Globus Automate

A Distributed Research Automation Platform

PI: Ian Foster (foster@anl.gov), Kyle Chard, Blasé Ur | Grant Number 1835890

globus.org

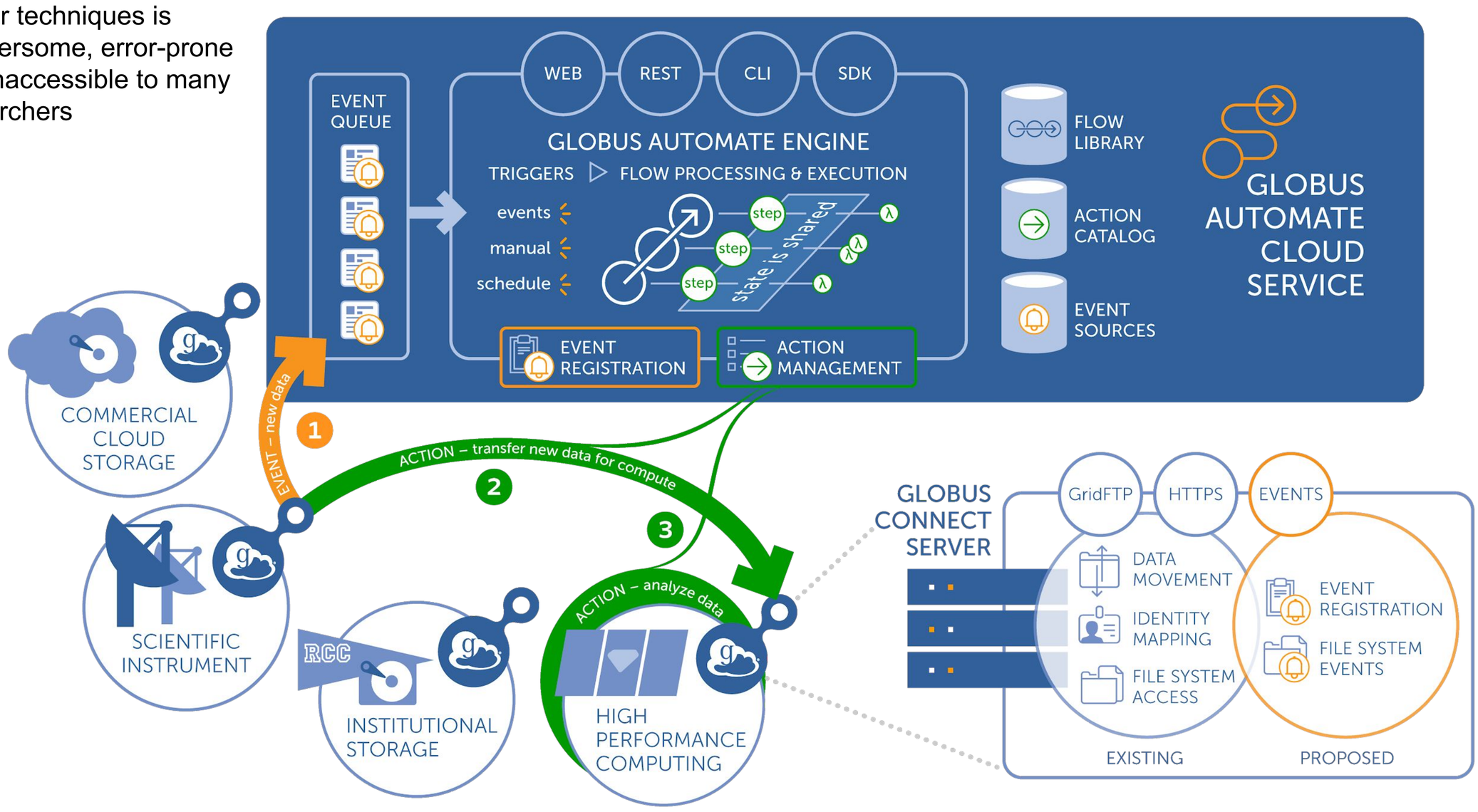
THE UNIVERSITY OF CHICAGO

Motivation/Challenges

- Research processes span locations, collaborations, time scales; it is increasingly infeasible for researchers to manually manage these processes
- Automation via scripting or similar techniques is cumbersome, error-prone and inaccessible to many researchers
- Automation requires monitoring and response to different types of events (e.g., data creation, periodic 'cron' triggers) and may comprise many different actions such as data movement, sharing, analysis, and publication.

