JUNE 11, 2025

AUTOMATING SCIENTIFIC INSIGHTS AT THE ADVANCED PHOTON SOURCE

HIGH-PERFORMANCE WORKFLOWS FOR ANALYZING EXPERIMENTAL DATA

HANNAH PARRAGA
Software Engineering Associate
Argonne National Laboratory

RYAN CHARD, JOHN HAMMONDS, PETER JEMIAN, STEVEN HENKE, NICKOLAUS SAINT, NICHOLAS SCHWARZ, SINIŠA VESELI

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

Argonne NATIONAL LABORATORY

OVERVIEW

Engineers at the Advanced Photon Source are creating automated data processing pipelines leveraging high performance computing to address increasing data needs and enable scientific discovery.

X-RAY LIGHT SOURCES

Indispensable tools in the exploration of matter

- Wavelengths of the emitted photons are similar to the atomic spacing of materials and biological cells
- Advanced Photon Source is one of 5 scientific user facilities funded by the US Department of Energy
- Serves thousands of users per year from academia, government, and industry
- Diverse communities: materials research, biology, geosciences, life sciences, security, and many more

Advanced Photon Source (APS) at Argonne National Laboratory

OBJECTIVE

Enable facility users to gain **meaningful scientific insights** from their experiments as quickly as possible by transforming raw X-ray detector data with some software appropriate for the experiment type (while **minimizing the burden of data processing** on the user).

SCALE OF THE PROBLEM

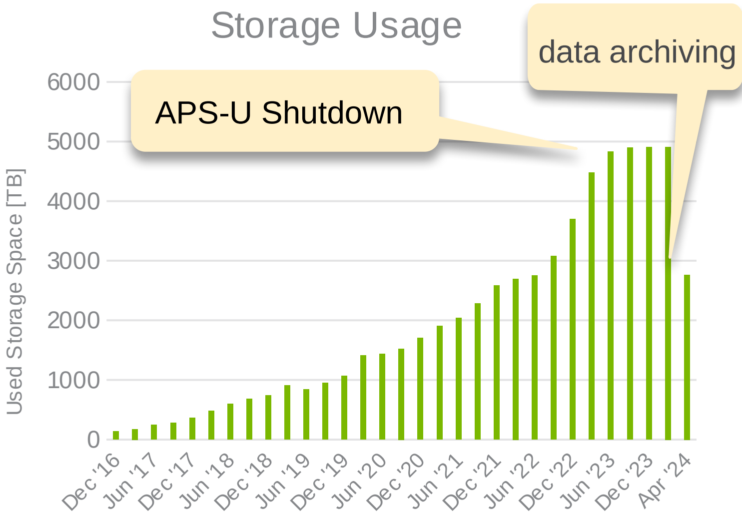
Multiple order-of-magnitude increase in demand for computing resources

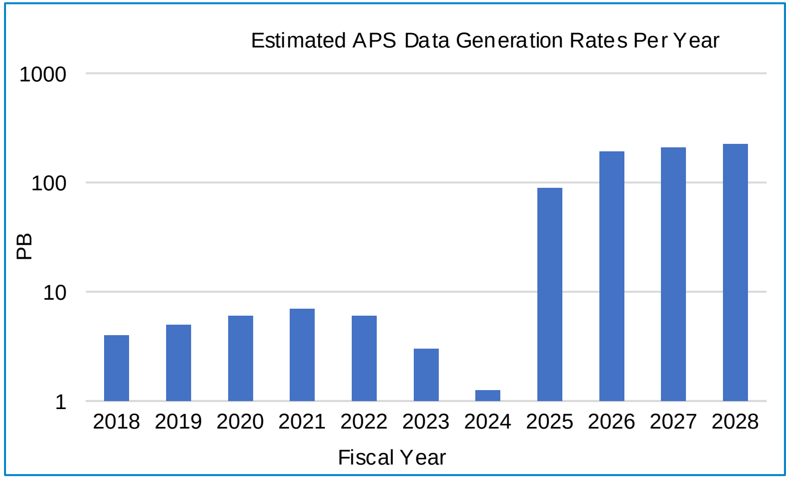
Over the past decade the APS has

- Created over 9000 experiments in the Data Management database
- Used 4.9PB of storage space

