



# Globus for Researchers

Vas Vasiliadis - [vas@uchicago.edu](mailto:vas@uchicago.edu)

May 23, 2023



THE UNIVERSITY OF  
CHICAGO





Word of the Day

# ocellated

*(OS-uh-lay-tid)*

**adjective: eyelike;  
having eyelike spots**



From: A.Word.A.Day (wordsmith.org)



Globus is ...

a non-profit service  
developed and operated by



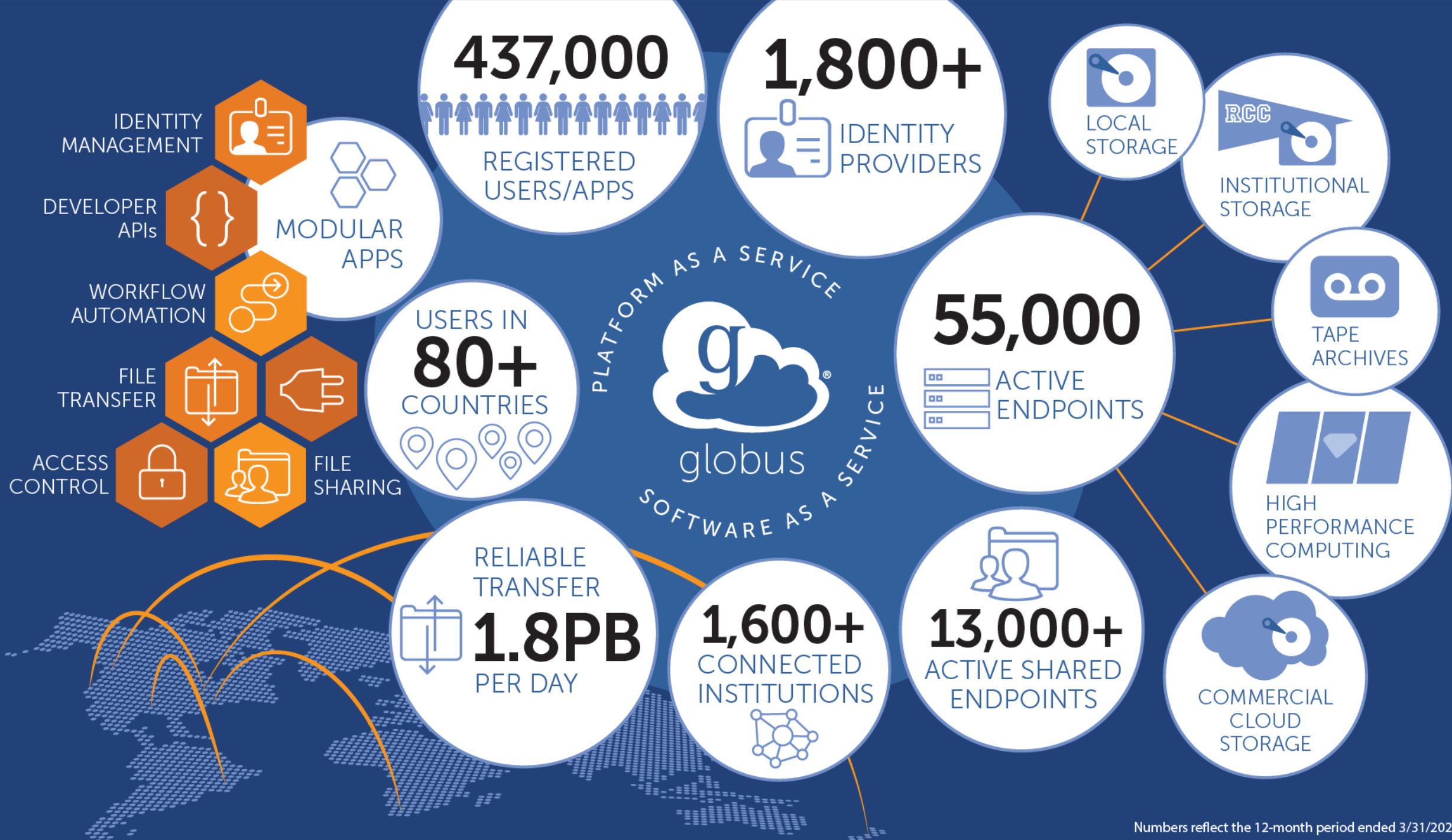
THE UNIVERSITY OF  
CHICAGO



Our mission is to...

increase the efficiency and  
effectiveness of researchers  
engaged in data-driven  
science and scholarship  
through *sustainable* software.







