Academy at SLAC

Stateful Agents for Autonomous Science

APeX Summit 2026 · Academy, Parsl, and Globus Compute · September 14, 2026

Tim J. Dunn

Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory

Lightning talk, presented remotely · Demonstrated, shared facility capability, the plan for durable experiment state

U.S. DEPARTMENT OF ENERGY | Stanford University

SLAC NATIONAL ACCELERATOR LABORATORY

Where this runs: SLAC, SSRL, and a beamline

Context for the computer scientists in the room · the terms the rest of the talk leans on

SLAC and SSRL

SLAC is a DOE Office of Science national laboratory in Menlo Park, California, operated by Stanford. It runs LCLS, an X-ray free-electron laser, and SSRL, a synchrotron.

SSRL is a DOE user facility. The SPEAR3 storage ring (3 GeV) feeds X-ray beamlines, numbered by their position on the ring: BL 6-2b, BL 17-2, BL 2-1 in this talk. Visiting scientists are awarded beamtime and run with a beamline scientist and an operator.

Me. I run the computing stack for SSRL's Materials Sciences Division and lead SSRL's data, compute, automation, and AI work.

A beamline experiment

The loop. A sample goes in the beam, a detector records, an acquisition system (SPEC today, Bluesky arriving) drives scans. A run is hours to days, and analysis can inform the next scan.

The data. One tomography scan at BL 6-2b is several gigabytes of raw projections; reconstruction and segmentation need GPUs, and historically finished weeks to months after the beamtime ended.

The people. The operator holds authority over the instrument. Anything autonomous has to fit under that.

The compute we reach

On site. S3DF, SLAC's shared scientific data facility: GPU nodes, storage, and Kubernetes for shared services.

DOE HPC and cloud. ALCF Polaris at Argonne, NERSC Perlmutter at Berkeley, OLCF Frontier at Oak Ridge, and AWS over ESnet, reached through IRI, AmSC, and Globus.

The programs. AmSC SUF IP supplies the shared infrastructure; SYNAPS-I, a six-facility Genesis Mission seedling, supplies the AI models and the science case. Battery tomography is where they meet at SSRL.

Terms. Beamtime is the scheduled run. The operator is the person with the instrument. SPEC is the acquisition system on most SSRL beamlines. Tiled is the Bluesky data-access service. IRI is DOE's Integrated Research Infrastructure API for submitting to facilities; AmSC is the American Science Cloud built on it. Globus moves the data and, through Globus Compute, launches the agents.

The experiment outlives any one job

What a beamline run carries. The current stage, the loaded model, pending work, references to intermediate results, and who holds authority over the instrument right now.

What a submission carries. Files, arguments, and metadata. Not the loaded model, the pending decisions, or who holds the instrument; those live in a person's head between stages, and a result gets evaluated before the next stage only if a person does it.

What Academy adds. An addressable service that holds the working context, exposes remotely invocable actions and monitoring loops, and coordinates what happens between stages, reachable across facilities through a cloud exchange. State becomes something the system holds, not something the operator remembers.

Conventional path

Mechanism. IRI API or Globus Flows to PBS, one job per stage, poll, transfer out.

Good. Simple, direct REST, no framework dependency.

Trade-off. No context carried between stages as we run it; evaluation and retry are the operator's job.

Agentic path

Mechanism. Academy agents, launched by Globus Compute into the allocation.

