

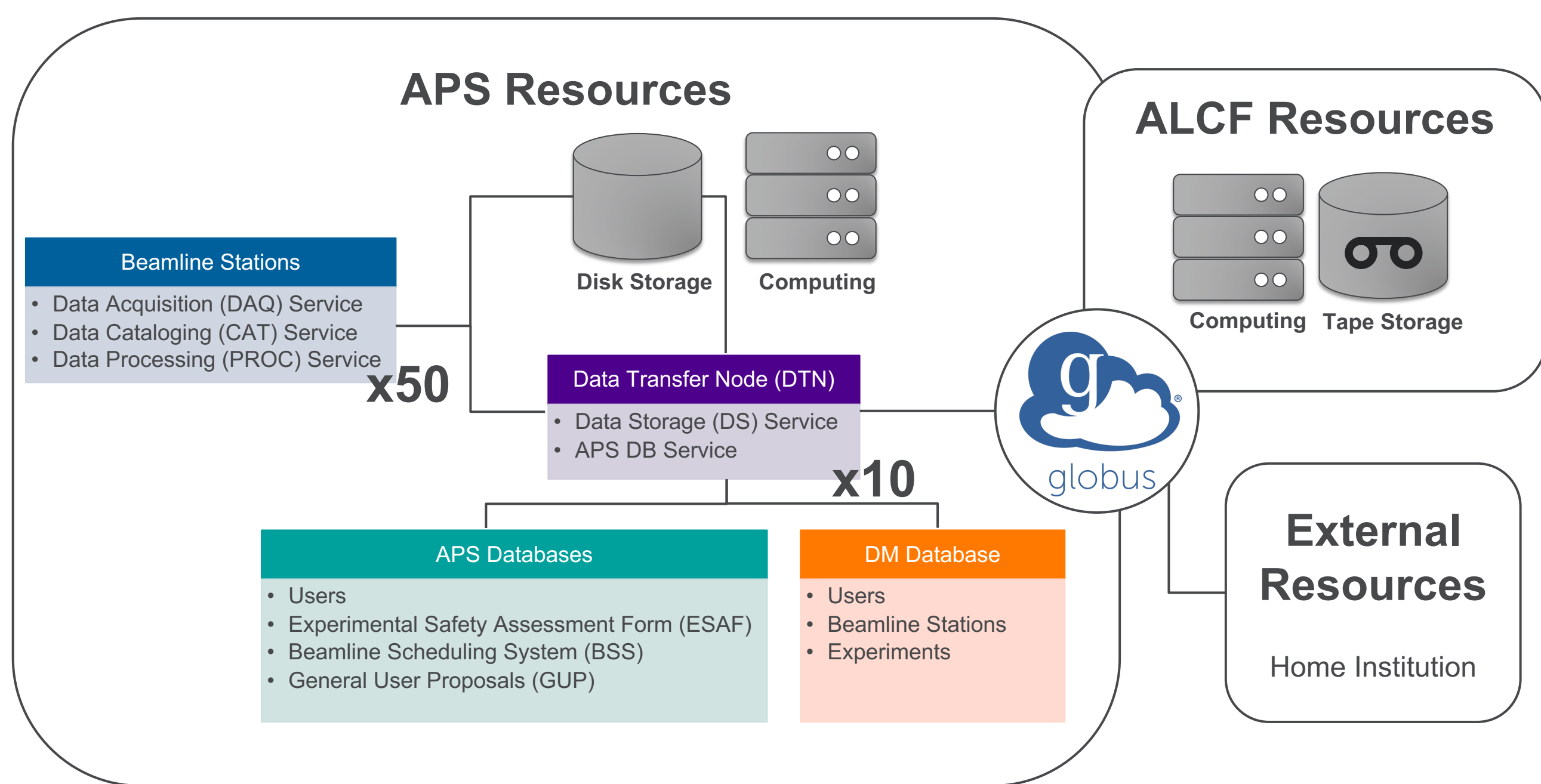
DATA MANAGEMENT SYSTEM ENHANCEMENTS FOR THE APS UPGRADE

Integrating APS Data & Computing Infrastructure to Enable Next-Generation Synchrotron Science