Over the next decade the APS will

- Generate 100s of petabytes (PBs) of raw data per year
- Require 10s of petaflop/s of on-demand computing power





REQUIREMENTS

Process experiment data…

Automatically with data collection

When a user triggers

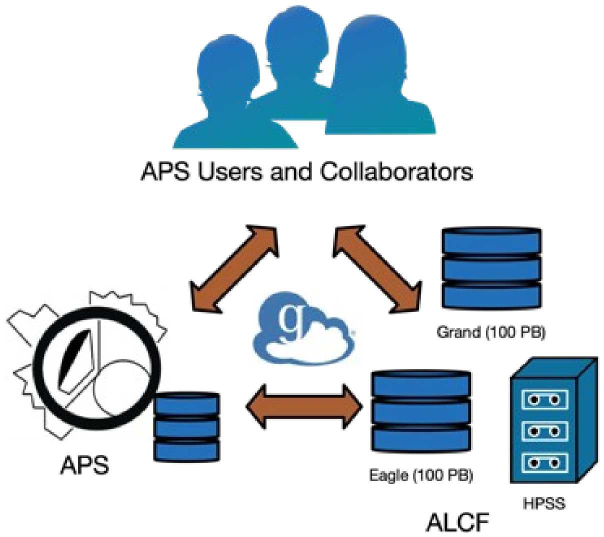
In different places

With a variety of software packages

THE APS DATA MANAGEMENT SYSTEM

Facility-wide software and hardware system for managing data

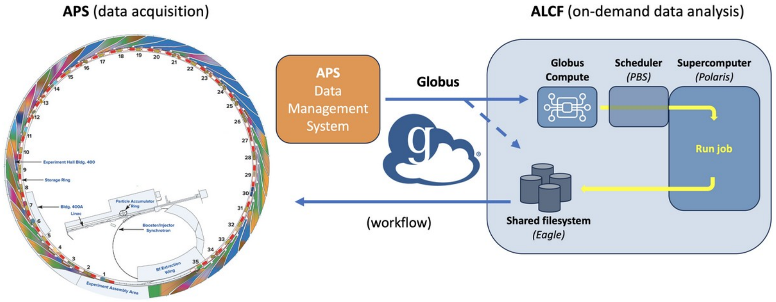
- Tools to automate **transfer** of data, manage storage **access** permissions, and **metadata** catalog
- Workflow tools automate data **processing** via plug-ins
  - Automated or user-initiated jobs
  - Flexible execution of any commands
  - Execute jobs locally or remotely
  - Integrated with Bluesky (beamline control software) and Globus

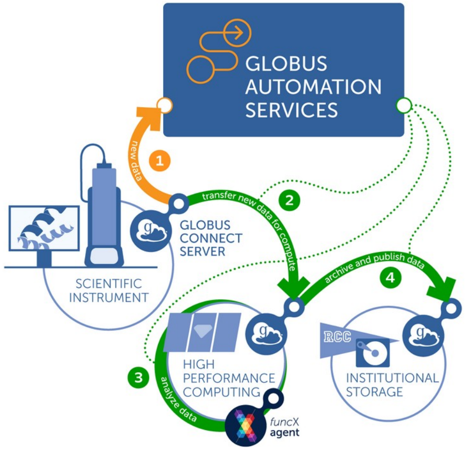


BRIDGING LARGE-SCALE FACILITIES

Advances from Leadership Computing Facilities and Globus

- Facility (ALCF/NERSC) collaboration means APS users do not need to navigate HPC
  - Priority “on-demand” queue for real time processing at ALCF
  - Service/collaboration accounts for beamlines
  - Computing allocations for APS
- Globus Compute
  - Function-as-a-service platform for remote job execution
  - Endpoints deployed at ALCF/NERSC
  - Secure access to data and compute resources





