Using the Globus Platform for Secure Data Sharing

Brigitte Raumann
braumann@uchicago.edu

AARNet and Globus Technical Catch-up
March 29, 2023

THE UNIVERSITY OF CHICAGO

Argonne NATIONAL LABORATORY

globus

1

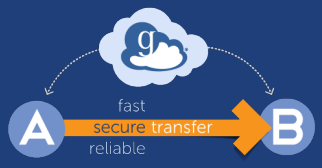
Globus is …

a non-profit service developed and operated by

THE UNIVERSITY OF CHICAGO

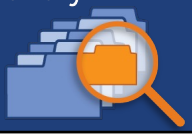
2

Globus platform for researchers

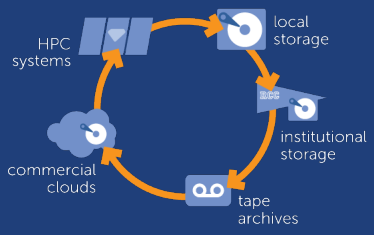


Data Transfer & Sharing

Publication & Discovery



Unified Data Access



Reliable Automation

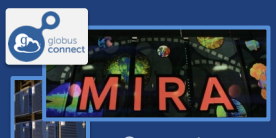
Platform-as-a-Service

Remote Execution

3

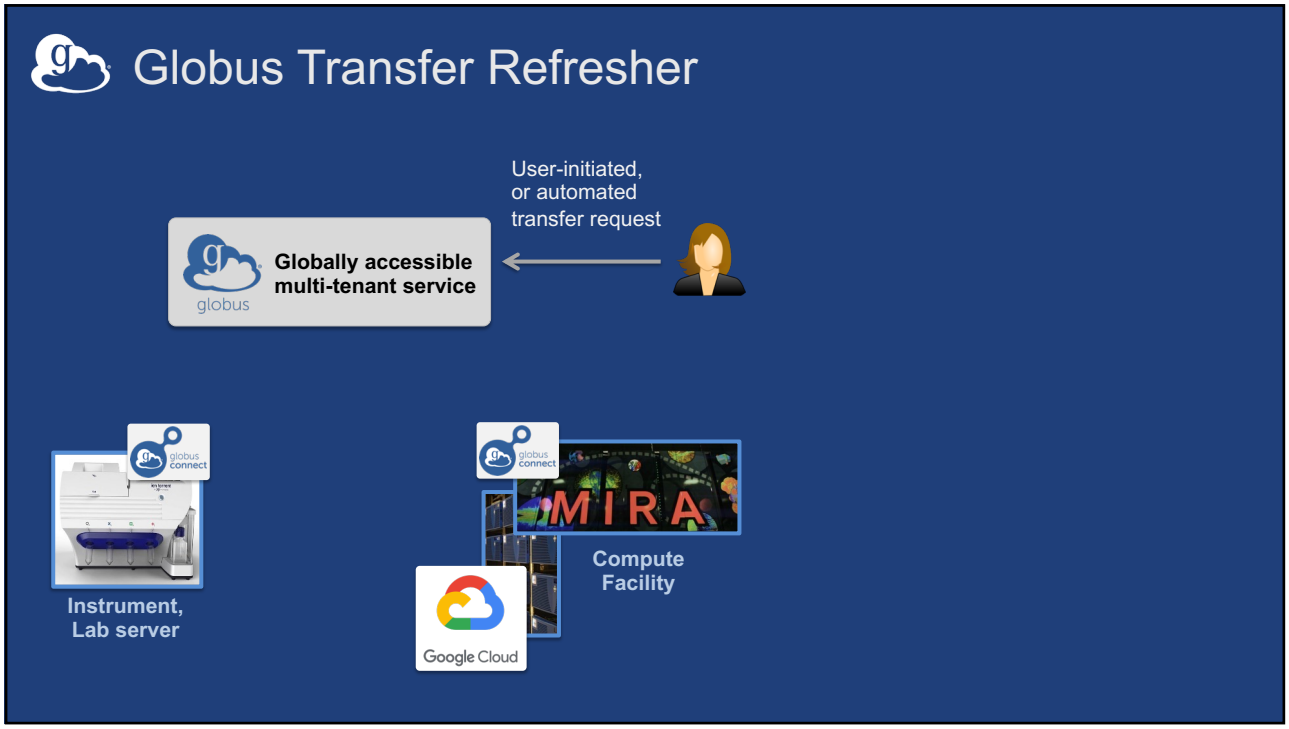
Globus Transfer Refresher

Instrument, Lab server

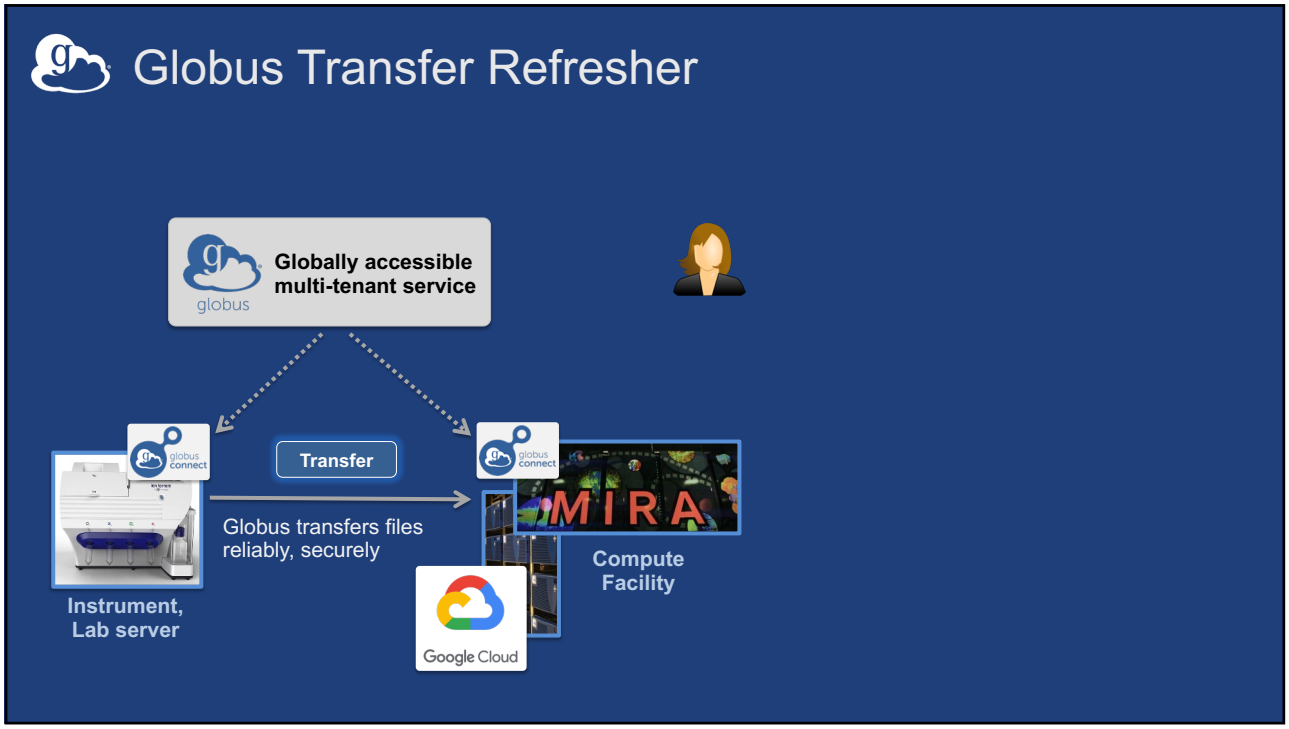


Compute Facility

4

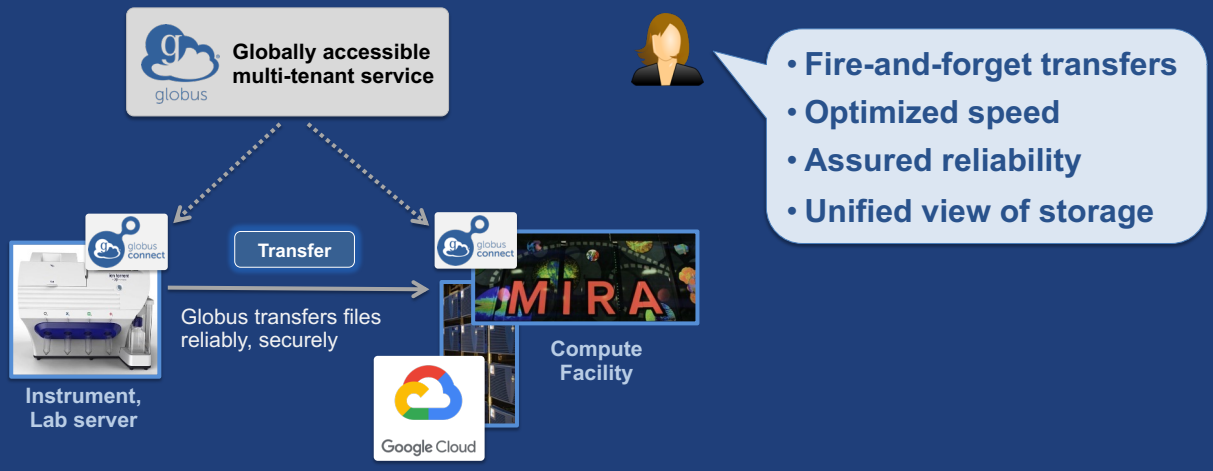

Globus Transfer Refresher
User-initiated,
or automated
transfer request
Globally accessible
multi-tenant service
globus
Instrument,
Lab server
MIRA
Compute
Facility
Google Cloud

5


Globus Transfer Refresher
Globally accessible
multi-tenant service
globus
Transfer
Globus transfers files
reliably, securely
Instrument,
Lab server
MIRA
Compute
Facility
Google Cloud