John Hammonds, Steven Henke, Hannah Parraga, Siniša Veseli, Nicholas Schwarz, X-ray Science Division, Advanced Photon Source, Argonne National Laboratory, 9700 S Cass Ave, Argonne, IL 60439, USA

APS DATA MANAGEMENT SYSTEM

The Advanced Photon Source (APS) Data Management System is a data platform that improves the productivity of beamline scientists by automating data management, facilitating data sharing, and executing data-driven workflows. These capabilities are realized by a set of Python web services that interface with facility databases, storage systems, Globus endpoints, and local and remote computing centers to simplify the management and processing of scientific data from experiments. System adoption has increased to 50 of the 68 APS beamlines since its first deployment in 2015.

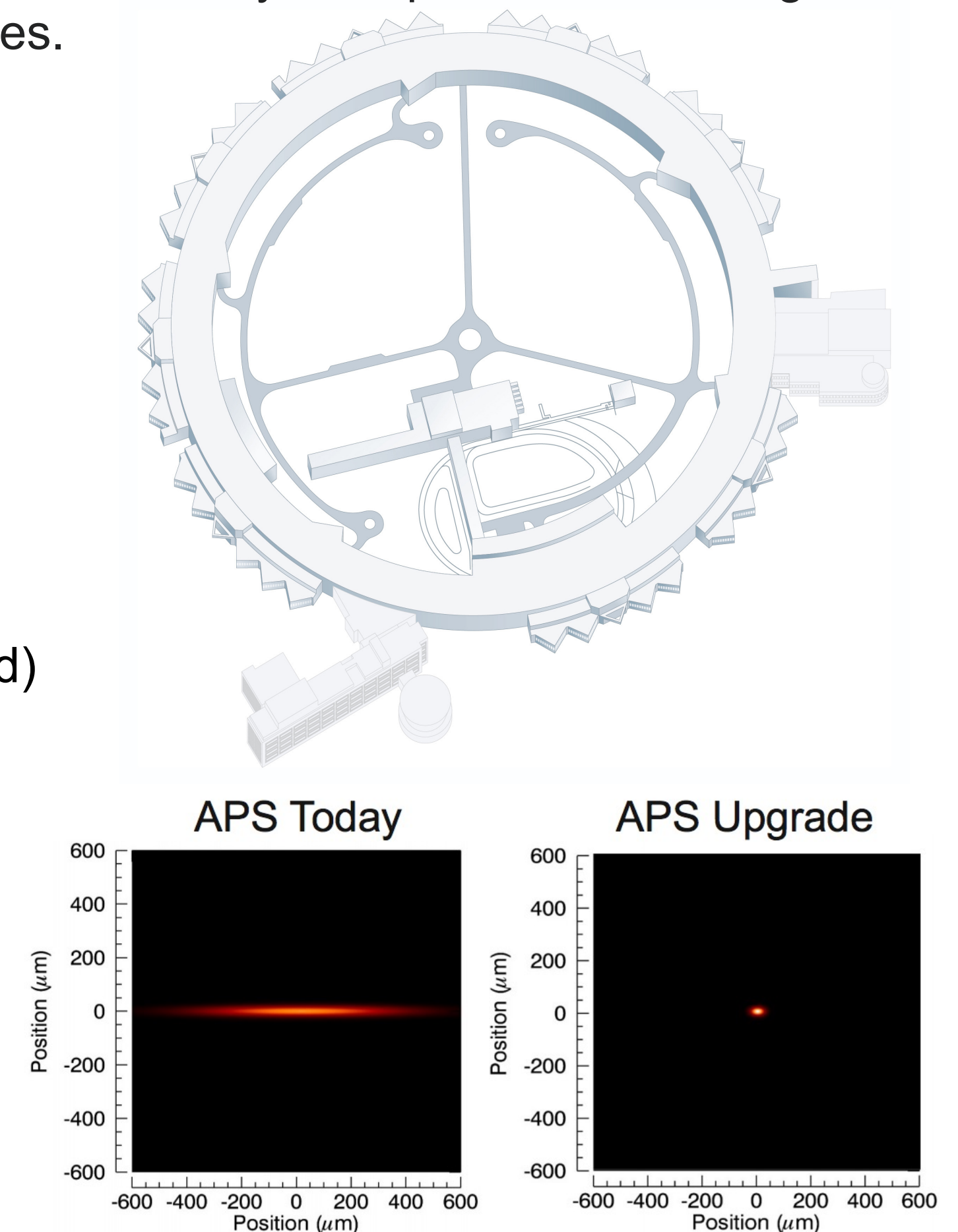


ADVANCED PHOTON SOURCE UPGRADE (APS-U)

An upgrade project is underway that will dramatically improve the APS light source and update beamline stations with next-generation detectors that support higher spatial and temporal resolution. As a result, facility data production rates are expected to increase by multiple orders-of-magnitude when the upgrade project completes.

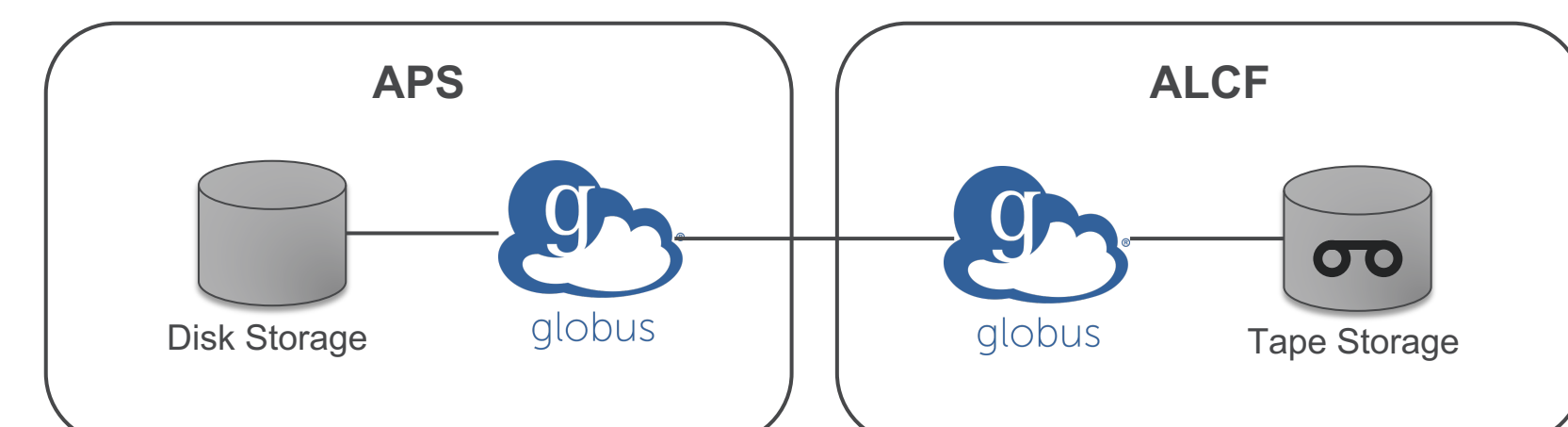
Example: 8-ID-I Beamline (XPCS)

- Lambda 750K detector
 - 1.5MB frame, 2000 fps
 - About 300 MB/s data rate (≈10% non-zero pixels)
- Production rates today
 - 1-2TB/week on average
 - 8-10TB/week max (compressed)
- Production rates post-upgrade
 - Estimated 500 TB/week
- Rigaku 500K detector
 - 50K frames per second
 - Can sustain 1GB/s data rate



