Enable an interoperable ecosystem of research applications and services

Transform how research applications and services are...

- created, used and delivered
- orchestrated to achieve automation
- sustained
Globus platform services

• **Identity and Access Management (IAM)**
  – Auth
  – Groups

• **Data Services**
  – Connect
  – Transfer
  – Manifest

• **Search**

• **Identifiers (collaboration with DataCite)**

• **Flows**
Use(r)-appropriate interfaces

GET /endpoint/go%23ep1
PUT /endpoint/vas#my_endpt
200 OK
X-Transfer-API-Version: 0.10
Content-Type: application/json
...

Globus service

Web

CLI

Rest

API
“Simple” Automation Use Cases

• Data backup – as user, as system
• Stage data in or out as part of a compute job
• Portal/science gateway submits a transfer of compute results as the user
• Portal/science gateway monitors users transfer, and initiates processing or backup of data.

Recurring transfers with sync option

Copy /ingest Daily @ 3:30am
Globus CLI and SDK

• CLI
  – Scripting and interactive use cases
  – Easy install and updates
  – “globus login” gets access tokens and refresh tokens
  – All interactions with the service use the tokens

• Python SDK
  – All interactions with Globus services: Auth and Transfer
Managing Instrument Data Delivery

Use Globus to deliver 100s of TB of genomic data to researchers

Credits: Joe George, University of Michigan
Automate at scale in heterogeneous environments
Globus Automate

A platform for defining, applying, and sharing distributed research automation **flows**

**Triggers** to start flows based on subscribed events

Flows call **Action Providers** to perform tasks
CityCOVID

• Integrated COVID-19 pandemic monitoring, modeling, and analysis capability.

• CityCOVID is a city-scale agent-based model

• Automate flow
  – Scrape daily Chicago reports
  – Perform simulations at ALCF
  – Postprocess data at LCRC

Jonathan Ozik, Nick Collier, and Charles Macal
Enabling serial crystallography at scale

- Serially image chips with thousands of embedded crystals
- Quality control first 1,000 to report failures
- Analyze batches of images as they are collected
- Report statistics and images during experiment
- Return crystal structure to scientist

Darren Sherrell, Gyorgy Babnigg, Andrzej Joachimiak
SSX Automation

Data capture

- High quality FAIR

Image processing

- QA Process
- Stop?
- Threshold
- Transfer data
- Analyze
- Generate crystal map

Publishing

- Publish results
- Index
- Catalog
- Return results
- Visualize

Data capture

- X-rays

High quality FAIR

- Publish
- Index
- Catalog
- Return results
- Visualize
XPCS: X-ray Photon Correlation Spectroscopy

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

Suresh Narayanan, Nicholas Schwarz
Automating XPCS

- Auth: Get credentials
- Transfer: Transfer HDF5
- Transfer: Transfer IMM
- funcX: Run Corr
- Share: Set ACL
- Transfer: Return Results
- Search: Ingest
- funcX: Plot Results
Materials Data Facility

- 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 services

> 40 TB of data
> 400 datasets
> 320 published authors
PaaS: develop custom action providers

- Directly use the platform to build and run extensible flows
- Develop action providers
  - Fit for purpose
  - Developed and deployed by the project
  - Plugged into their flows
- Action Provider Development toolkit
SaaS: instrument data management

- Templated solution
- Configurable...
  - Set transfer triggers
  - Select destination(s)
  - Define metadata
- Extensible...
  - Add/remove actions
  - Change action providers
- No development required

Cryo EM
Lightsheet Sequencer

Image reconstruction, analysis, visualization

Automated egress from device

Transfer

Xtract

Indexing for search

--/cohort045
--/cohort096
--/cohort127
SaaS: Data Management Plans

• “Turnkey” DMP enablement
• Select dataset (collection)…
• …add metadata for indexing
• …generate persistent ID (DOI, ARK, etc.)

“Point & Click” to findable and accessible data
Data portals leveraging the platform
Questions?
Thank you, funders...