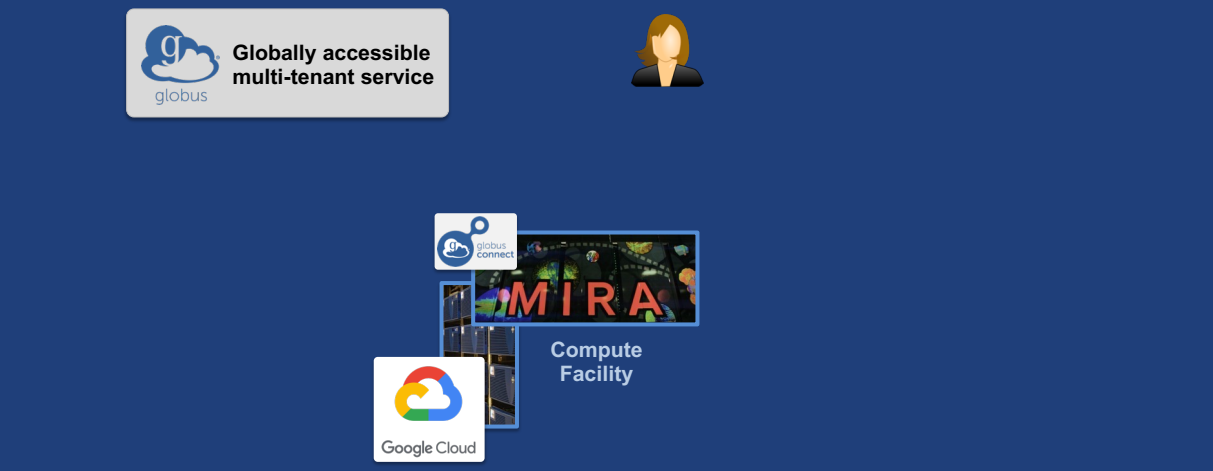
6

Globus Transfer Refresher

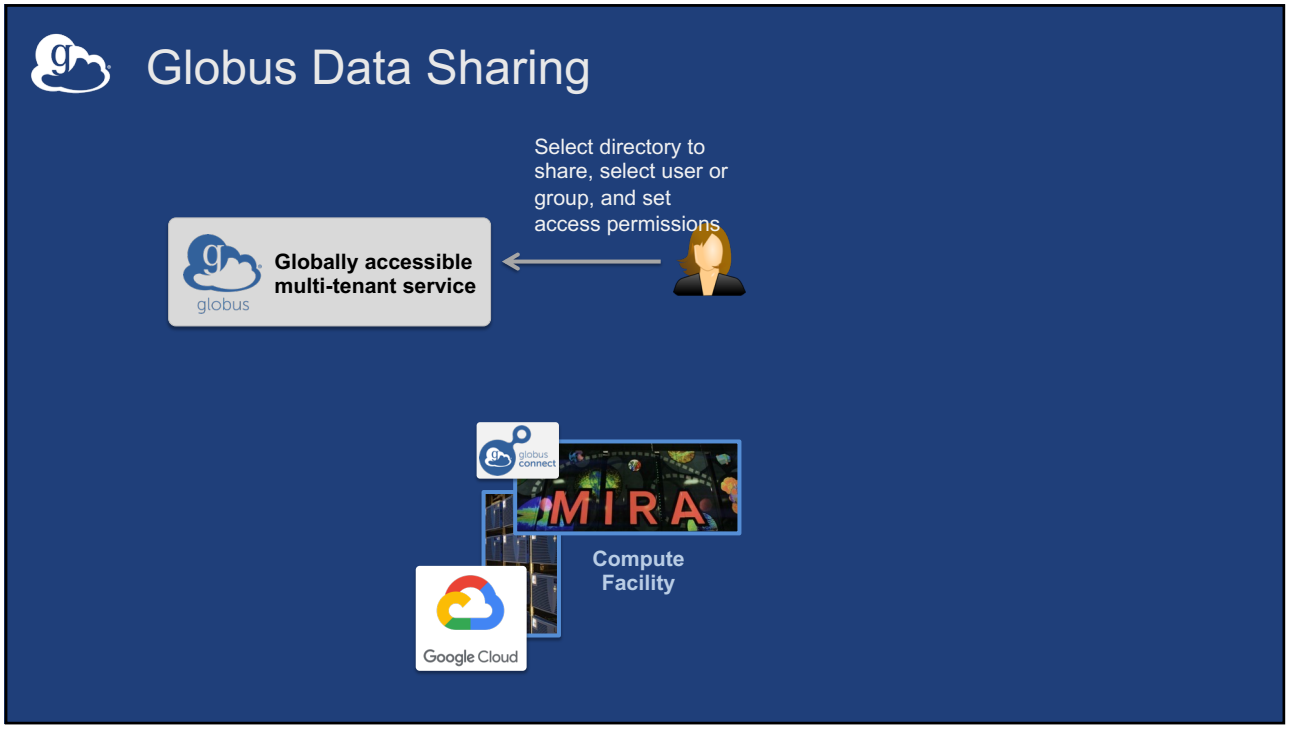


7

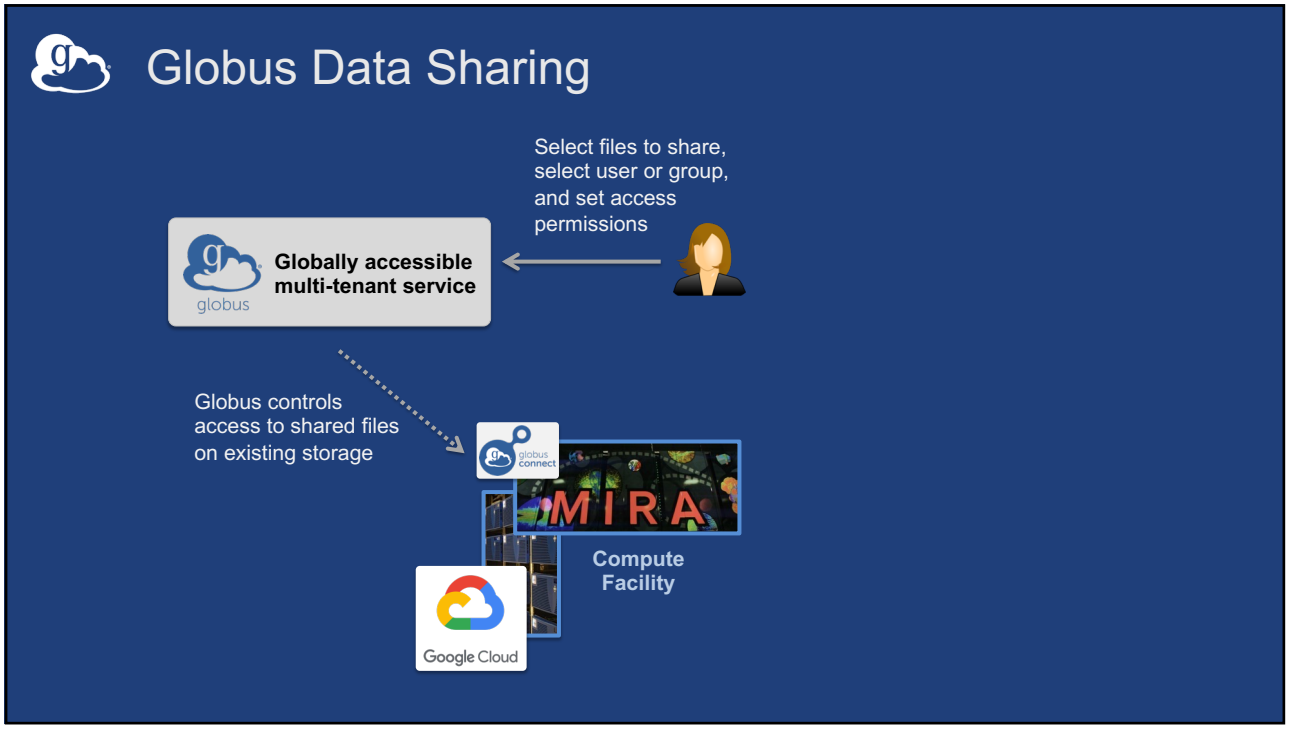
Globus Data Sharing



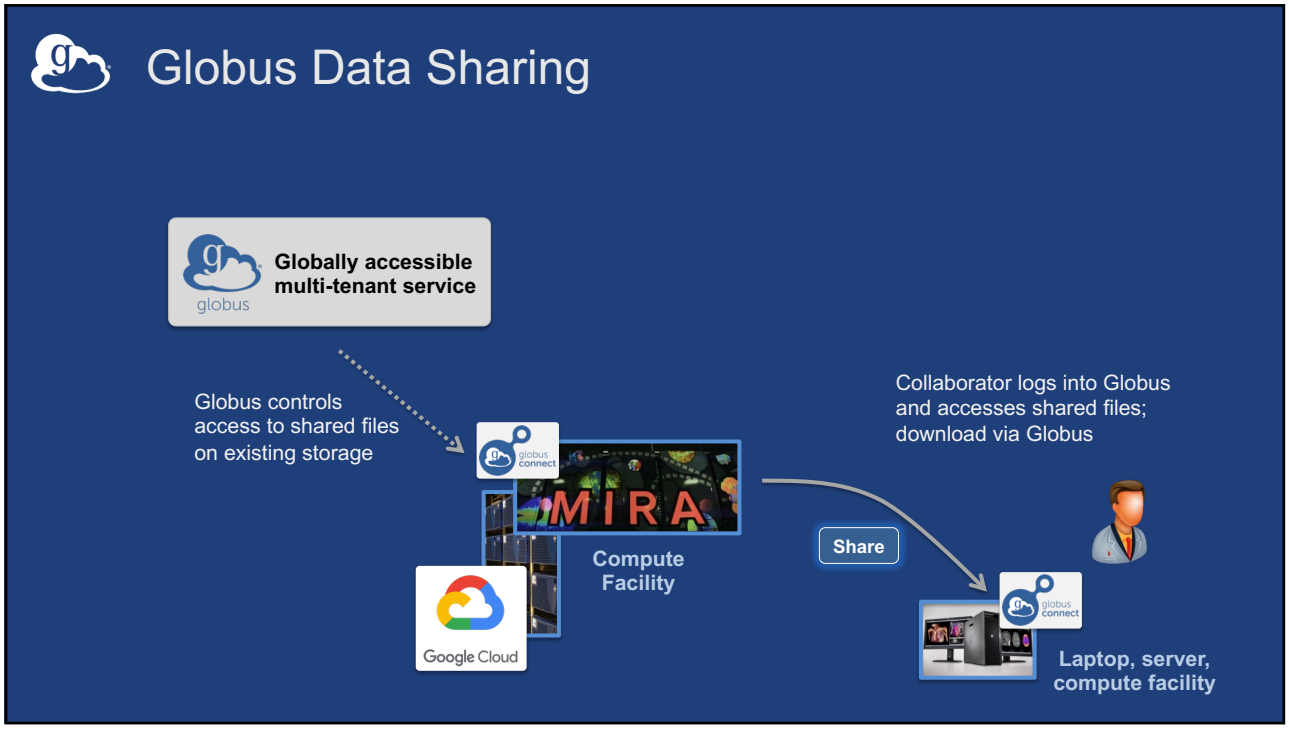
8


Globus Data Sharing
Select directory to share, select user or group, and set access permissions
Globally accessible multi-tenant service
globus
globus connect
MIRA
Compute Facility
Google Cloud

