M) which allows streaming to RabbitMQ or Redis at OLCF
- Globus is planning to release a streaming solution between Globus Compute endpoints



LSLStream



PvaPy

EJFAT

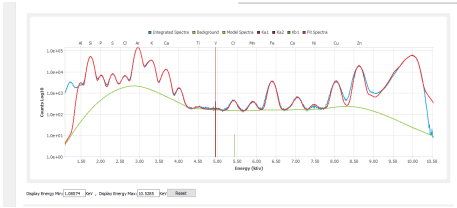
ESnet-Jefferson Lab FPGA Accelerated Transport

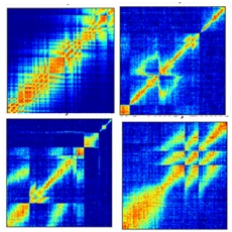
UPLIFTING COMMUNITY SCIENTIFIC SOFTWARE

The Challenge

- APS supports multiple types of X-ray science
  - tomography, ptychography, spectroscopy, diffraction, and more...
- There are existing community scientific software for each type of x-ray data analysis
- Users and staff already know and use these packages
- Most existing software was originally designed for single workstation execution, not optimized for GPU or multiple nodes

TomocuPy
GPU-based Tomographic reconstruction in Python





DIALS
Diffraction Integration for Advanced Light Sources

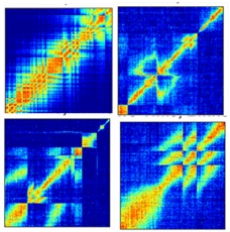
UPLIFTING COMMUNITY SCIENTIFIC SOFTWARE

What's Needed

X-ray analysis software in a modern IRI ecosystem must:

- Process data streams
- Efficiently use multiple nodes and GPUs
- Deploys in standardized containerized environments that provide consistent software stacks across facilities

TomocuPy
GPU-based Tomographic reconstruction in Python



DIALS
Diffraction Integration for Advanced Light Sources

Need to engage with scientific software developers to connect them with:

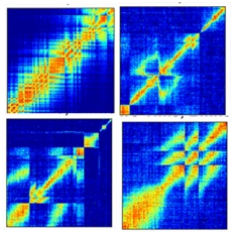
- Container experts
- HPC experts at ALCF and NERSC
- Developers of streaming frameworks

UPLIFTING COMMUNITY SCIENTIFIC SOFTWARE

Current Work

- APS has started a containers working group
- Nikitin et al. (2022) demonstrated streaming tomographic reconstruction
- Prince et al. (2023) demonstrated data processing for Laue microdiffraction optimized to run on ALCF's Polaris using CUDA and MPI
- Ptychodus ptychography processor demonstrated with pvaPy streaming framework
- ALCF and NERSC designate a facility point of contact for projects to ask for help running at their facilities