Key Features

- Composition and execution service for automation and orchestration of research processes
- Trigger-action programming model with high level specification and authoring tools
- Automatic and reliable invocation of processes in response to events
- Modular architecture with open APIs to support arbitrary event sources and actions



Globus Automate Model

Events

Detect and respond to changes:

- Files created, updated, or deleted
- Instruments creating data
- Time elapsing
- *Action* completion

Triggers

Bind *Events* to *Flows* or *Actions*

- *Event* payload transformed to *Action* input
- *Action* invoked with identity of user creating the *Trigger*

Actions

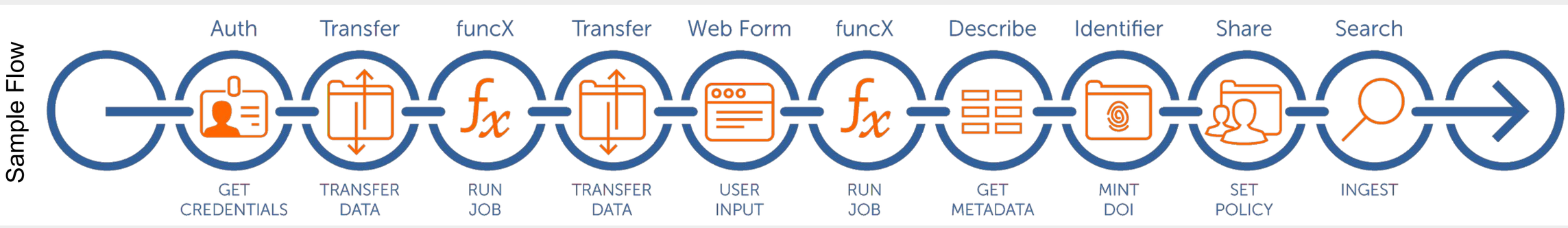
Implement reliable, uniform method for running and/or monitoring an activity

- Can be long running tasks
- Any user may provide an *Action* for use by all users

Flows

User-defined combination of *Actions* to perform complex activities

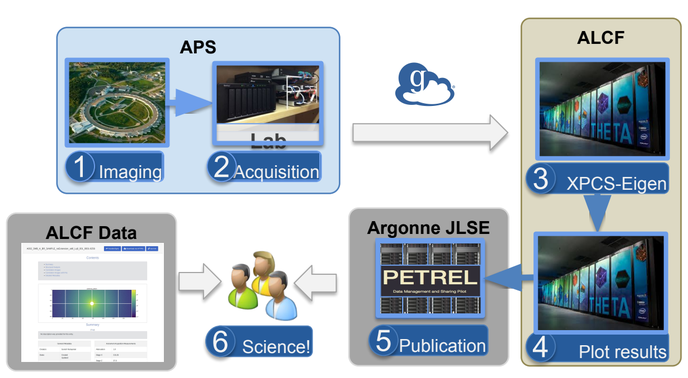
- Simple, UI-friendly format
- Branching/error-handling/ concurrency



Instrument Data Management & Analysis

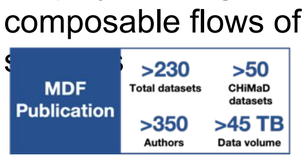
- Automate flows stage data to ALCF for on-demand analysis and publication
- Metadata and plots are dynamically extracted and published into a search catalog
- Scientists can select datasets and initiate flows to perform batch analysis tasks

X-ray Photon Correlation Spectroscopy at the Advanced Photon Source uses Globus Automate to offload analysis tasks to Argonne Leadership Computing Facility

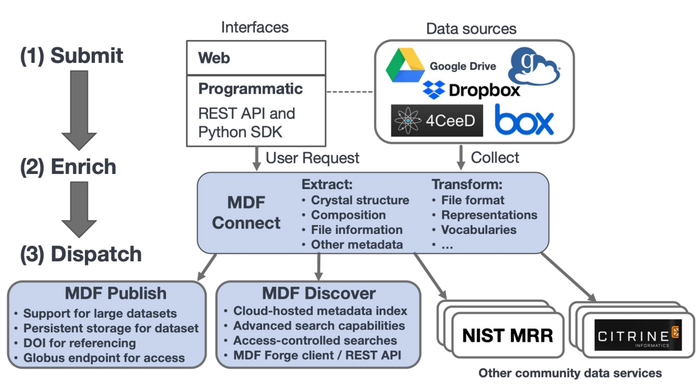


The Materials Data Facility (MDF)

- **Accept data from many locations** with flexible interfaces
- **Index dataset contents** in science-aware ways
- **Dispatch data** to the community
- Using Automate to simplify building composable flows of



An openly flowing materials data ecosystem to support machine learning applications and materials discovery



https://www.materialsdatafacility.org

Contributors: Rachana Ananthakrishnan, Ben Blaiszik, Ryan Chard, Jonathan Gaff, Mattias Lidman, Jim Pruyne, Rudy Richter, Jason Williams, and Vytas Čuplinskas