Automating Data Management Flows with iRODS and Globus

Vas Vasiliadis

iRODS UGM – June 8, 2021

THE UNIVERSITY OF CHICAGO

Argonne NATIONAL LABORATORY

Does your institution use (or plan to use) Globus?

Text to 22333:
UGM1 for “Yes”, **UGM2** for “No”

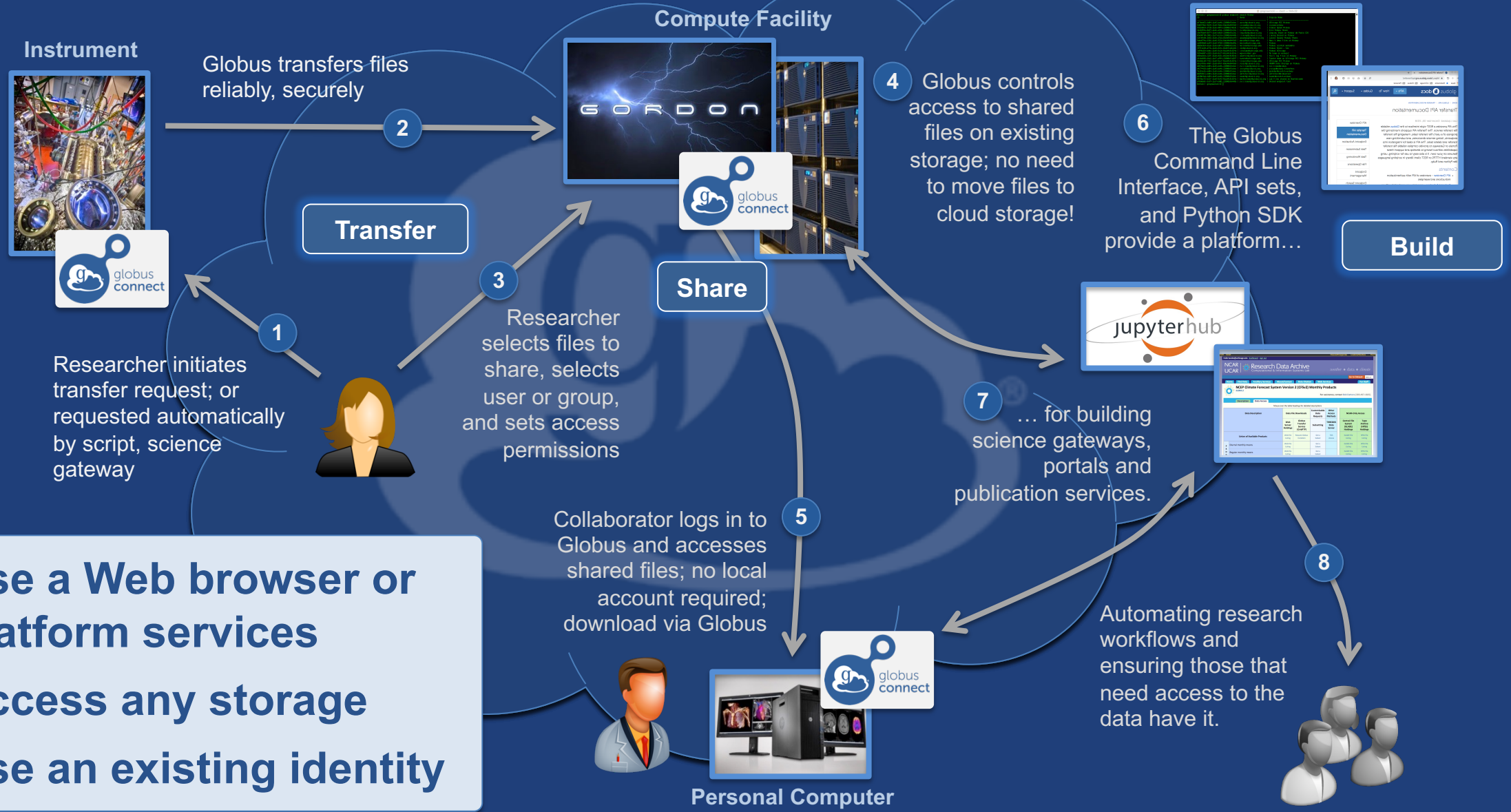
Reply at: **bit.ly/globuspoll**

Globus is …

a non-profit data management service developed and operated by

THE UNIVERSITY OF CHICAGO

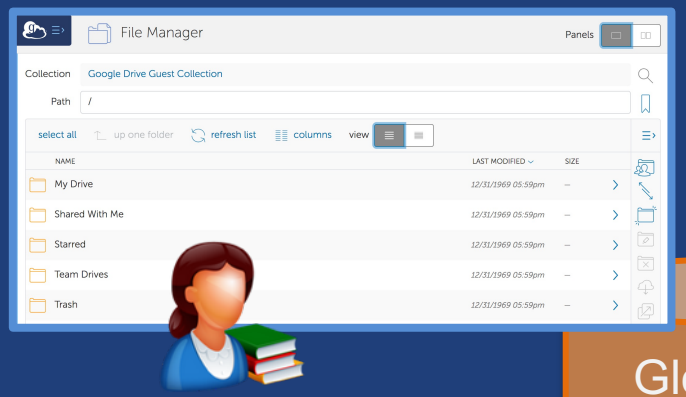
The Globus data management platform

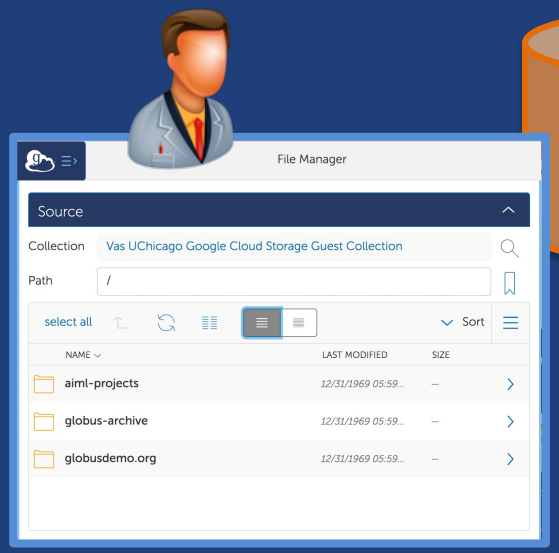


- **Use a Web browser or platform services**
- **Access any storage**
- **Use an existing identity**

Globus accessible storage systems

iRODS + Globus deployment





iRODS

Globus Collection

Globus Collection

Globus Endpoint

Data Transfer Node(s)