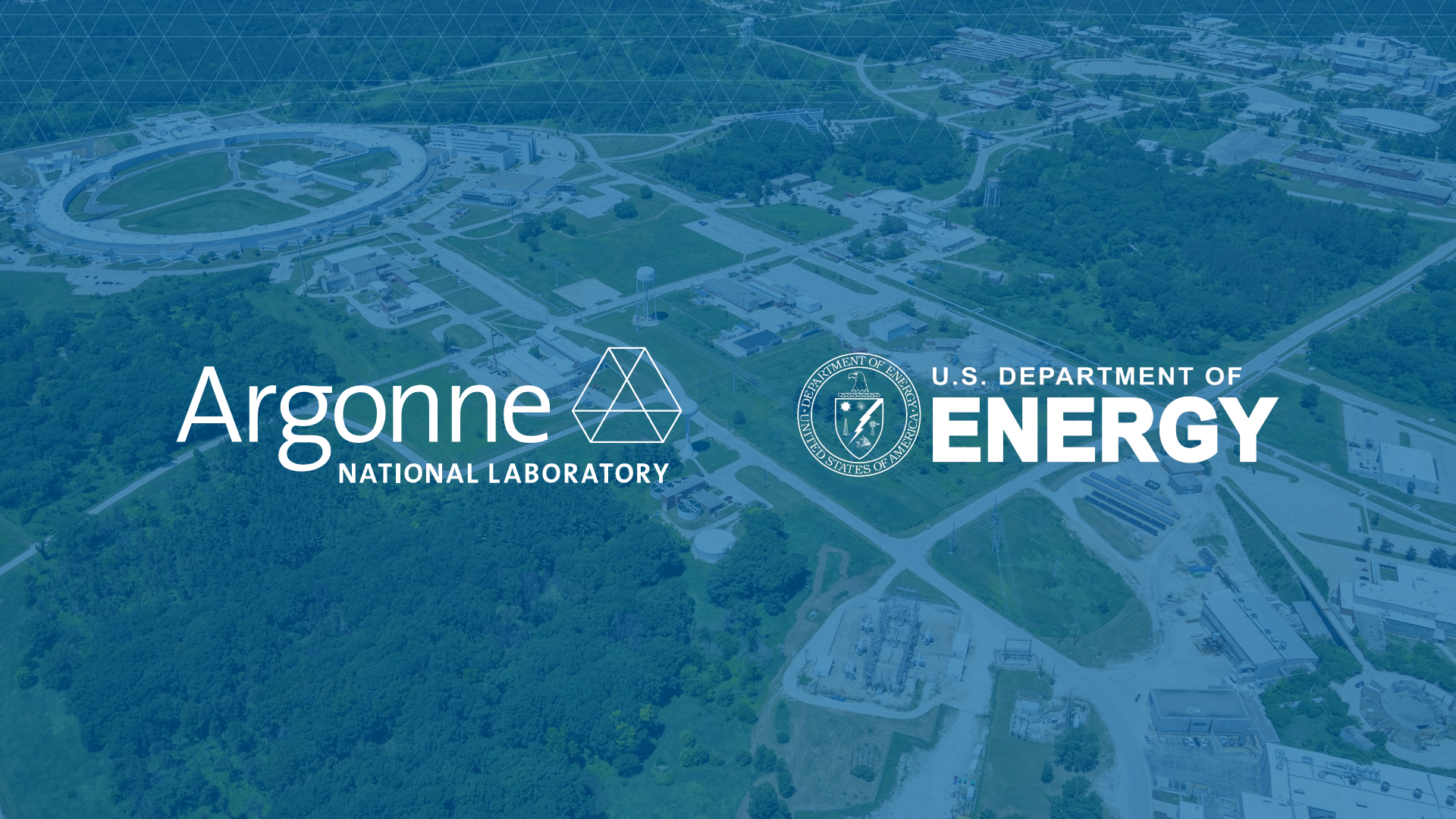
TomocuPy
GPU-based Tomographic reconstruction in Python



DIALS
Diffraction Integration for Advanced Light Sources

CONCLUSION

Opportunities remain for creating a productive software ecosystem for scientific user facilities like the Advanced Photon Source by strengthening cross-facility awareness, improving data streaming infrastructure, and uplifting community scientific software.


