**Linking scientific instruments and computation: Patterns, technologies, experiences**

Ian Foster

*The University of Chicago, Chicago, Illinois, USA & Argonne National Laboratory, Lemont, Illinois, USA*

foster@uchicago.edu

**Keywords:** Automation, big data, synchrotron light sources, high-performance computing

Powerful detectors at modern experimental facilities routinely collect data at multiple GB/s. Online analysis methods are needed to enable the collection of only interesting subsets of such massive data streams, such as by explicitly discarding some data elements or by directing instruments to relevant areas of experimental space. Such online analyses require methods for configuring and running high-performance distributed computing pipelines—what we call here *flows*—linking instruments, data center computers (e.g., for analysis, simulation, AI model training), edge computing (for analysis), data stores, data catalogs, high-speed networks, and other resources.

I review common patterns associated with such flows, describe methods for instantiating those patterns, and present experiences with the application of these methods to the processing of data from a range of light source beamlines, each of which engages HPC resources for data inversion, machine learning model training, or other purposes. I also discuss implications of these methods for operators and users of scientific facilities. More details on this work are provided in a recent article [1].

**Figure 1.** *Left:* Cloud-hosted Globus automation services are used in this work for reliable and secure orchestration of flows, often triggered by data availability, and linking various resources including, as shown here, scientific instruments, high-performance computers, and storage systems. Globus and funcX agents provide access to storage and compute resources, respectively.

*Right:* Example flows used at Advanced Photon Source (APS), Argonne Leadership Computing Facility (ALCF), and Stanford Synchrotron Radiation Lightsource (SSRL) for x-ray photon correlation spectroscopy (XPCS), serial synchrotron crystallography (SSX-Stills, SSX-Prime, SSX-Publish [2]), ptychography (Psycho), high energy diffraction microscopy (BraggNN [3], HEDM).


This work was supported in part by the US Department of Energy under contract DE-AC02-06CH11357, including by the Office of Advanced Scientific Computing Research’s Braid project; by NSF grants OAC-1835890 and OAC-2004894; and by award 70NANB14H012 from the US Department of Commerce, National Institute of Standards and Technology, as part of the Center for Hierarchical Material Design (CHiMaD). We are grateful to staff at the APS, ALCF, University of Chicago Globus group, and SSRL for assistance with this work.