Globus Connectors

globus connect

iRODS

Other Globus Endpoints

globus connect

Data

Control

Control

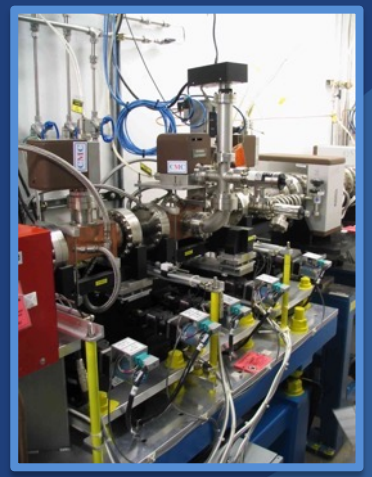
globus

Demonstration

Accessing iRODS via Globus…

Common Use Cases

Next-Gen Sequencer



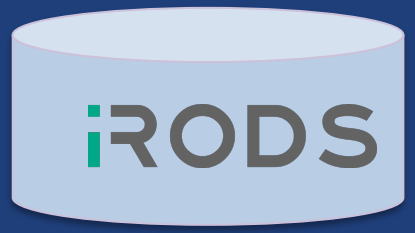
Advanced Light Source

Cryo-EM

MRI

Light Sheet Microscope

Analysis store



Data Archival/ Distribution

Remote visualization

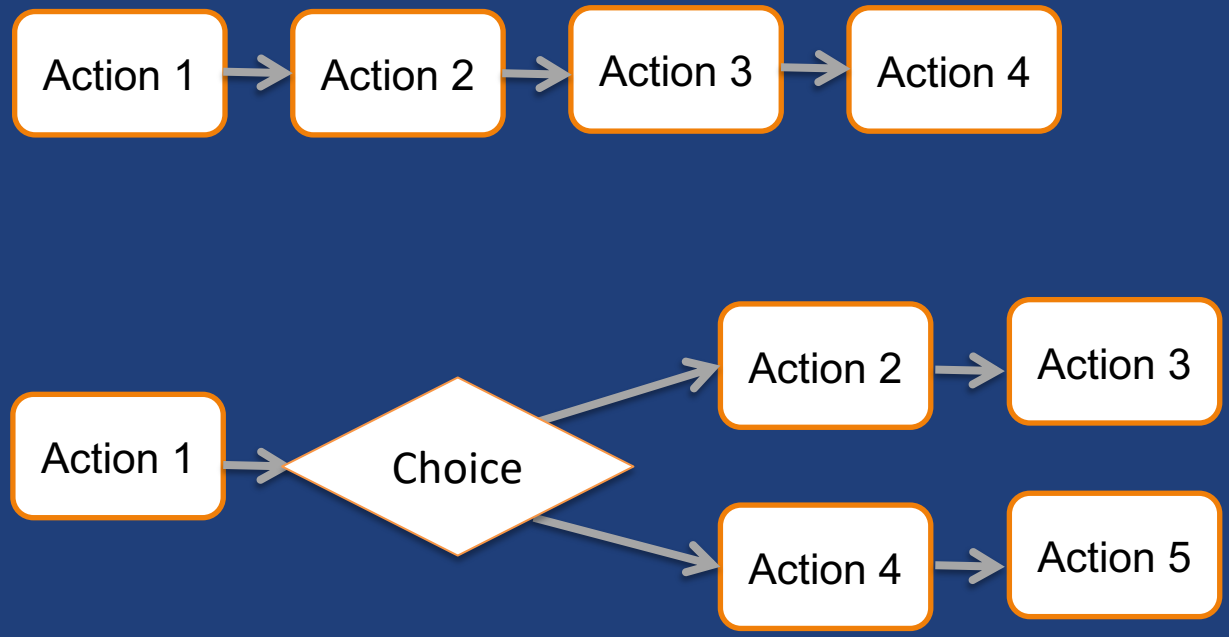
Personal system

Multiple ways to “automate” data management

- **Scripts using the Globus CLI (+ cron?)**
- **Globus Timer service → scheduled and recurring transfers**
- **Your own code calling the Globus APIs (ugh!)**