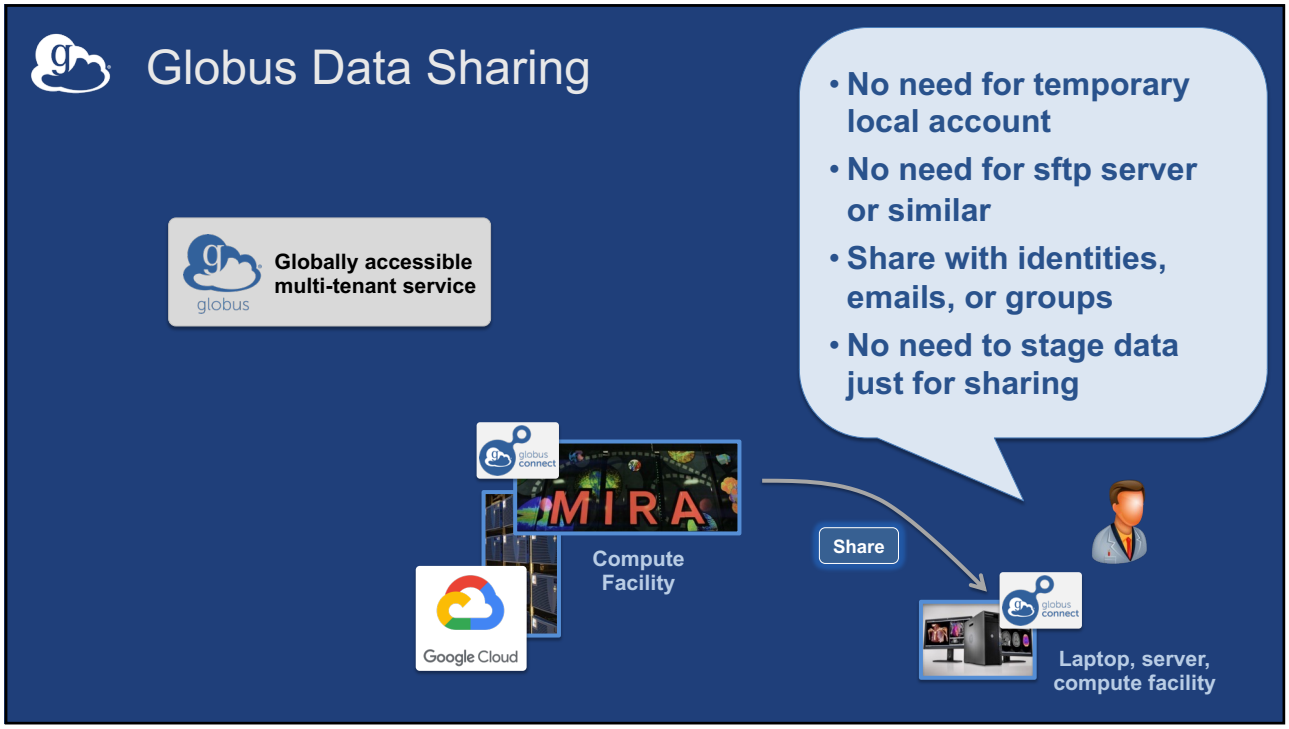
9


Globus Data Sharing
Select files to share, select user or group, and set access permissions
Globally accessible multi-tenant service
globus
Globus controls access to shared files on existing storage
globus connect
MIRA
Compute Facility
Google Cloud

10


Globus Data Sharing
Globally accessible multi-tenant service
globus
Globus controls access to shared files on existing storage
Collaborator logs into Globus and accesses shared files; download via Globus
MIRA
Compute Facility
Google Cloud
Share
Laptop, server, compute facility

11


Globus Data Sharing
Globally accessible multi-tenant service
globus
• No need for temporary local account
• No need for sftp server or similar
• Share with identities, emails, or groups
• No need to stage data just for sharing
MIRA
Compute Facility
Google Cloud
Share
Laptop, server, compute facility

12

What does “Globus Sharing” mean?

- Globus sharing is a way to grant Globus users that may not have a local account on your filesystem access to files and directories.
- Advantages of Globus sharing
  - No need to provision temporary local accounts
  - No need to move data to cloud storage, “sharing in place”
  - No need to set up sftp (or similar) services
  - Users can grant permissions and roles to other users, to groups, or to applications (client identities)
  - Access manager role grants others the rights to manage permissions

13

How does Globus enable sharing?

- Globus uses “overlay permissions”, allowing the specified Globus user (guest user) to act as the local user (mapped user) to read or read / write on a specific directory only.
- Local permissions are always enforced.
- Many levels of sharing controls are available.

14

Steps to share data using Globus

1. **Create a guest collection**
2. **Select directory to be shared**
3. **Select user or group to be shared with**
4. **Repeat steps 2 and 3 as needed**

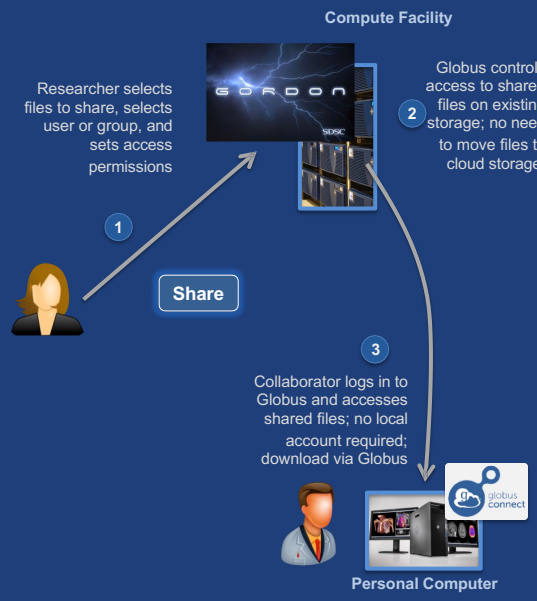
15

Demo

16

Ad hoc data sharing

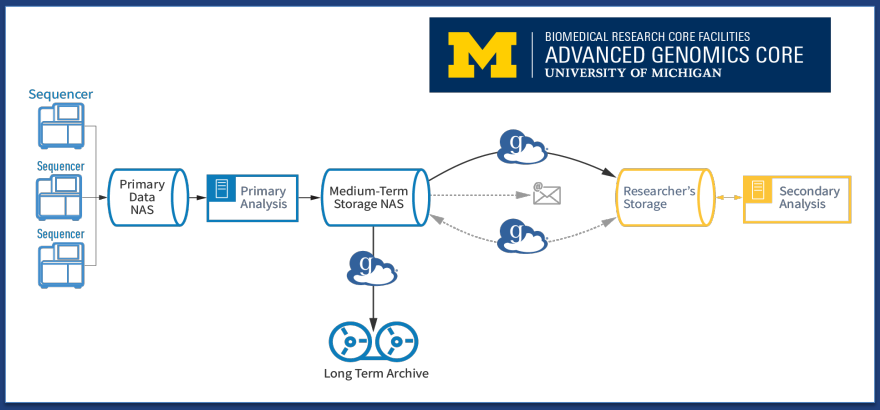
- **Researchers sharing with collaborators or community**
- **Making data publicly available**