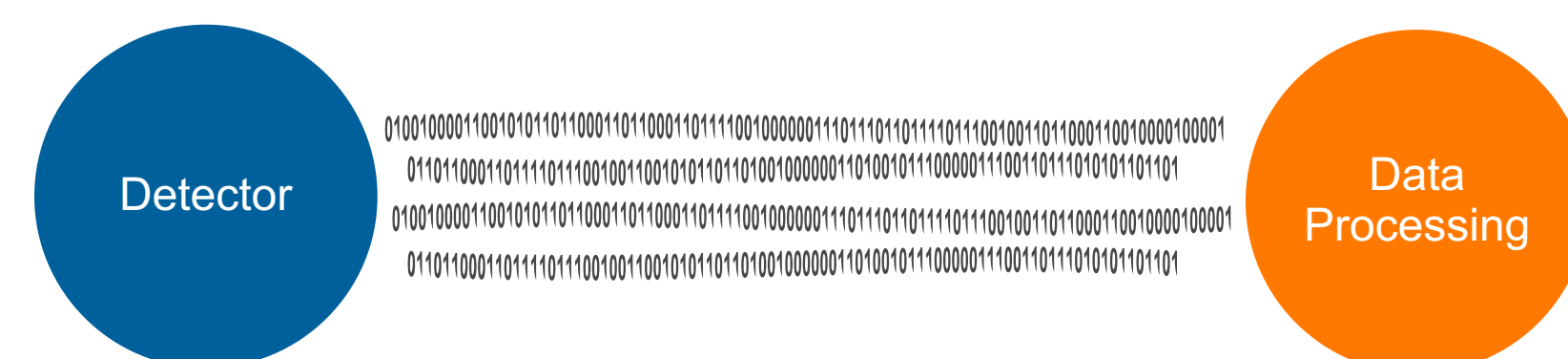
To meet APS-U demands, we are actively developing and deploying new data management capabilities:

DATA ARCHIVING



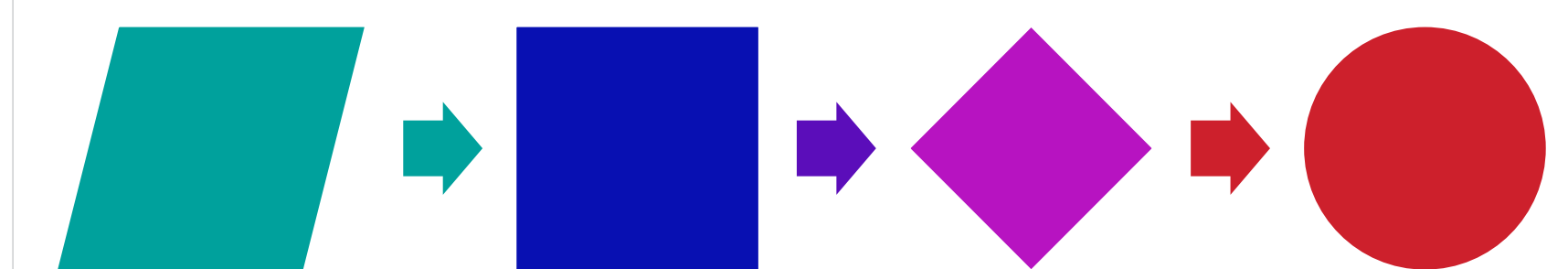
We are implementing a data archiving capability that uses the Argonne Leadership Computing Facility (ALCF) to more efficiently scale to the expected data volumes. Our solution makes use of an allocation on the ALCF High Performance Storage System (HPSS) tape libraries that is accessed using the APS and ALCF Globus Connect Server 5 endpoints.