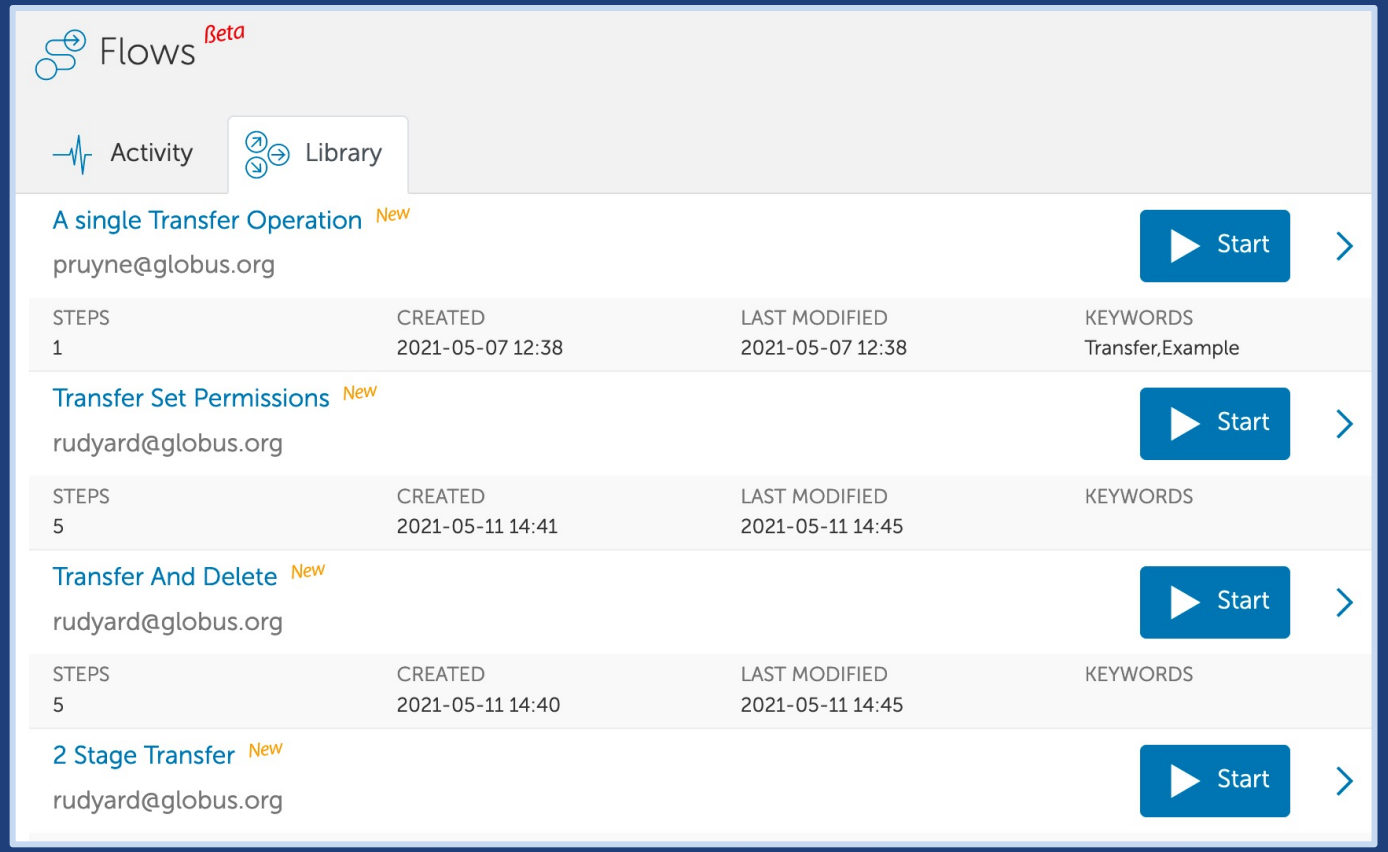
The best way to automate: Globus Flows

- **Flows comprise actions**
- **Define flow (using declarative language, JSON or YAML)**
- **Set policy: `visibility`, `runnable by`**
- **Deploy to Flows service**