INTEGRATING CONTROLS WITH ANALYSIS

- Bluesky uses information on experiment conditions and where data is located to inform downstream analysis
- Bluesky plans use the APS Data Management API to launch workflows
- Users automatically get processed data when they run a Bluesky plan without extra effort

apstools / apstools / devices / aps_data_management.py

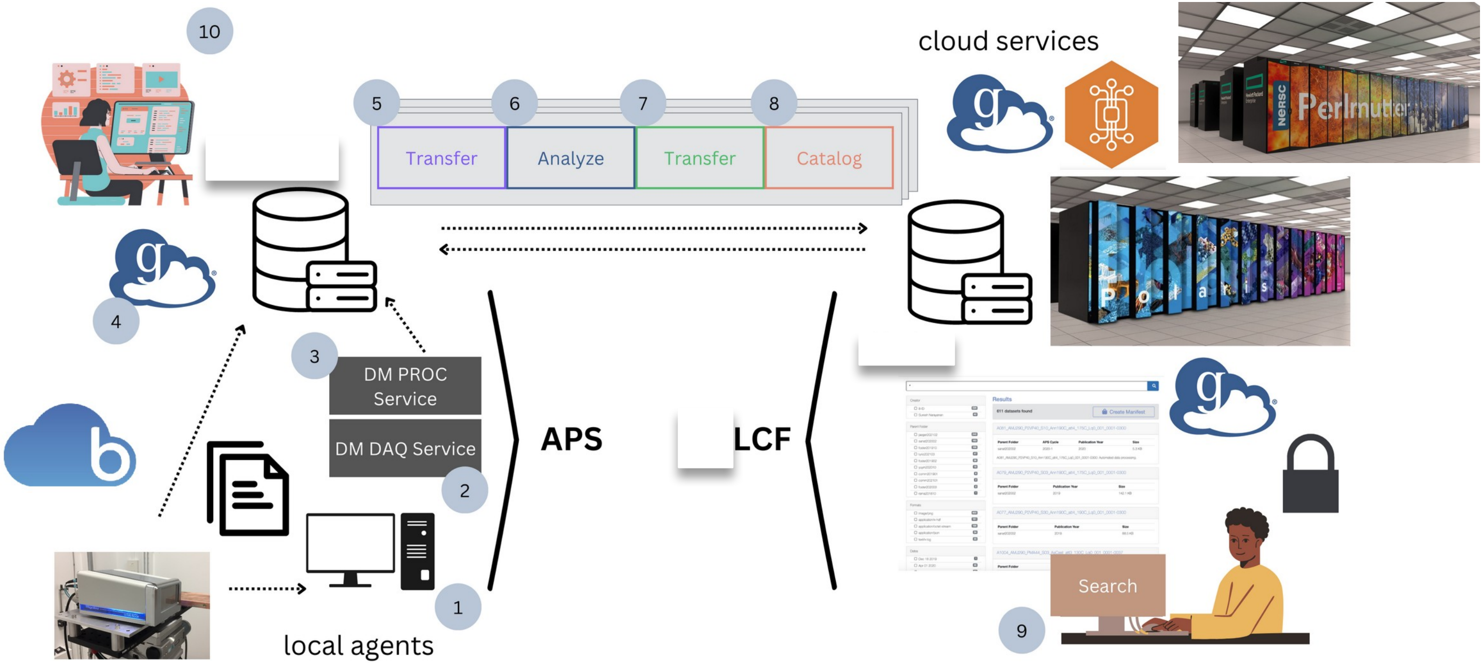
prjemian MNT #872

Code Blame 347 lines (287 loc) · 11.4 KB

```
"""
Connect with APS Data Management workflows.

Example::

    import bluesky
    from apstools.devices import DM_WorkflowConnector

    RE = bluesky.RunEngine()

    dm_workflow = DM_WorkflowConnector(name="dm_workflow", labels=["DM"])
    RE(
        dm_workflow.run_as_plan(
            workflow="example-01",
            filePath="/home/beams/S1IDTEST/
        )
    )
```

Bluesky tools for APS by Peter Jemian and Eric Codrea

DATA LIFE CYCLE



1. Users fill form about experiment
2. Experiment database info populated from form and directory created in central storage
3. Start monitor for files
4. Acquire data and writes metadata
5. Transfer files to storage
6. Select workflow and arguments
7. Data processed locally or with HPC
8. Result published to portal
9. After time, data archived to tape