DATA STREAMING



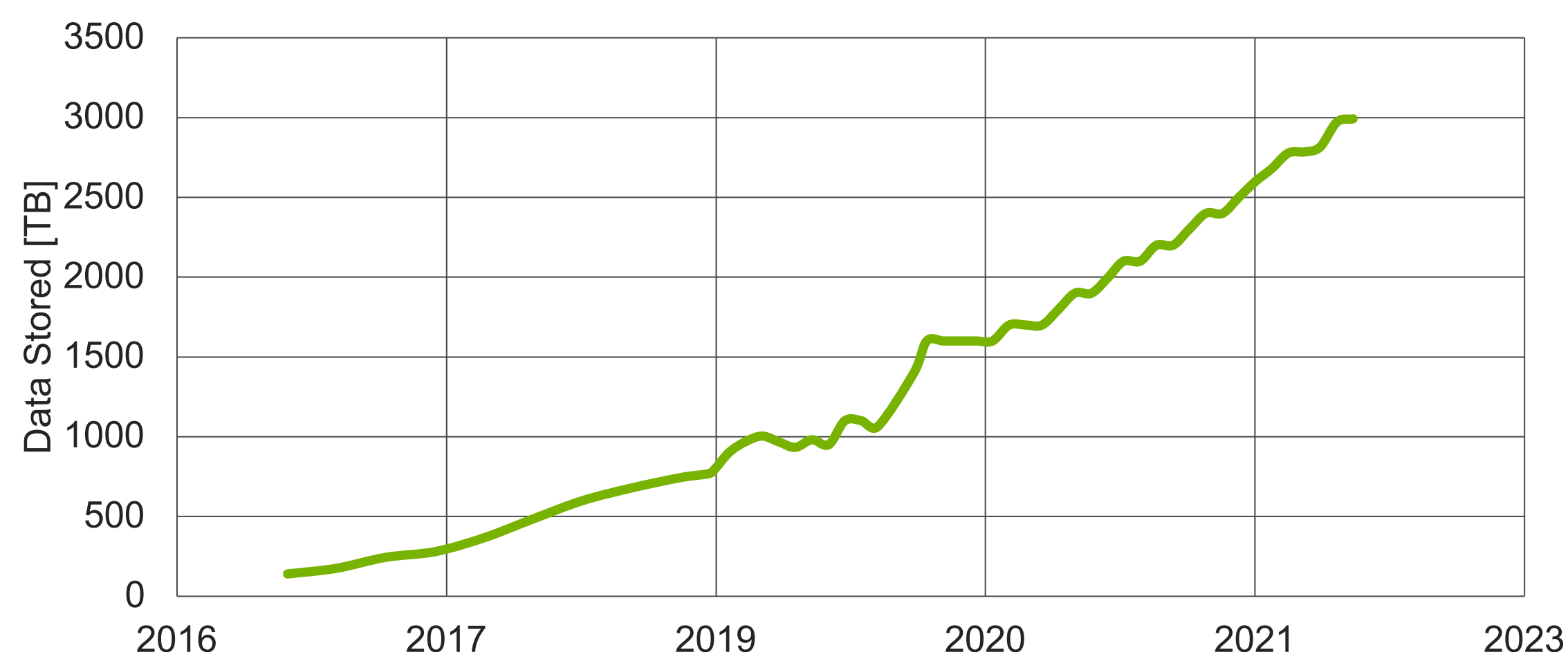
We are developing a prototype data streaming framework into a full-fledged feature that can keep up with the data rates of next-generation detectors. Our solution relies on EPICSS protocols/APIs, which are widely used at APS and other light sources. With our approach, data never needs to touch the filesystem, except possibly for storing processing output.

WORKFLOW IMPROVEMENTS



We are improving the existing workflow system to simplify the creation and execution of analysis tasks. Workflows can include tasks that are executed on local compute resources or at the ALCF. Our plans include GUI improvements to facilitate workflow management, a processing job resubmission capability, and the ability to retain workflow information.