17

Instrument data distribution

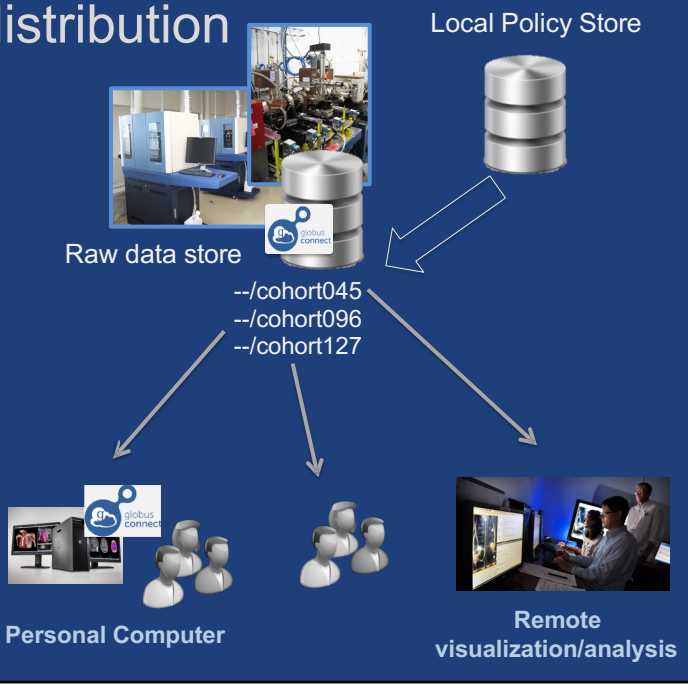
- **Secure data delivery to instrument core customer**
- **Automated data egress or ingress across tiers**



18

User facility data distribution

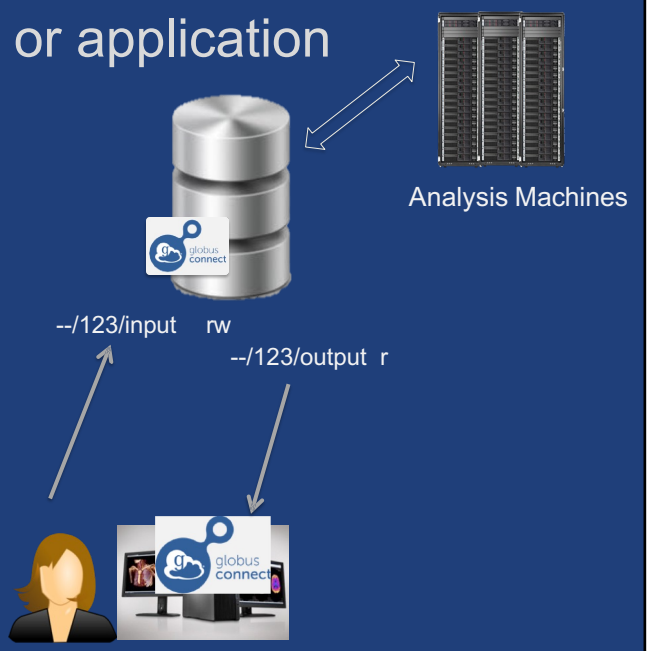
- Access permissions based on policy information at site
- Self managed by researcher
- Provide near-real time access to data



19

Data processing core or application

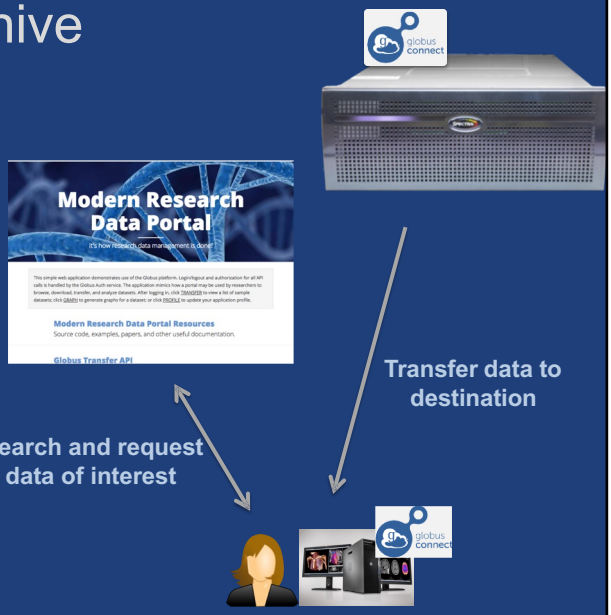
- Allow user to securely upload data for processing
- Post analysis make the user's results available to them
- Automate setup and tear down of folders and permissions



20

Data from provider/archive

- **Portal/science gateway to distribute data**
- **Interface for search and gathering of data of interest**
- **Data transfer**
  - Asynchronous
  - Directly to the user's machine
  - With authentication & authorization



21

Automating data sharing

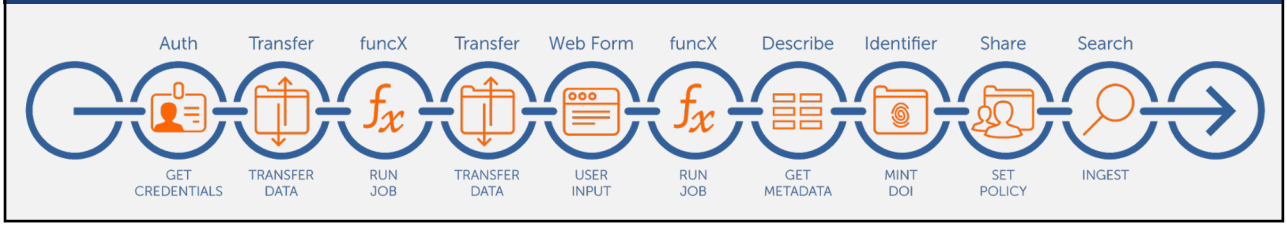
Permissions management can be automated

- **Scripts**
  - Globus CLI can be used to script sharing
  - Step by step video creating script to transfer and set permissions. https://www.youtube.com/watch?v=qIQTC6YOvrE
- **Globus Flows**
  - Managed task orchestration
- **Applications**
  - Custom application for automated permission management
  - App uses client identity and secret
  - Grant client the Group admin role or Access Manager role