# Reimagining research IT with Globus



**Managed transfer & sync**



**Publication & discovery**



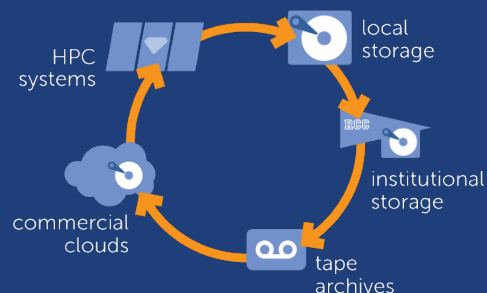
**Platform-as-a-Service**



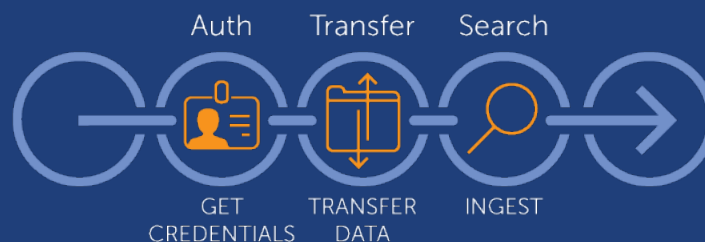
**Collaborative data sharing**



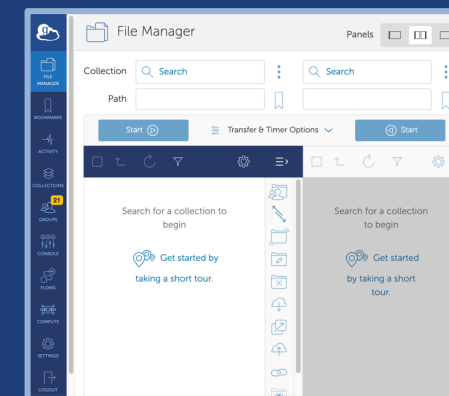
**Managed remote execution**



**Unified data access**



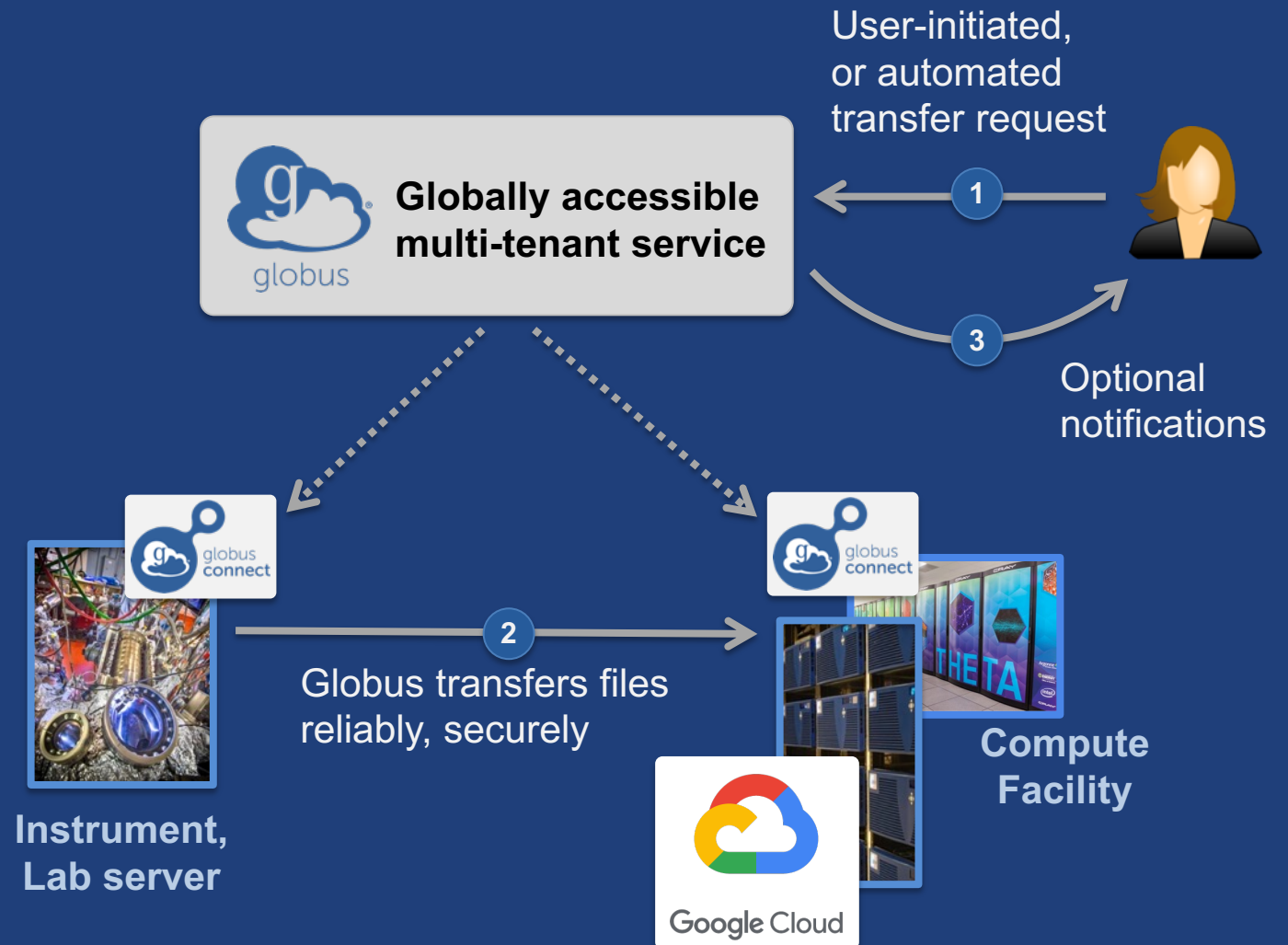
**Reliable automation**



**Software-as-a-Service**

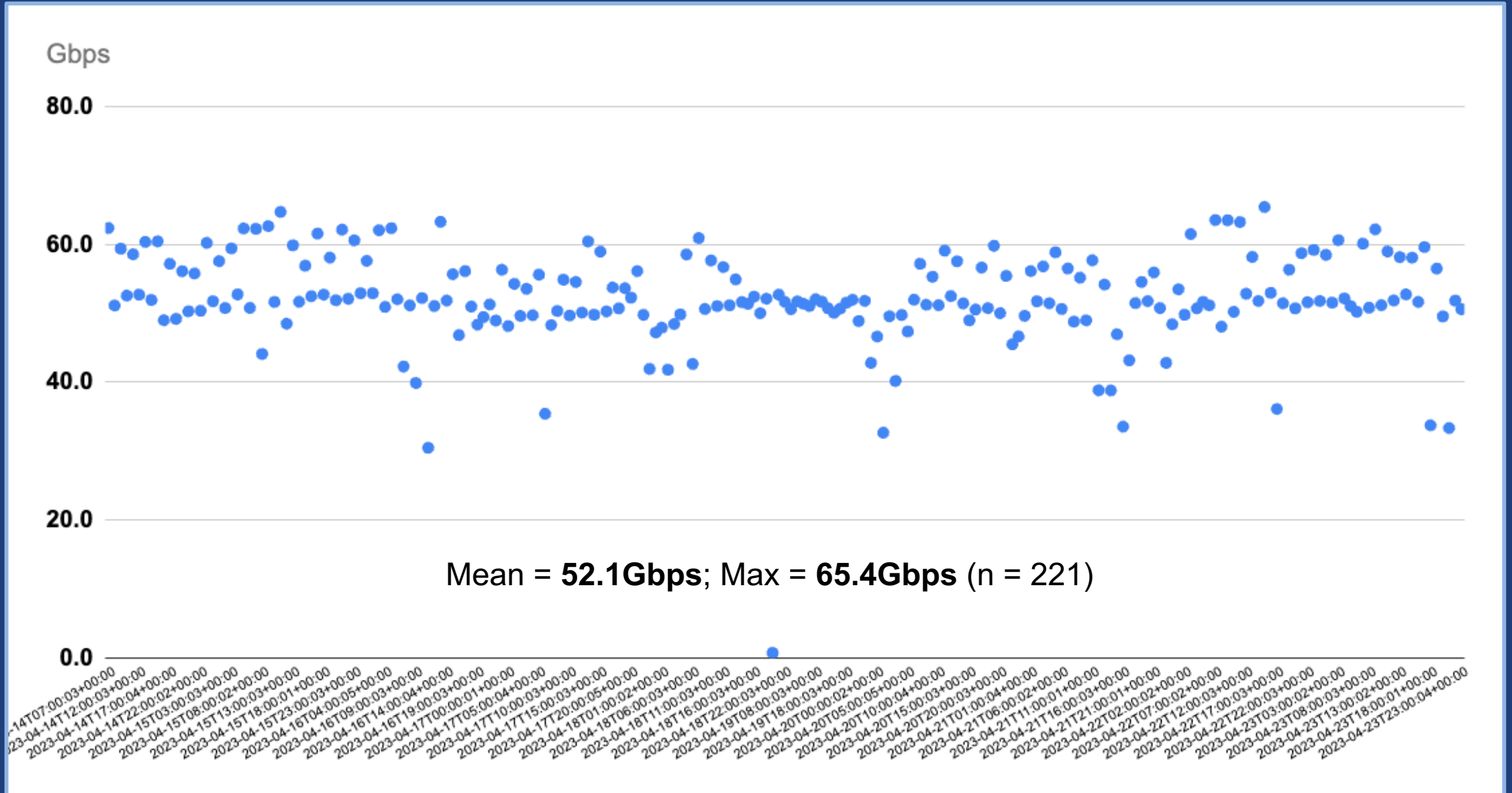
# Fast, reliable file transfer ...from any to any system

- Fire-and-forget transfer, sync
- Optimized speed
- Assured reliability
- Unified view of storage
- HTTP/S access to data





# NCAR–Boulder, CO $\leftrightarrow$ ANL–Darien, IL (RTT ~23ms)





# Globus is reliable, secure ...and fast!

Activity List



RDA to ALCF noverify

transfer completed

## 72.8Gbps



Overview



Event Log

Task Label	RDA to ALCF noverify
Source	NCAR RDA Dataset Archive ⓘ
Destination	DME PerfTest - Argonne ⓘ
Task ID	20ebf766-a46d-11eb-8a95-d70d98a40c8d
Owner	Vas Vasiliadis (vas@globusid.org)
Condition	SUCCEEDED
Requested	2021-04-23 02:50 pm
Completed	2021-04-23 02:53 pm
Duration	2 minutes 47 seconds
Transfer Settings	<ul style="list-style-type: none"><li>• transfer is not encrypted</li><li>• overwriting all files on destination</li></ul>