Starting and managing runs

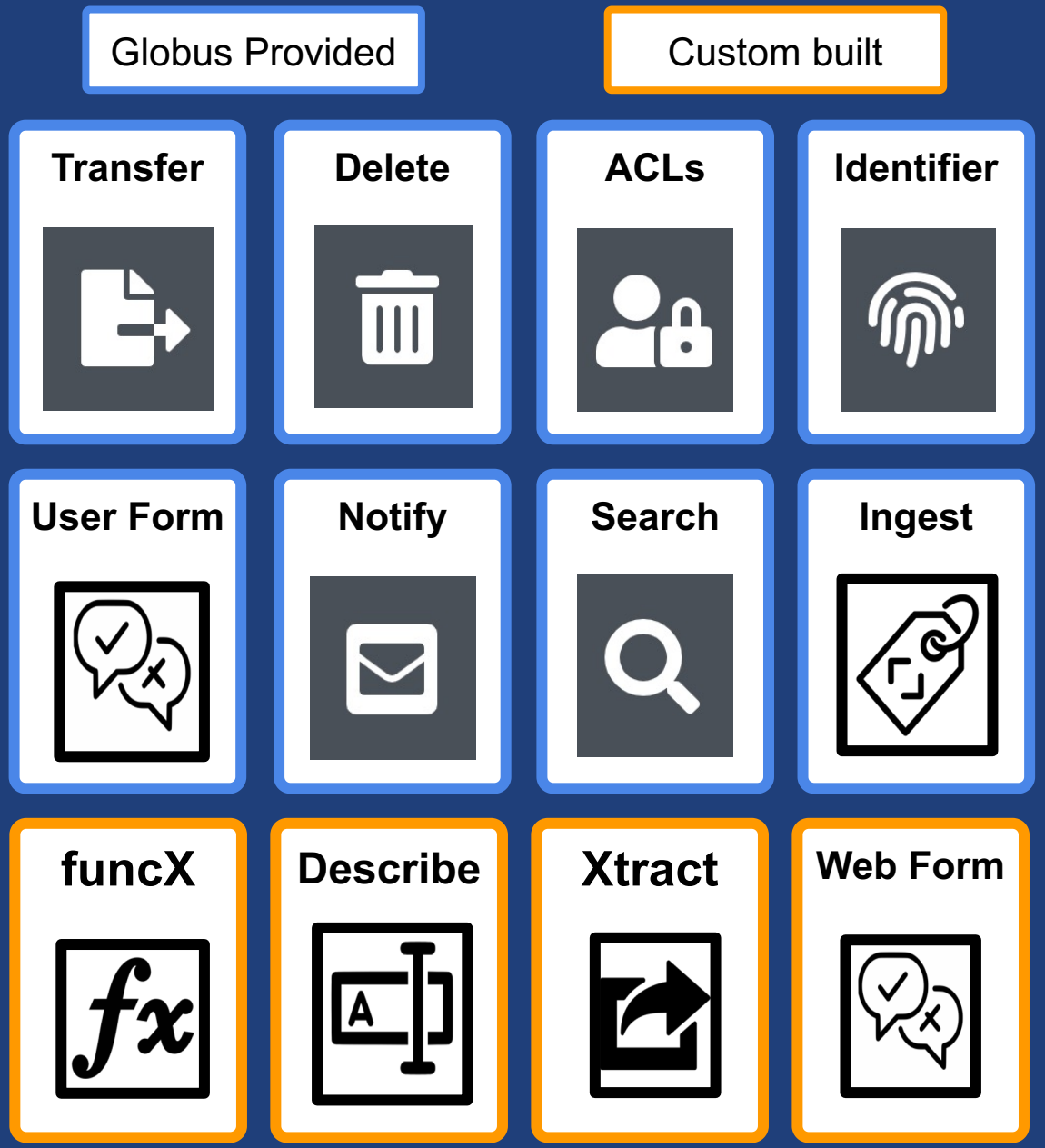
- **Start Flow execution**
  - Supply input parameters, e.g., collection IDs
- **Check run status**
- **Cancel run**
- **Set policy: monitor, manager**
- **Triggers to start flows**