22

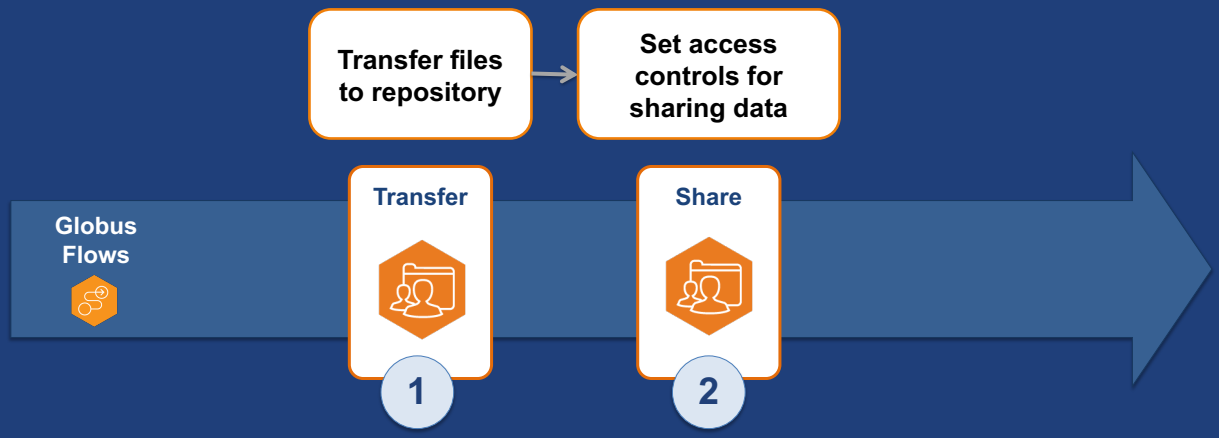
Globus Flows

Managed, secure, reliable task orchestration across **heterogenous resources**, using a declarative language for composition and an **event driven** execution model, **extensible** via custom actions, for automation at **scale**