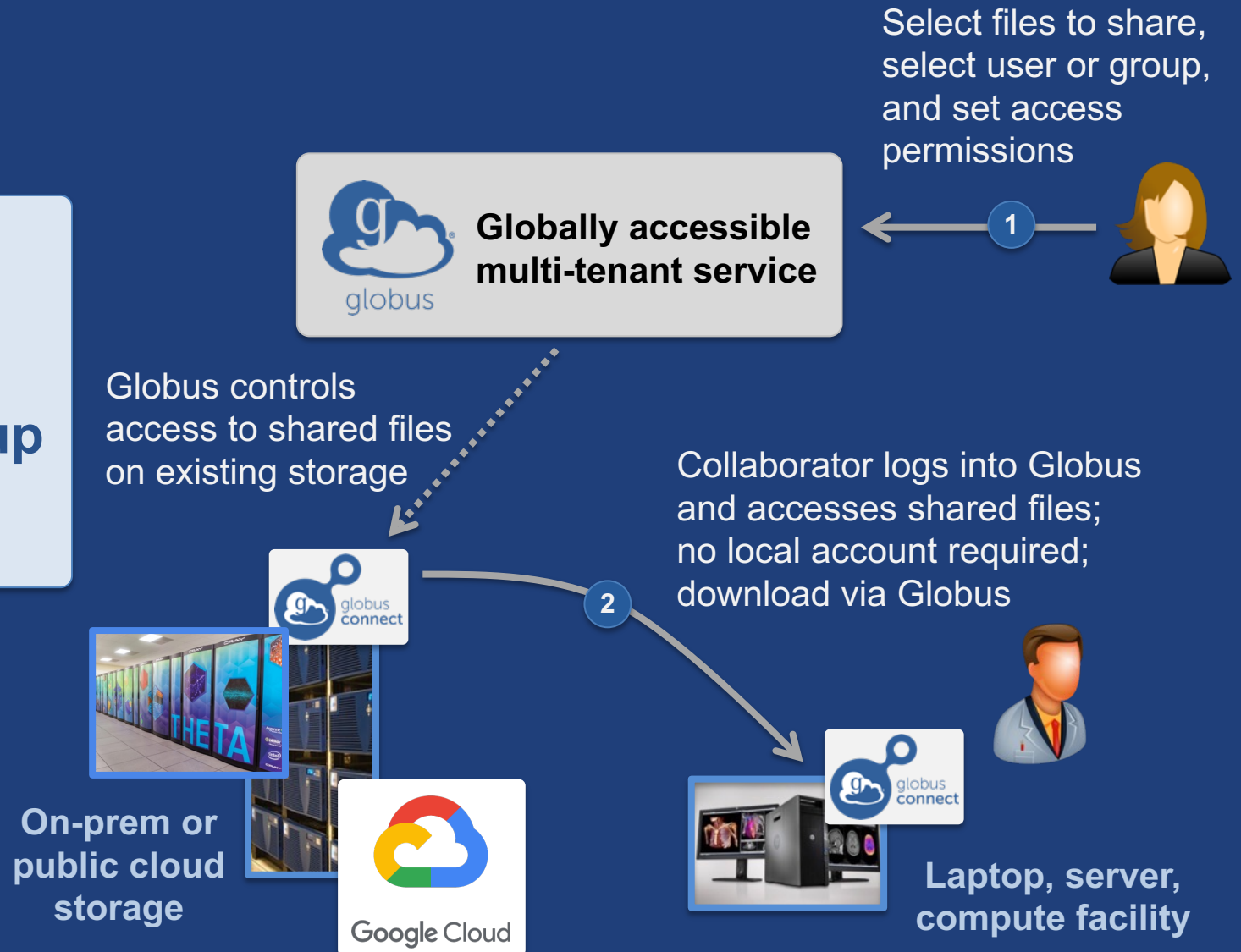
6151	Files
2	Directories
1.51 TB	Bytes Transferred
9.10 GB/s	Effective Speed
0	Skipped files on sync
0	Skipped files on error

[View debug data](#)



# Secure data sharing ...from any storage

- Fine-grained access control
- Storage system “overlay”
- Share with identity/email/group
- No data staging required