DATA PROCESSING

Standardized processing for common X-ray techniques

= operational
= in progress

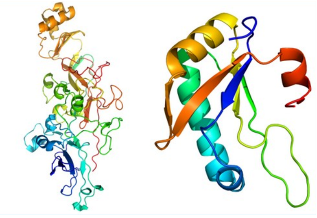
High-energy Diffraction Microscopy
1ID, 20ID
https://github.com/marinerhemant/MIDAS

Wide Angle and Small Angle X-ray Scattering
1ID, 20ID, 12ID
https://github.com/marinerhemant/MIDAS

X-ray Photon Correlation Spectroscopy
8ID, 9ID
https://github.com/AdvancedPhotonSource/boost_corr

Crystallography
23ID
https://www.gmca.aps.anl.gov/

Laue Micro-diffraction
34ID
Prince *et al.*, 2023

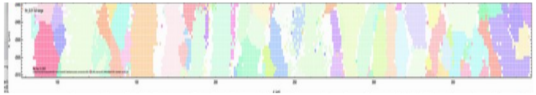


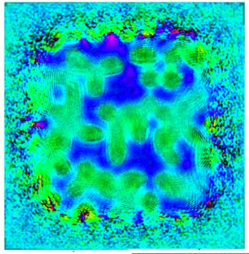
Ptychography
2ID, 4ID, 9ID, 12ID,19ID, 26ID, 28ID, 31ID, 33ID
https://github.com/AdvancedPhotonSource/ptychodus

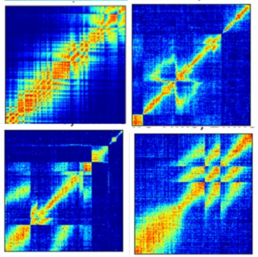
Grazing Incidence X-ray Scattering
9ID
Werzer *et al.*, 2024

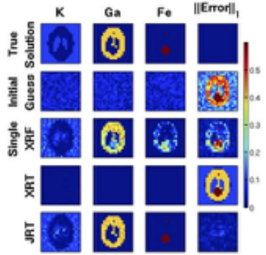
Tomography/Laminography
1ID, 2ID, 2BM, 7BM, 19ID, 20ID, 31ID, 32ID
https://github.com/tomography/tomocupy/

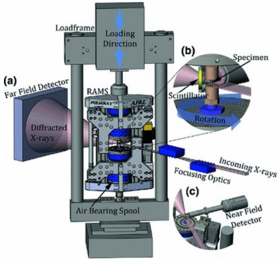
X-ray Fluorescence Microscopy
2ID, 4ID, 12ID, 19ID, 33ID
https://github.com/AdvancedPhotonSource/XRF-Maps











MULTI-TIERED APPROACH

Utilize local and remote resources

High-end compute resources

- Large data processing tasks, ML training, post-processing, and data refinement

Local compute resources

- Perform pre-analysis/data reduction
  - Compression and running ML models
  - Quality control and experiment steering
- May include a GPU workstation at a beamline or the APS computing cluster

Argonne Leadership Computing Facility (ALCF)

Polaris
~44 PFLOP/s

Aurora
> 1 EXAFLOP/s

~4 PFLOP/s of Polaris is prioritized for prototype on-demand use by experimental and observational facilities; when Aurora is in User operations, all of Polaris will be prioritized for on-demand use

Synergy
Planning is underway for the next generation on-demand system prioritized for experimental and observational facilities

Next Generation Supercomputer
Planning is underway for the next generation leadership class supercomputer

Argonne Laboratory Computing Resource Center (LCRC)

Improv
~2.51 PFLOP/s
825 nodes with 2 AMD EPYC CPUs each

Bebop
~1.75 PFLOP/s
672 nodes with 36 Intel Broadwell cores each

Swing
~925 TFLOP/s
48 NVIDIA A100s | 768 AMD EPYC cores

Advanced Photon Source (APS)