Extending flows

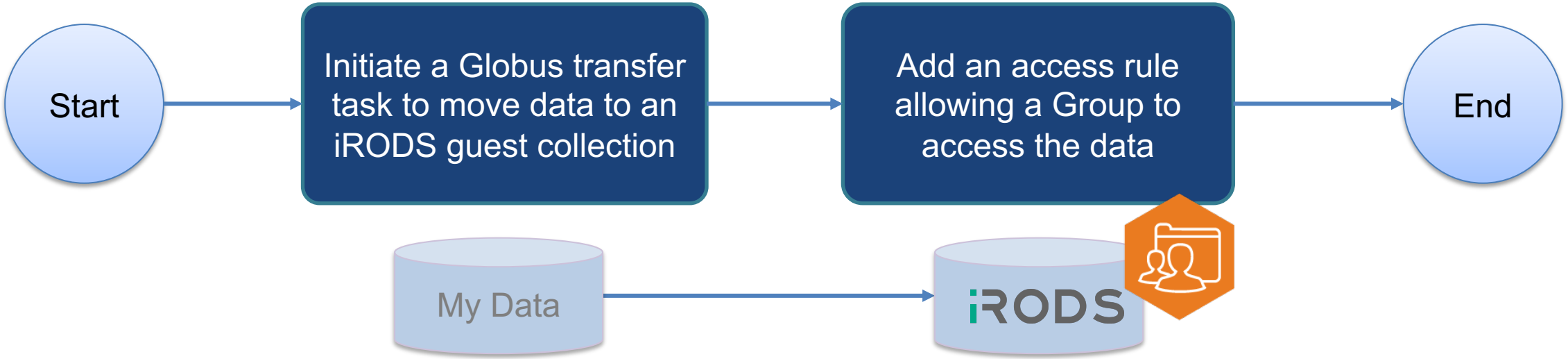
- **Actions execute against an Action Provider service endpoint**
- **Pre-built action providers for Globus services**
- **Build your own Action Providers using toolkit**

action-provider-tools.readthedocs.io



Demonstration

Move/share data on iRODS…



Demonstration

Make the data findable…

Resources

- Documentation: docs.globus.org
- YouTube channel: youtube.com/GlobusOnline
- Helpdesk and issue escalation: support@globus.org
- Mailing list: discuss@globus.org
- Open source clients/examples: github.com/globus
- Globus customer team: outreach@globus.org