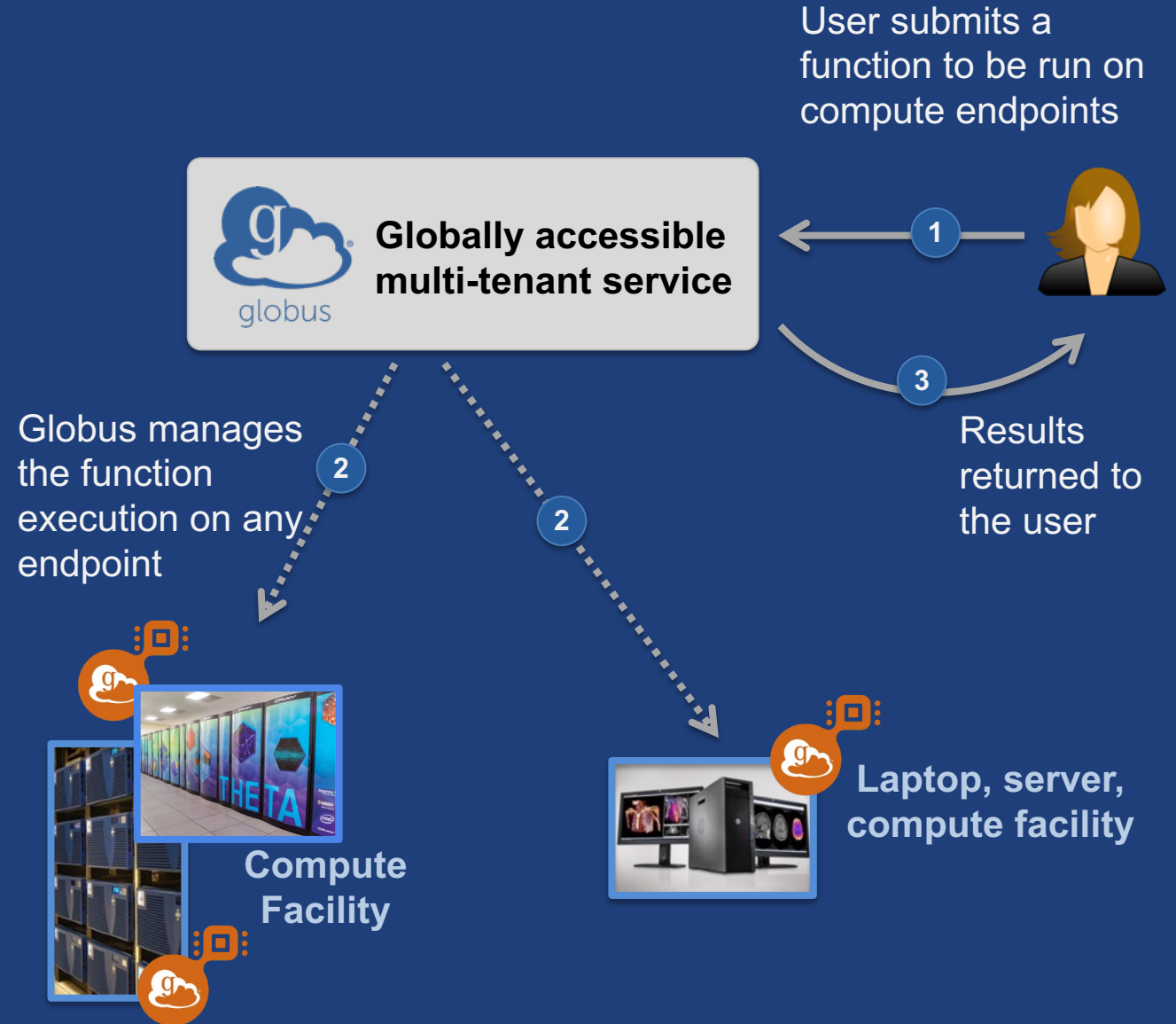


# Globus supports diverse storage systems



# Managed compute ...on any system

- Fire and forget function execution
- Federated authentication, and local access control
- Uniform interface to diverse compute resources
- Supports Python functions
- ...and everything else by wrapping as Python :-)

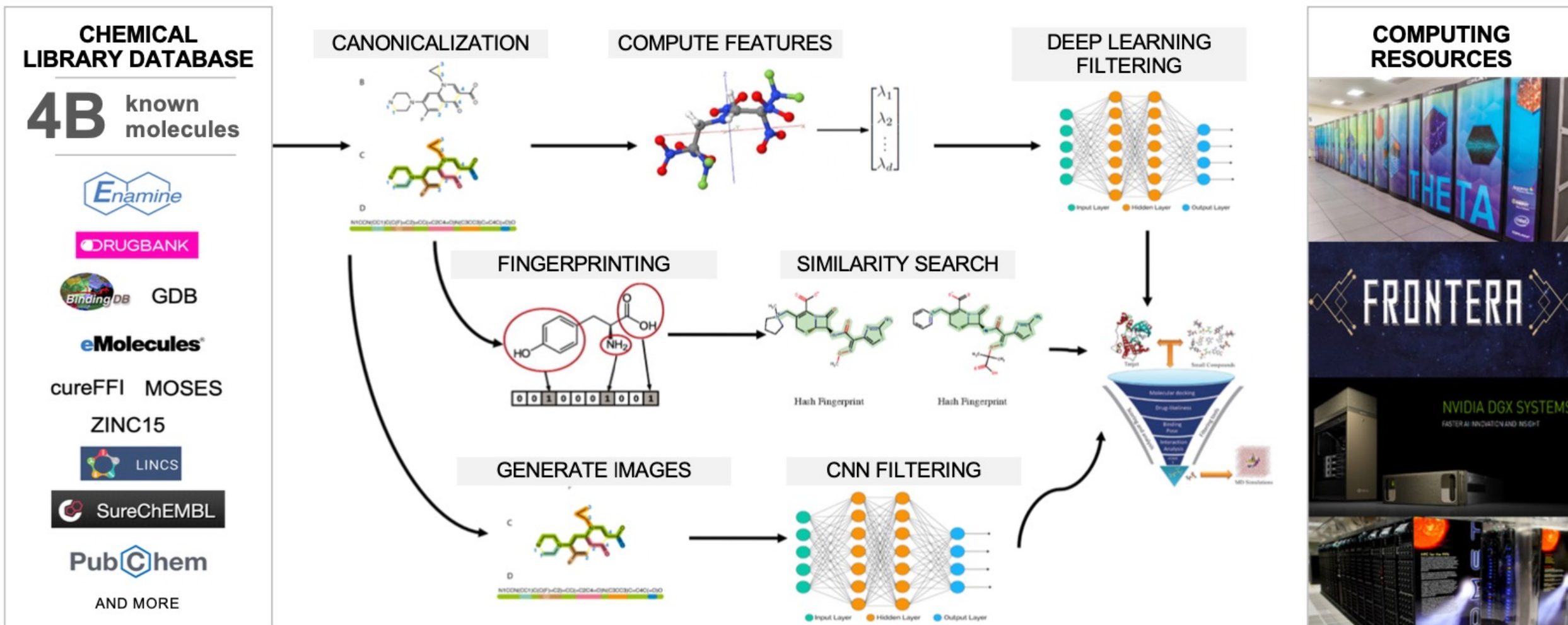




# The Globus Compute model

- **Compute service** — Highly available cloud-hosted service for managed, fire-and-forget function execution
- **Compute endpoint** — Abstracts access to compute resources, from edge device to supercomputer
- **Compute SDK** — Python interface for interacting with the service, suing the familiar (to many) Globus look and feel
- **Security** — Leverages Globus Auth; Compute endpoints are resource servers, authN and access via tokens