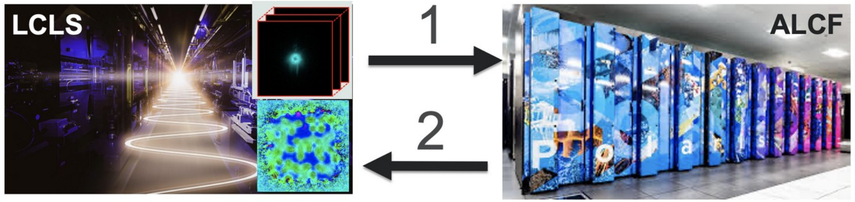
APS general purpose distributed-memory compute cluster
- ~20 TFLOP/s CPU cores

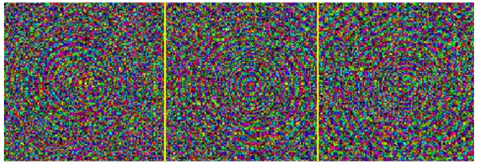
~50 High-Performance Computing Workstations
- Califone – 8 H100s
- Ecto – 8 RTX A6000s
- Refiner – 4 A100s
- Many others…

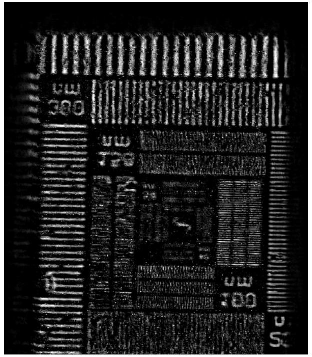
Edge Computing Devices
- 1 x NVIDIA Jetson AGX Orin
- 2 x NVIDIA BlueField-3 DPUs

PORTABLE CROSS-FACILITY WORKFLOWS FOR X-RAY PTYCHOGRAPHY

- Ptychography data volumes are expected to increase by *orders of magnitude* at leading X-ray research facilities due to next-generation upgrades
- Ptychography benefits from access to GPU computing resources
- We demonstrate cross-facility capabilities by deploying software at the **Linac Coherent Light Source (LCLS)**, packaging data and transferring it to **ALCF** for processing
- Used the ptychodus package







Vong *et al.*, 2024

FUTURE WORK

- Expand use of portals and integrate data management features into web pages
- Develop additional workflows and deploy at more beamlines
  - GSAS II, DIALS, laminoAlign and more
- Develop streaming workflows using PvaPy Streaming Framework

ACKNOWLEDGEMENTS

ALCF

- William Allcock
- Benoit Cote
- Jennifer Francis
- Carissa Holohan
- Ryan Milner
- Michael Papka
- Paul Rich
- George Rojas
- Haritha Siddabathuni Som

NERSC

- Johannes Blaschke
- Nicholas Tyler

DSL

- Tekin Bicer
- Ian Foster
- Rajkumar Kettimuthu

Globus

- Rachana Ananthakrishnan
- Benjamin Blaiszik
- Ryan Chard
- James Pruyne
- Nickolaus Saint
- Rafael Vescovi

ACKNOWLEDGEMENTS

3DMN & Atomic

- Jon Tischler
- Dina Sheyfer
- Wenjun Liu
- Ross Harder
- Wonsuk Cha

CHEX

- Hua Zhou
- Dillon Fong
- Matthew Highland
- Hawoong Hong
- Stephan Hruszkewycz
- Zhan Zhang

HEXM

- Jon Almer
- Andrew Chuang
- Peter Kenesei
- Jun-Sang Park
- Victoria Cooley
- Leighanne Gallington
- John Okasinski

Imaging

- Francesco De Carlo
- Alan Kastengren
- Viktor Nikitin

Polar

- Daniel Haskel
- Yongseong Choi
- Gilberto Fabbris
- Joerg Strempfer
- David Gagliano
- Jong Woo Kim
- Srutarshi Banerjee

PtychoProbe & ISN

- Luxi Li
- Volker Rose
- Junjing Deng
- Barry Lai
- Curt Preissner
- Jorg Maser
- Zhonghou Cai
- Sarah Wieghold
- Olga Antipova
- Si Chen
- Yi Jiang
- Fabricio Marin
- Yanqi Luo