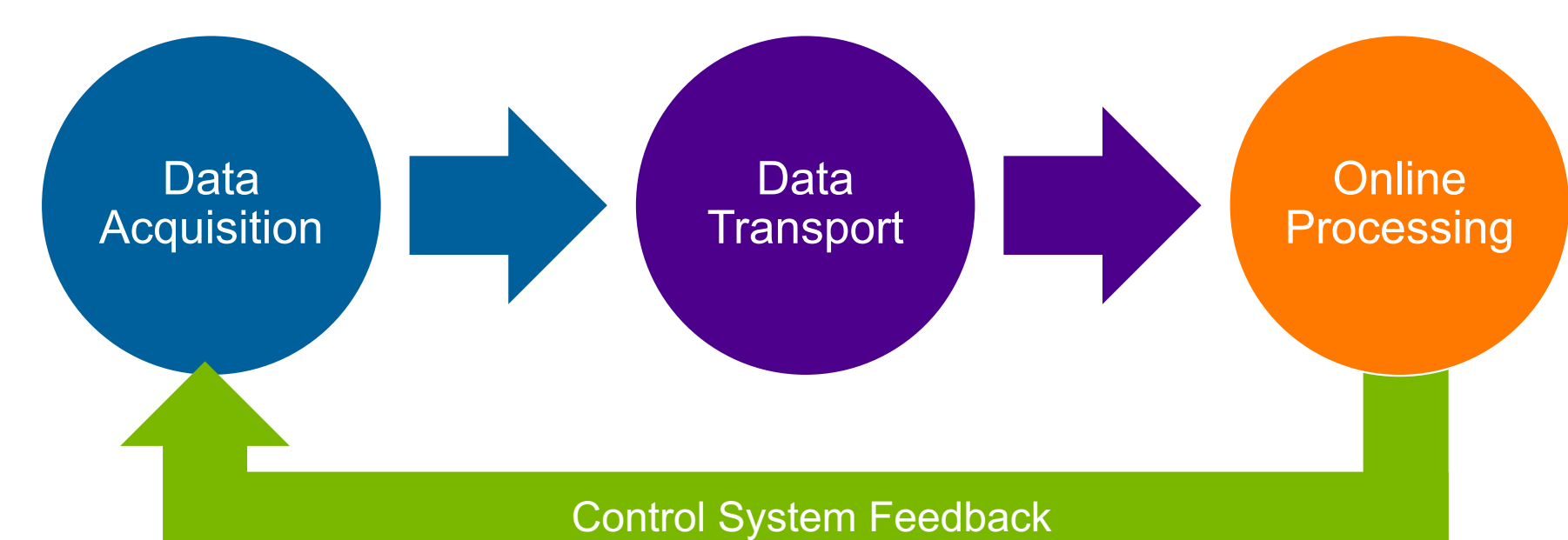
RESOURCE UTILIZATION



- 10/15: First DM software deployment at 6-ID-D
- 12/16: 5 beamlines; 150 TB stored
- 01/20: 35 beamlines; 1.1 PB stored; 2,400+ experiments in DM DB
- 05/22: 50 beamlines; 3.0 PB stored; 6,800+ experiments in DM DB

NEXT-GENERATION BEAMLINE DATA PIPELINE

When incorporated into the beamline operation, these new features can be combined with AI/ML tools to provide rapid insights to the beamline scientists during the data collection process, and ultimately can be integrated into automated control system feedback loops that are tailored to each measurement technique. In this way, the data management system is an important productivity tool for APS beamline scientists that will become essential to making optimal use of APS resources in the coming years.



- Veseli, S., Schwarz, N. & Schmitz, C. (2018). J. Synchrotron Rad. 25, 1574-1580.
- Park, J.-S., Horn, C., Ramanathan, P., Kenesei, P. & Veseli, S. (2019). J. Synchrotron Rad. 26, 373-381.
- Schwarz, N., Veseli, S. & Jarosz, D. (2019). Synchrotron Rad. News 32-3, 13-18.
- Zhang, Q., Dufresne, E. M., Nakaye, Y., Jemian, P. R., Sakumura, T., Sakuma, Y., Ferrara, J. D., Maj, P., Hassan, A., Bahadur, D., Ramakrishnan, S., Khan, F., Veseli, S., Sandy, A. R., Schwarz, N. & Narayanan, S. (2021). J. Synchrotron Rad. 28, 259-265.