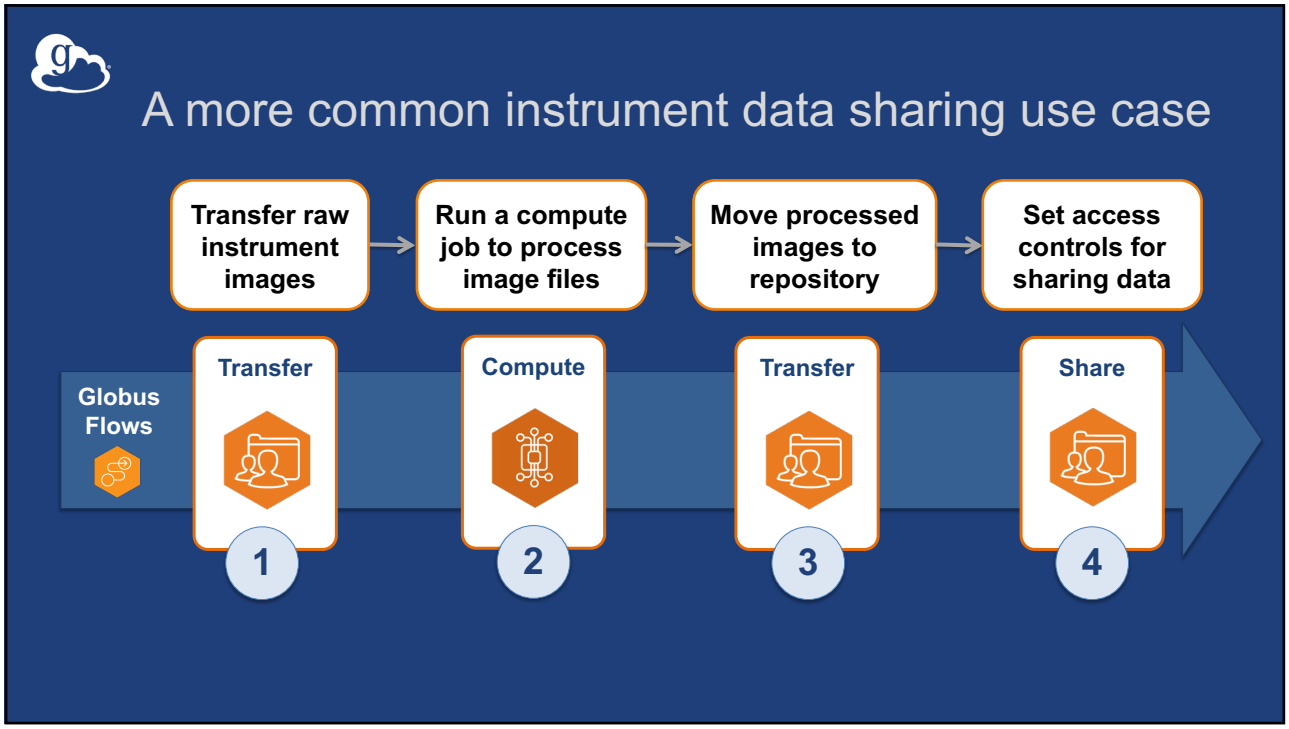


23

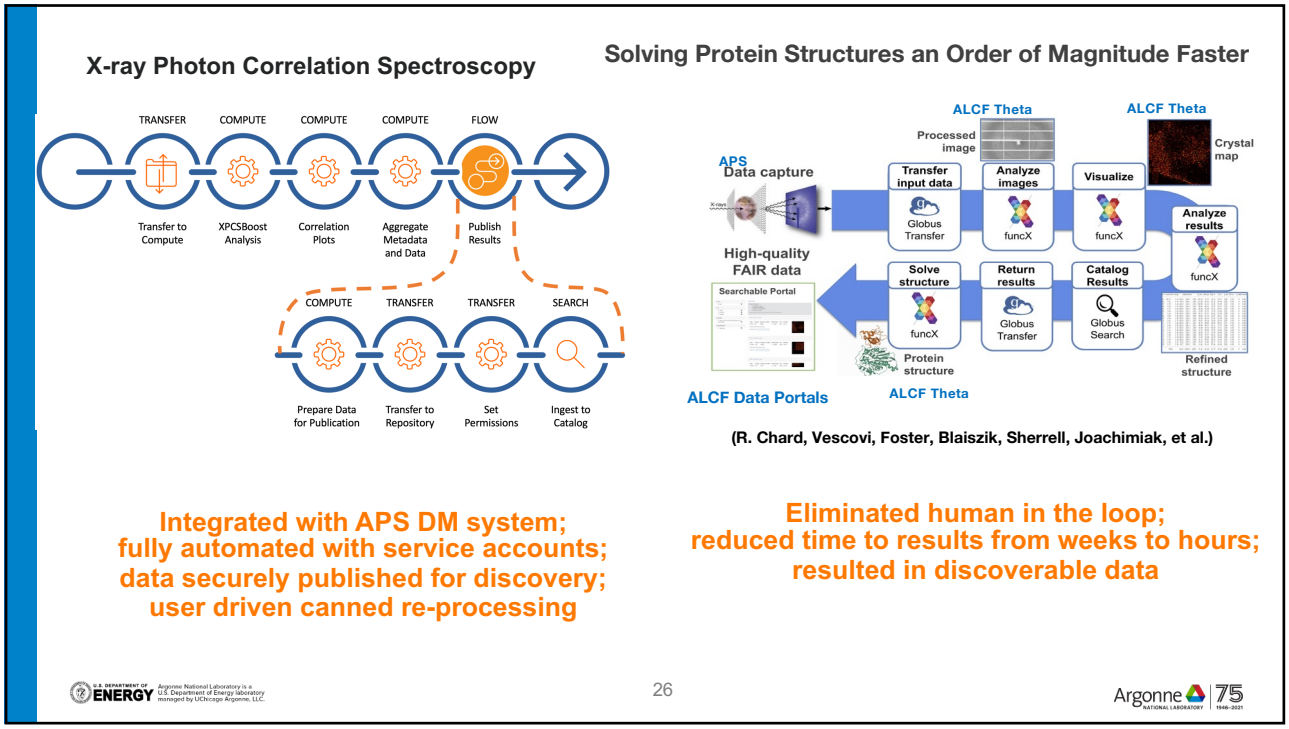
Automating simple sharing



24


A more common instrument data sharing use case
Transfer raw instrument images
Run a compute job to process image files
Move processed images to repository
Set access controls for sharing data
Globus Flows
Transfer
Compute
Transfer
Share
1
2
3
4

25


X-ray Photon Correlation Spectroscopy
TRANSFER
COMPUTE
COMPUTE
COMPUTE
FLOW
Transfer to Compute
XPCSBoost Analysis
Correlation Plots
Aggregate Metadata and Data
Publish Results
COMPUTE
TRANSFER
TRANSFER
SEARCH
Prepare Data for Publication
Transfer to Repository
Set Permissions
Ingest to Catalog
Integrated with APS DM system; fully automated with service accounts; data securely published for discovery; user driven canned re-processing
Solving Protein Structures an Order of Magnitude Faster
ALCF Theta
ALCF Theta
Processed image
Crystal map
APS
Data capture
Transfer input data
Analyze images
Visualize
Analyze results
Globus Transfer
funcX
funcX
funcX
High-quality FAIR data
Searchable Portal
Solve structure
Return results
Catalog Results
funcX
Globus Transfer
Globus Search
Protein structure
Refined structure
ALCF Data Portals
ALCF Theta
(R. Chard, Vescovi, Foster, Blaiszik, Sherrell, Joachimiak, et al.)
Eliminated human in the loop; reduced time to results from weeks to hours; resulted in discoverable data
26

26

Demo

27

Storage admin sharing controls

- **What can be shared**
  - Which collections, which paths, what level
- **Who can share**
  - Which users/groups
- **Who can share what**
  - Which users/groups can share which paths
- **Whom can be shared with**
  - What identity domains must guests have
- **View and delete sharing permissions**
- **File system permissions control guest access**

28

User sharing controls

- **What is shared**
  - What path, what level
- **With whom it is shared**
  - Anonymous access
  - All authenticated users
  - Specific identities/groups
- **Who can share**
  - Access managers

29

Sharing support resources

- **How to share data using Globus web app**
  - https://docs.globus.org/how-to/share-files/
- **Secure Sharing for Globus Endpoint Administrators and Subscribers**
  - https://www.youtube.com/watch?v=Vr5lXxcE1xE&t=523s
- **Someone just shared data with me, now what?**
  - https://www.youtube.com/watch?v=jdeMtubk8D4&t=15s
- **Jupyter Notebook for writing transfer and share flow**
  - https://github.com/globus/globus-jupyter-notebooks/blob/master/Automation_Using_Globus_Flows.ipynb
- **Example Flows for transfer and sharing**
  - https://github.com/globus/globus-flows-trigger-examples

30

Support resources

- **Globus documentation: docs.globus.org**
- **YouTube channel: youtube.com/GlobusOnline**
- **Helpdesk: support@globus.org**
- **Mailing Lists: globus.org/mailing-lists**
- **Customer engagement team (office hours)**
- **Professional services team (advisory, custom work)**

31