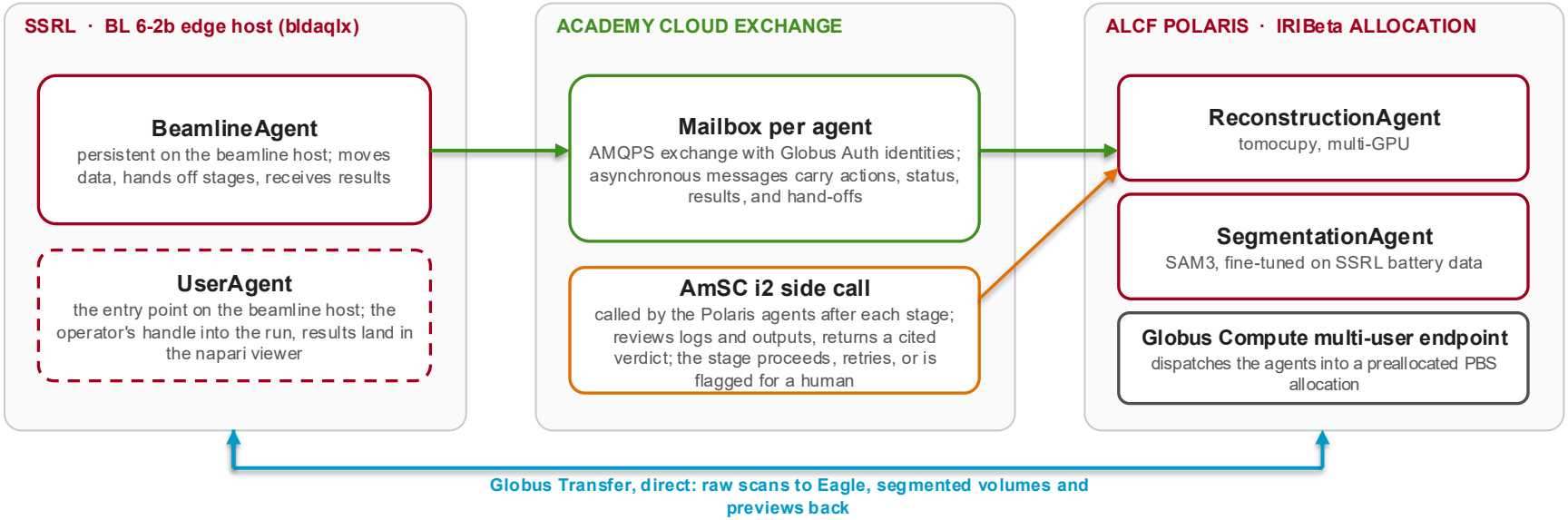
Good. Stateful, self-evaluating with AmSC i2 in the loop, cross-facility by construction.

Cost. One more moving part, the exchange.

Same Polaris-side scripts and kernels under both paths. Academy adds coordination around the existing science code; the agent layer is an option, not a rewrite.

Four agents, two sites, one exchange

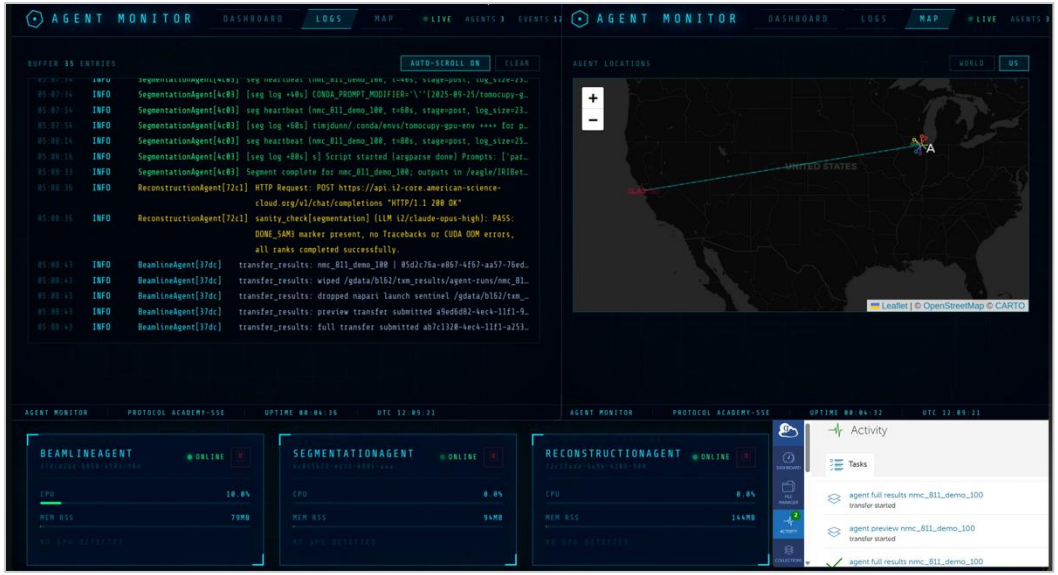
Demonstrated · SSRL to Polaris · Lightsources SYNAPS team, ALCF GPU hackathon, April 2026 · CONFAB26 plenary, May 13, 2026



Globus Compute dispatches the Polaris agents into a PBS allocation; Academy routes action calls and results between the running agents; Globus Transfer moves the data separately. No SSH, no qsub, no manual log tail; a heartbeat coroutine keeps the exchange channel warm through a long stage. The language model serves only the bounded per-stage verdict, and the instrument stays with the operator.