XPCS & CSSI

- Suresh Narayanan
- Joe Strzalka
- Qingteng Zhang
- Eric Dufresne
- Zhang Jiang
- Jin Wang
- Ashish Tripathi
- Peco Myint

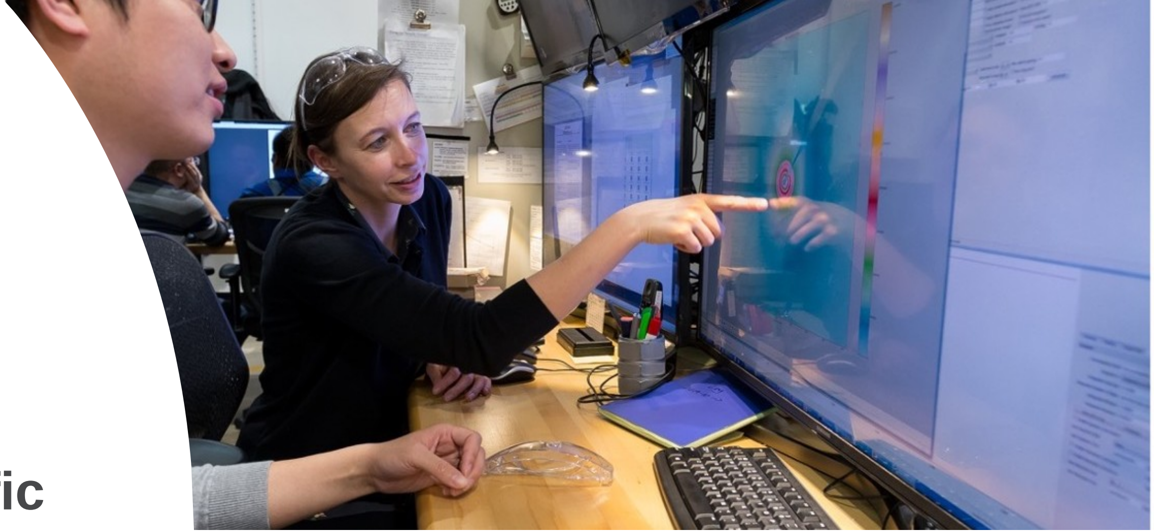
ACKNOWLEDGEMENTS

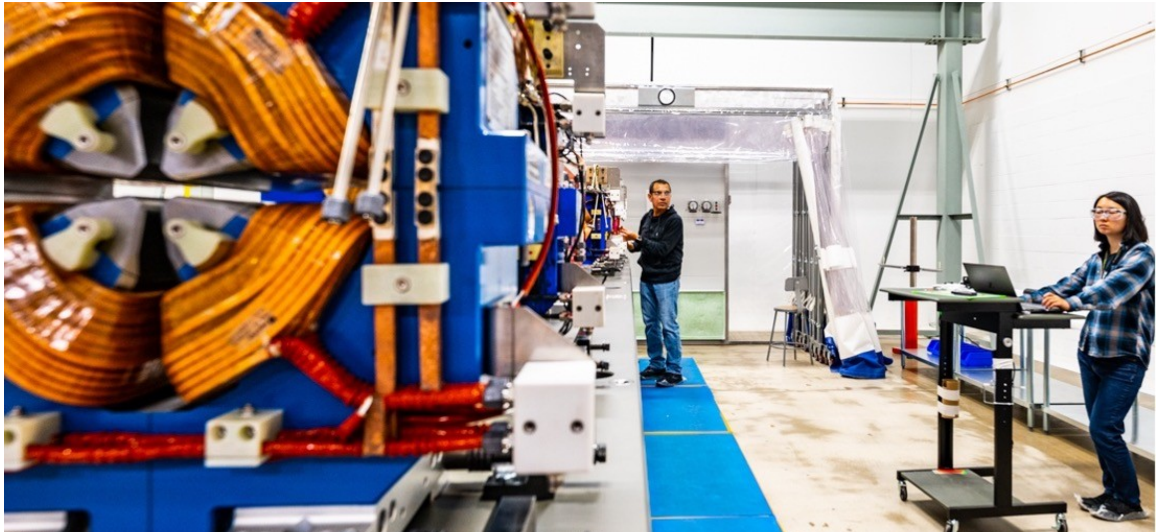
APS:

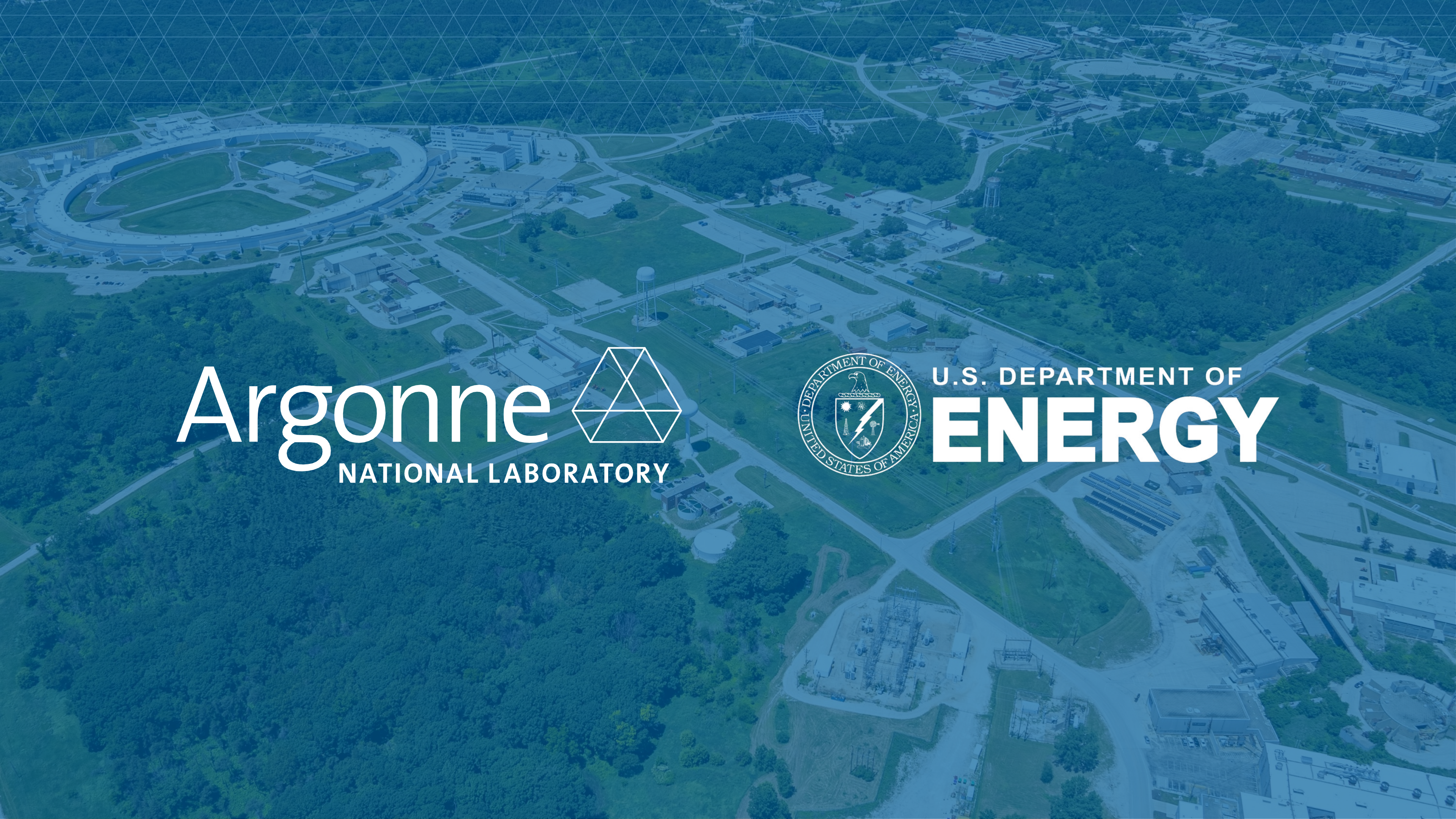
- Ryan Aydelott
- Eric Codrea
- Mathew Cherukara
- Miaoqi Chu
- Barbara Frosik
- Arthur Glowacki
- Doga Gursoy
- Tejas Guruswamy
- John Hammonds
- Steven Henke
- Pete Jemian
- Jonathan Lang
- Keenan Lang
- Alex Lavens
- David Leibfritz
- Antonino Miceli
- Tim Mooney
- Tom Naughton
- Michael Prince
- Alec Sandy
- Nicholas Schwarz
- Roger Sersted
- Hemant Sharma
- Kenneth Sidorowicz
- Joseph Sullivan
- David Wallis
- Mary Westbrook
- Max Wyman
- Sinisa Veseli
- Stefan Vogt
- Albert Vong
- Qingping Xu
- Xuan Zhang
- Tao Zhou