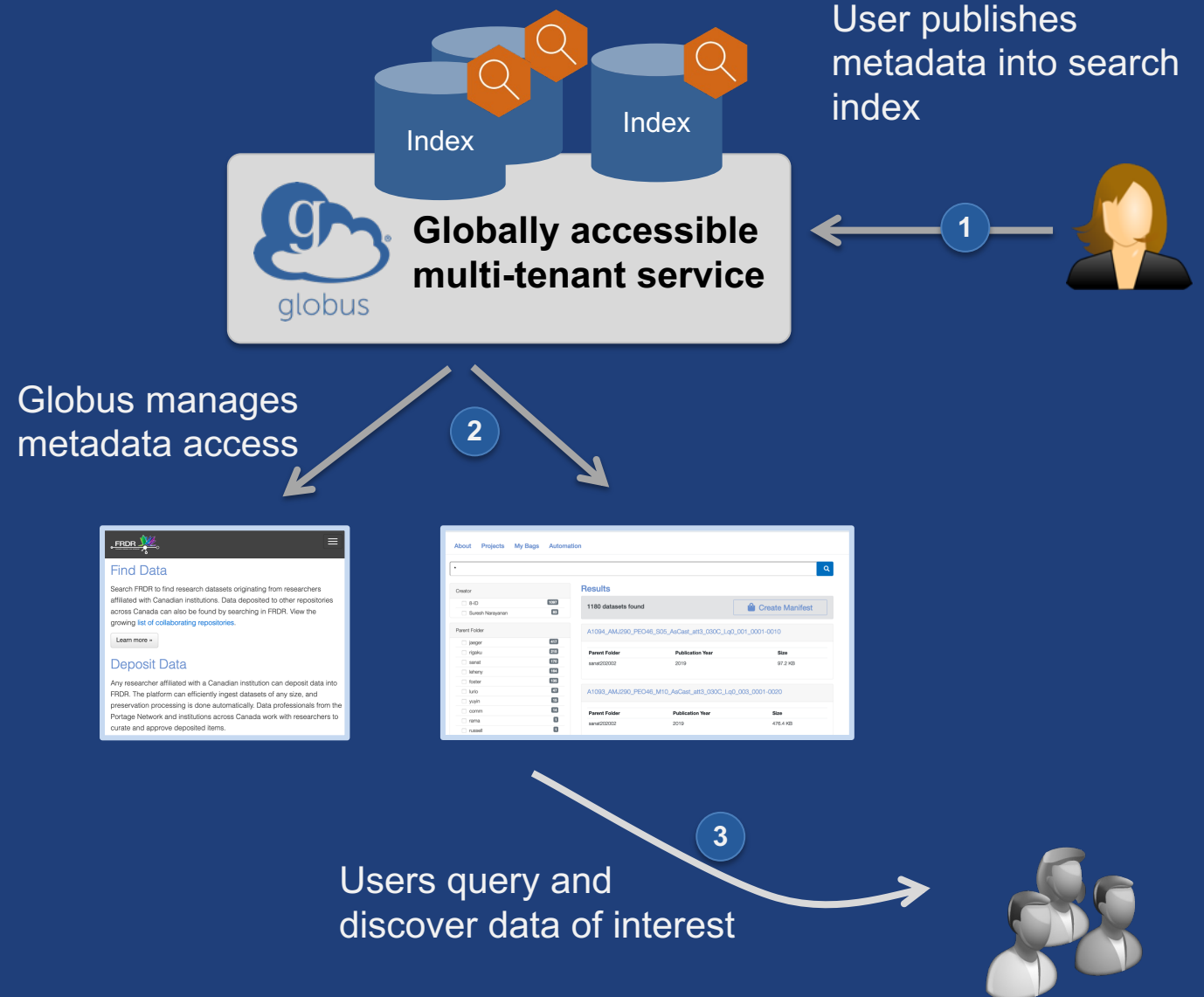
# ML-based drug screening





# Scalable data discovery ...for any domain

- Scalable metadata store
- Fine grained visibility controls
- Schema agnostic
- Federated auth integration
- Query API with facets





# Repository data distribution

Search results interface for ALCF Community Data Co-Op. The interface shows a search bar, filters for Creator (Suresh Narayanan), Parent Folder (sanat, foster, rama), APS Cycle (2018-3, 2019-1, 2020-1), and Dates (Dec 18 2019, Apr 01 2020, Apr 02 2020, Apr 07 2020, Mar 24 2021). The Results section shows 83 datasets found, with a list of dataset IDs and a 'Reprocess Datasets' button.

ALCF Community Data Co-Op interface showing various data categories and search options. Categories include APS 2D XRF, APS HEDM, APS Ptychography, APS XPCS, DLHub, Kanzus Beamline Chip Analysis, Neurocartography, and The Materials Data Facility. Each category has a 'Search This Project' button.