Argonne
NATIONAL LABORATORY
U.S. DEPARTMENT OF
ENERGY

REFERENCES

- Nikitin, Viktor, Aniket Tekawade, Anton Duchkov, Pavel Shevchenko, and Francesco De Carlo. "Real-time streaming tomographic reconstruction with on-demand data capturing and 3D zooming to regions of interest." *Journal of Synchrotron Radiation* 29, no. 3 (2022): 816–828.
- Hannah Parraga, John Hammonds, Steven Henke, Siniša Veseli, William Allcock, Benoit Côté, Ryan Chard, Suresh Narayanan, and Nicholas Schwarz. Empowering scientific discovery through computing at the advanced photon source. In Proceedings of the SC '23 Workshops of The International Conference on High Performance Computing, Network, Storage, and Analysis, SC-W '23, page 2126–213, 2023.
- Prince, Michael, Doğa Gürsoy, Dina Sheyfer, Ryan Chard, Benoit Côté et al. "Demonstrating cross-facility data processing at scale with Laue microdiffraction." In Proceedings of the SC'23 Workshops of the International Conference on High Performance Computing, Network, Storage, and Analysis, pp. 2133-2139. 2023
- Rogers, David, Valerio Mariani, Cong Wang, Ryan Coffee, Wilko Kroeger, Murali Shankar, Hans Thorsten Schwander, Tom Beck, Frédéric Poitevin, and Jana Thayer. "The LCLStream Ecosystem for Multi-Institutional Dataset Exploration." *arXiv preprint* arXiv:2510.04012 (2025).

REFERENCES

- Skluzacek, Tyler J., Paul Bryant, A. J. Ruckman, Daniel Rosendo, Suzanne Prentice, Michael J. Brim, Ryan Adamson, Sarp Oral, Mallikarjun Shankar, and Rafael Ferreira da Silva. "Secure API-driven research automation to accelerate scientific discovery." In *Proceedings of the 2025 ACM International Conference on High Performance Computing, Network, Storage, and Analysis*. Association for Computing Machinery, 2025.
- Vescovi, Rafael, Ryan Chard, Nickolaus D. Saint et al. "Linking scientific instruments and computation: Patterns, technologies, and experiences." Patterns 3, no. 10, 2022.
- Siniša Veseli, John Hammonds, Steven Henke, Madeline Miller, Hannah Parraga, Ilya Baldin, Derek Howard, Yatish Kumar, and Nicholas Schwarz, Streaming X-ray Detector Data to Remote Facilities Using EJFAT. To appear in Proceedings of the SC ’25 Workshops of the International Conference for High Performance Computing, Networking, Storage and Analysis, SC-W ’25, 2025.
- Veseli, Siniša, John Hammonds, Steven Henke, Hannah Parraga et al. "PvaPy streaming framework for real-time data processing." Synchrotron Radiation 32, no. 3 (2025).

REFERENCES

- Siniša Veseli, John Hammonds, Steven Henke, Hannah Parraga, and Nicholas Schwarz. Streaming data from experimental facilities to supercomputers for real-time data processing. In Proceedings of the SC ’23 Workshops of The International Conference on High Performance Computing, Network, Storage, and Analysis, SC-W ’23, page 2110–2117, 2023.
- Veseli, Siniša, Nicholas Schwarz, and Collin Schmitz. "APS data management system." Synchrotron Radiation 25, no. 5 (2018): 1574-1580.
- Albert Vong, Steven Henke, Oliver Hoidn, Hannah Parraga, Tao Zhou, Apurva Mehta, and Nicholas Schwarz. Portable cross-facility workflows for x-ray ptychography. In Proceedings of the SC ’24 Workshops of The International Conference on High Performance Computing, Network, Storage, and Analysis, SC-W ’24, page 2135–2143, 2024.

- This research used resources of the Advanced Photon Source (APS) and the Argonne Leadership Computing Facility (ALCF), both U.S. DOE Office of Science user facilities operated by Argonne National Laboratory under Contract No. DE-AC02-06CH11357.
- This work received support from the DOE Office of Science ASCR Leadership Computing Challenge (ALCC) through the 2025–2026 award "Enhancing APS-Enabled Research through Integrated Research Infrastructure."
- This work is supported by the U.S. Department of Energy (DOE) Office of Science, Advanced Scientific Computing Research (ASCR) and Basic Energy Sciences (BES) under the award "ILLUMINE - Intelligent Learning for Light Source and Neutron Source User Measurements Including Navigation and Experiment Steering."
- This work is also supported by the U.S. Department of Energy (DOE) Office of Science-Basic Energy Sciences awards Collaborative Machine Learning Platform for Scientific Discovery and Collaborative Machine Learning Platform for Scientific Discovery 2.0.