SUMMARY

The Advanced Photon Source is enabling scientific discovery and addressing increasing data needs by creating automated data processing pipelines leveraging high performance computing.






Argonne
NATIONAL LABORATORY
U.S. DEPARTMENT OF
ENERGY

REFERENCES

- Siniša Veseli, John Hammonds, Steven Henke, Hannah Parraga, and Nicholas Schwarz. 2023. 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). Association for Computing Machinery, New York, NY, USA, 2110–2117. https://doi.org/10.1145/3624062.3624610
- Babu, A.V., Zhou, T., Kandel, S. *et al.* Deep learning at the edge enables real-time streaming ptychographic imaging. *Nat Commun* **14**, 7059 (2023). https://doi.org/10.1038/s41467-023-41496-z
- R. Pokharel, "Overview of High-Energy X-Ray Diffraction Microscopy (HEDM) for Mesoscale Material Characterization in Three-Dimensions," *Materials Discovery and Design By Means of Data Science and Optimal Learning*, Springer, 2018, 167 - 201. DOI: 10.1007/978-3-319-99465-9_7
- A. Vong *et al*., "Portable Cross-Facility Workflows for X-ray Ptychography," *SC24-W: Workshops of the International Conference for High Performance Computing, Networking, Storage and Analysis*, Atlanta, GA, USA, 2024, pp. 2135-2143, doi: 10.1109/SCW63240.2024.00267.
- Z. Li et al., "funcX: Federated Function as a Service for Science," in IEEE Transactions on Parallel and Distributed Systems, vol. 33, no. 12, pp. 4948-4963, 1 Dec. 2022, doi: 10.1109/TPDS.2022.3208767.
- Michael Prince, Doğa Gürsoy, Dina Sheyfer, Ryan Chard, Benoit Côté, Hannah Parraga, Barbara Frosik, Jon Tischler, and Nicholas Schwarz. 2023. 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 (SC-W '23). Association for Computing Machinery, New York, NY, USA, 2133–2139. https://doi.org/10.1145/3624062.3624613
- D. Sheyfer, Q. Zhang et al., Phys. Rev. Lett. 125, 125504 (2020)
- Z. Liu, T. Bicer, R. Kettimuthu, D. Gursoy, F. De Carlo, and I. Foster. "TomoGAN: Low-Dose X-Ray Tomography with Generative Adversarial Networks." preprint arXiv:1902.07582 (2019)
- Z. Liu, H. Sharma, J. Park, P. Kenesei, J. Almer, R. Kettimuthu, I. Foster. arXiv https://arxiv.org/abs/2008.08198
- S. Veseli, N. Schwarz and C. Schmitz, J. Synchrotron Rad. 25 1574 (2018)
- Q. Zhang et al., J. Synchrotron Rad. (2021)

REFERENCES

- https://www.globus.org
- https://www.aps.anl.gov/Dynamics-and-Structure
- https://www.gmca.aps.anl.gov/
- https://acdc.alcf.anl.gov/
- https://git.aps.anl.gov/DM/dm-docs
- https://github.com/globus-gladier/gladier-xpcs
- https://github.com/AdvancedPhotonSource/boost_corr
- https://github.com/AdvancedPhotonSource/ptychodus
- https://github.com/AdvancedPhotonSource/XRF-Maps
- https://github.com/Linked-Liszt/laue-parallel