Globus logo and text: Globally accessible multi-tenant service

aws | S3 logo



Example: **acdc.alcf.anl.gov**

Search, request data of interest

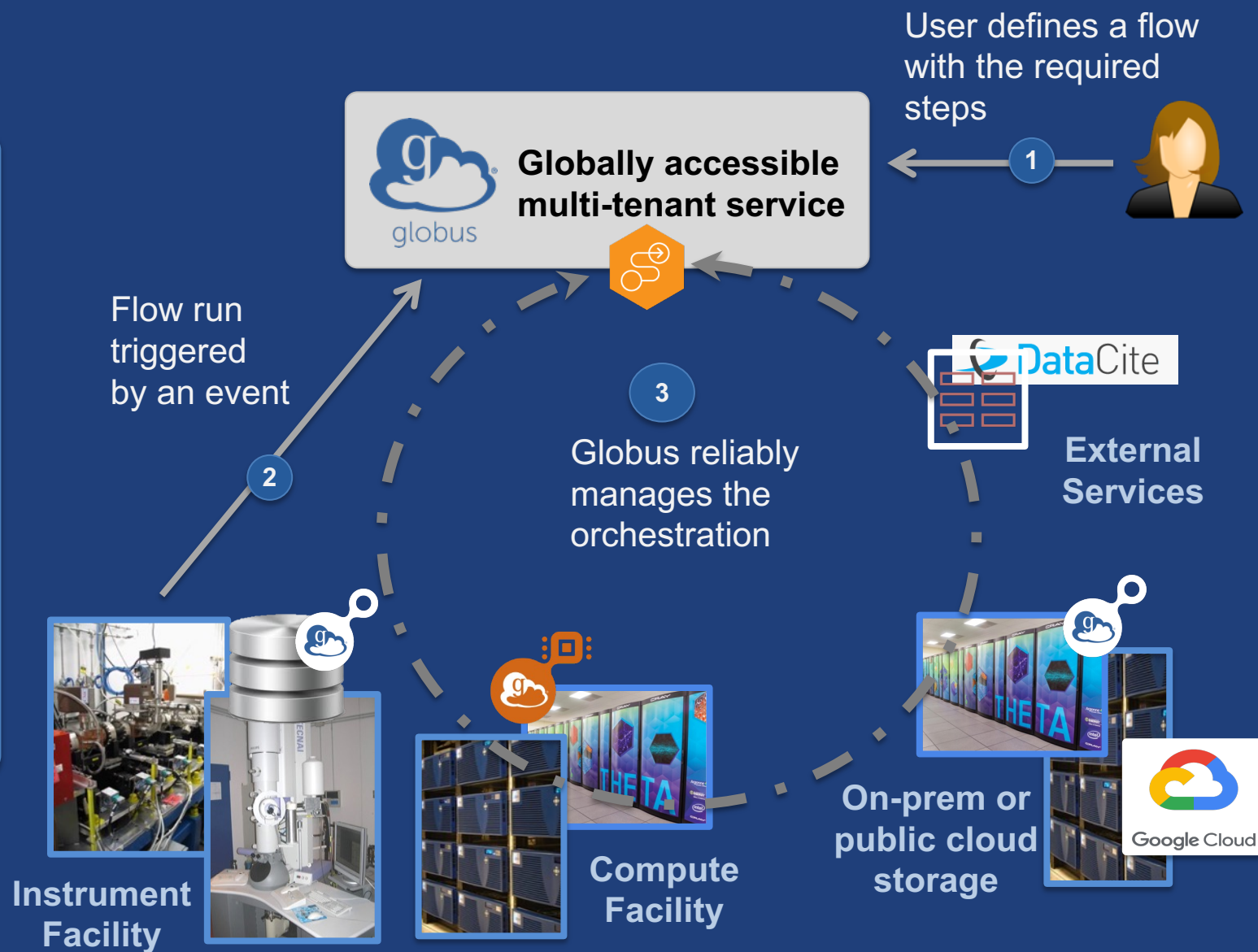
Browser based download

Bulk data transfer



# Reliable automation ...spanning all resources

- **Managed reliable task orchestration**
- **Declarative language for flow definition**
- **Event driven execution model**
- **Extensible for integrating external services**





# Automating cryoEM flows

Globus  
Flows

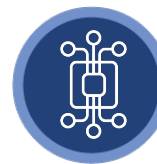


Transfer



Transfer  
raw files

Compute



Launch  
analysis job

Carbon!



Correct,  
classify, ...

Compute



Extract  
metadata

Share



Set access  
controls

Transfer



Move final  
files to repo

Search



Ingest to  
index

ALCF Community Data Co-Op / [redacted] / Serial Crystallography Search

Run

Run	Count
S9	14
S8	6
S10	5
SSX	5
S15	4
S14	4
S18	2
S12	1
S16	1
S17	1

Date Processed

Date Processed	Count
2020	45
2021	22

Results

Search Statistics

- 68 datasets found
- CBF Files: 23615.4
- Number of Ints: 2961.5

Statistics are approximate, and may deviate up to 10% from actual values

Brisk2\_images

Chip	Protein	CBFs Processed	CBF Range	Ints
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

Int Files: [redacted]

Full Size Image Preview: [redacted]

Created: Dec. 17, 2021, 7:54 p.m.