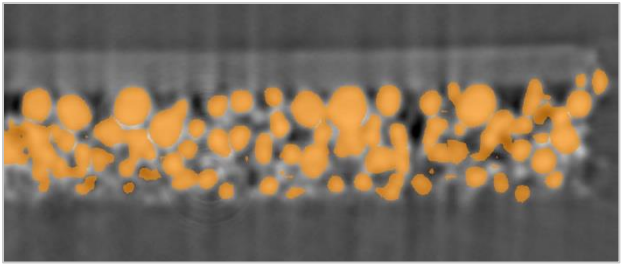
Results back while the run is still going

Demonstrated · what it looked like



The agent monitor during the CONFAB26 run. Segmentation completes, the i2 check returns PASS, the BeamlineAgent submits the result transfers, and the three agents report in below.

Minutes, not months. Reconstruction and segmentation that historically were completed weeks to months after an experiment now return to the beamline in minutes, during the run, in time to inform later scans.



Fine-tuned SAM3 segmentation (orange) over the NMC cathode reconstruction, in the napari viewer at the beamline.

186 to 200 particles

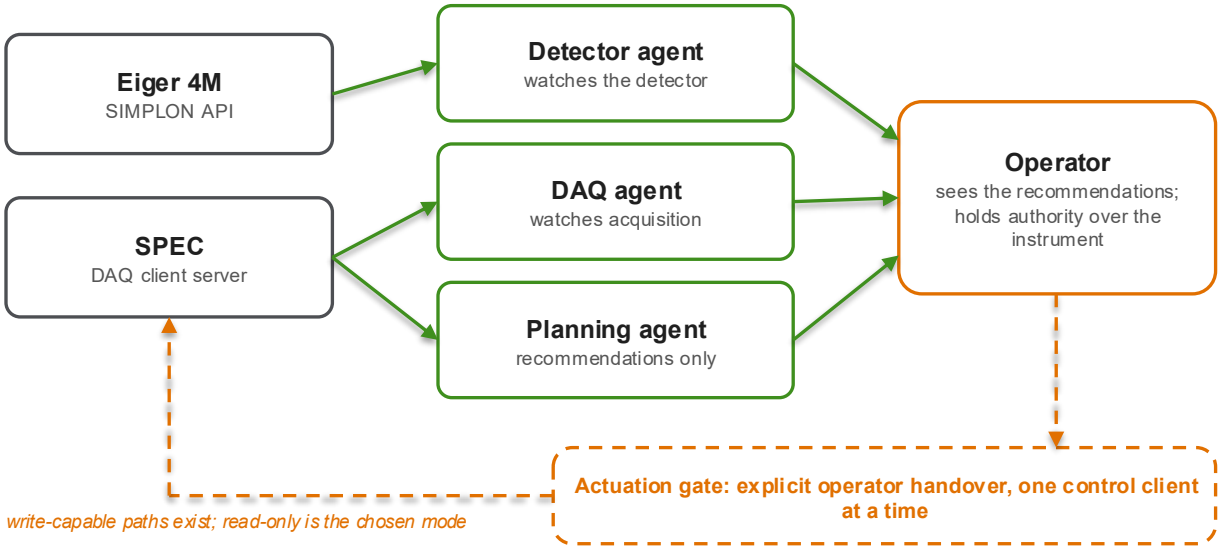
on the NMC cathode across three validated fine-tuned runs, consistent with the conventional pipeline

Same kernels, four targets

Separately, 155 to 157 labels on the 38.6 GB Sand2 reference volume at ALCF, NERSC, S3DF, and AWS; about 26 minutes of compute for a full run, transfers additional

A production beamline, operator in charge

Demonstrated · instrument side · in production on SSRL BL 17-2 since late June 2026



Monitoring continuously during operator-controlled experiments, through the user run and beyond.

Read-only is deliberate. A live test of the alignment routine reached SPEC and stopped at the safety gate. Operator handover governs actuation.

Next. Bounded, operator-authorized actions, alignment first. pyFAI reduction agents on S3DF through Globus Compute. Tiled registration for 17-2 products. More beamlines in the next deployment phase.

Not tomography-specific. Same agent layer, different instrument.

From one identity to SLAC-managed services

Shared capability · what makes it repeatable

S3DF, agent home and compute target

Now. L40S GPU node; workflow execution service and the tiled-ssrl Tiled service on Kubernetes, co-developed and jointly administered with SLAC TID.

Purchased 2026. Torino CPU and Hopper H200 GPU nodes (H200 expected January 2027), 4 PB of storage, a dedicated node for shared services.

Ready to deploy. Academy reduction agents on S3DF, launched through Globus Compute.

Tiled, the record agents write and read

Agents write it. The pipeline's last step registers its products in SSRL's Tiled catalog by reference, with FAIR metadata and provenance. A MetadataAgent, in progress, will record every stage: acquisition, reconstruction, segmentation, model, and job.

Agents read it. The next agent, or the next facility, takes a Tiled reference rather than a path. Records marked public also flow to the AmSC catalog through a policy gate (demonstrated August 20, 2026; SSRL public entries live).

Academy is not the record. Working context is transient; Tiled is what survives, so an agent can restart, or another facility's agent take over, without losing the science. Recovering an agent's pending work is the separate, still-open problem.

SLAC-wide Globus subscription, 2026

What it brings. SLAC identities, managed endpoints per beamline, service accounts, Globus Compute for Academy, an S3 interface to S3DF object storage.

Today. The demonstrations run under my personal ALCF-linked identity and one BL 6-2 endpoint.

After. Agents run under SLAC-owned services, authorized under each site's policies, and the next beamline adopts the workflow without rebuilding the plumbing.

Agents run beside instruments and compute resources. S3DF hosts the shared services. Tiled holds the durable record. That split is what lets a demonstration become a facility capability.

The plan: durable experiment state across facilities

Proposed work at SLAC, not yet started · Academy is the agent middleware · with UChicago (Chard, Babuji) on the open middleware and federated compute

Two gaps. Experiment state does not travel between facilities, and skills stay bound to one instrument. The plan closes both with a transferable record of the experiment and agents that live beside the resources they run on, across independently administered facilities, on the AmSC federation as the cross-lab compute substrate.

Digital sample-state package

The core transferable product: a schema core with versioned technique-specific extensions carrying custody, exposure, and calibrated uncertainty. What the next instrument reads before it touches the sample.

Resource-local agents with lifecycle management

Academy agents across independently administered facilities, runtime-neutral execution with pluggable launchers, an IRI execution adapter, and the AmSC federation as the cross-lab compute substrate.

Skill contracts with explicit authority

What an agent may do, on which instrument, under whose handover. The 17-2 gate, generalized into something another facility can read and enforce.

First. A first version of the sample-state package, the BL 2-1 to BL 17-2 handoff on robotic powder diffraction, scored against expert-operated human baselines with pre-registered task definitions, and a baseline release. The durable experiment record is the first need.

Then. An SLAC-to-APS inter-laboratory port, with partner facilities replicating the state and agent contracts on their own stacks (NSLS-II on Bluesky and Tiled). Benchmarked on time to first result and workflow-decision latency.

Where this stands, and where it goes

Summary · demonstrated, becoming a facility capability, and the plan

Demonstrated

SSRL to Polaris. Four Academy agents across two sites, launched by Globus Compute, coordinated over the exchange, data by Globus Transfer, an LLM verdict after each stage. SAM3 segmentation back at the beamline during the run, 186 to 200 particles across three validated runs.

BL 17-2. Detector, DAQ, and planning agents in production since June, monitor-only by choice, with a fail-closed actuation gate and operator handover.

Becoming a facility capability

S3DF. Agent home and compute target: workflow service and Tiled on Kubernetes with SLAC TID; reduction agents ready to deploy; a MetadataAgent so what agents produce outlives them.

SLAC Globus. SLAC-owned identities, managed endpoints, and service accounts, so the agents stop running as one person.

Where it goes

Proposed at SLAC, with UChicago on the middleware: a digital sample-state package that travels with the sample, resource-local Academy agents at each facility, and skill contracts with explicit authority.

First, a BL 2-1 to BL 17-2 handoff scored against human baselines. Then an SLAC-to-APS port, with partner facilities replicating the contracts on their own stacks.

Autonomy is added to an experiment the way shifts are added to a new operator. Observation first, supervised actuation second, independence only after that, with the record to justify each step.

timdunn@slac.stanford.edu

Thanks to Kyle Chard and Globus Labs; ALCF; Bo Su (fine-tuned SAM3); SLAC TID; and the SYNAPS-I and AmSC teams.

This work grew out of the ALCF GPU hackathon proposal Lightsources SYNAPS (Tim Dunn, PI, with Alex Hexemer and David Abramov), mentored by Yadu Babuji and Christine Simpson.

A. Kamatar et al., Empowering Scientific Workflows with Federated Agents, arXiv:2505.05428. T. J. Dunn et al., Portable AI Workflows and Federated Agents for Experiment-in-the-Loop Synchrotron Science, manuscript, 2026.