cryoSPARC

RELIION

cryoDRGN

Statistics

	This week	This month
Projects	5	4
Workspaces	14	13
Jobs	119	86
Completed Jobs	102	76

Change Log

My Recent Jobs

Links

- General Reference
- Tutorials
- Discussion Forum



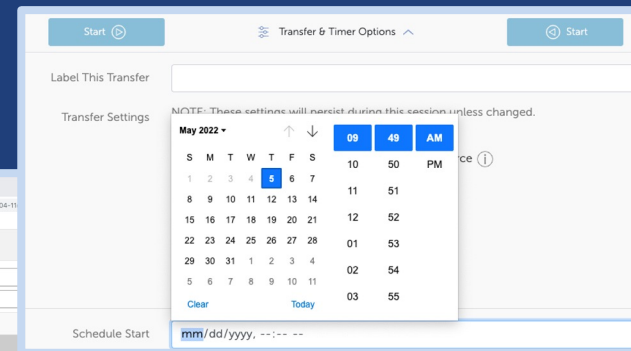
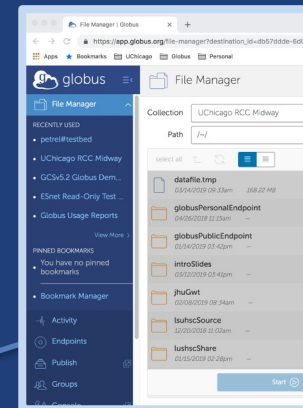
# One service, many interfaces

```
(globus-cli) jupiter:~ vas$ globus
Usage: globus [OPTIONS] COMMAND [ARGS]...

Options:
  -v, --verbose          Control level of output
  -h, --help             Show this message and exit.
  -F, --format [json|text] Output format for stdout. Defaults to text
  --map-http-status TEXT Map HTTP statuses to any of these exit codes:
                        0,1,50-99. e.g. "404=50,403=51"

Commands:
  bookmark      Manage Endpoint Bookmarks
  config        Modify, view, and manage your Globus CLI config.
```

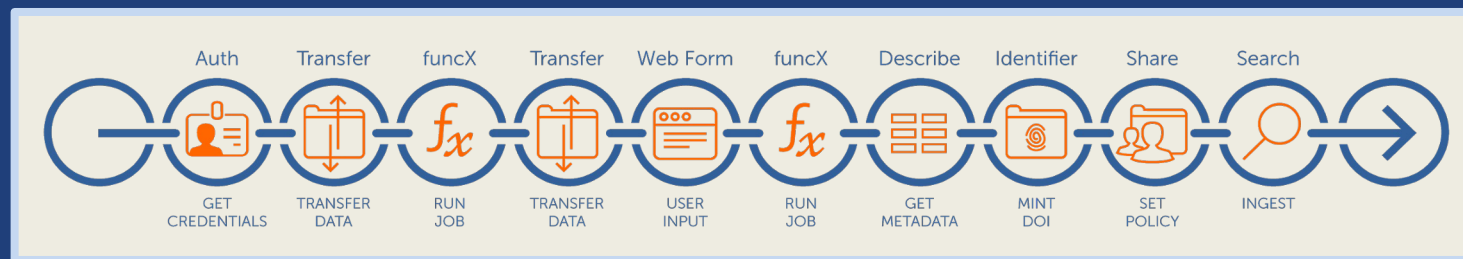
CLI



Web



Globus service



Flows

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

Rest  
API



# Plugging into the Globus ecosystem

- **Globus Connect Server**  
– Data access on multi-user systems   
[docs.globus.org/globus-connect-server](https://docs.globus.org/globus-connect-server)
- **Globus Connect Personal**     
– Data access on personal systems  
[globus.org/globus-connect-personal](https://globus.org/globus-connect-personal)
- **Globus Compute Agent**  
– Compute access in single-user environments  
[globus.org/compute](https://globus.org/compute)





# Hand-on time!

## Follow along on laptop virtual machine



## Watch demonstration (and do it yourself)



# To participate...

Recent version of web browser

SSH client for later sessions





# Let's see how to...



**Authenticate**

**Search collections and transfer files**

**View transfer activity and logs**

**Up/download files via browser**



# Install Globus Connect Personal on your laptop







# Some cool stuff...

Sharing data

Linking identities

Creating bookmarks





# Globus core security features



- **Access Control**
  - Identities provided and managed by institution
  - Institution controls all access policies
  - Globus is identity broker; no access to/storage of user credentials
- **Data remain at institutions, no storage/routing via Globus**
- **Integrity checks of transferred data**
- **Enforced encryption of Globus control data**
- **Institution-configured encryption of user data in transit**



# Globus High Assurance features



- **Additional authentication assurance**
  - Reauthenticate after specified time period
  - Authenticate with a *specific identity* within session
- **Per application, per session auth context isolation**
- **Enforced encryption of all user data in transit**
- **Local audit logging**
- **Option to require MFA**
- **Blocked anonymous or open sharing**



# Globus High Assurance for managing protected data

## Security controls

- NIST 800-53
- NIST 800-171



## Restricted data handling

- PHI, PII, CUI
- Compliant data sharing

## BAA w/Uchicago

- UChicago BAA with Amazon

# Move without (worrying about) limits

- API request rates
- File size
- Data volume
- Third-party tools cannot circumvent...
- ...but Globus lets you “fire-and-forget”
- → it will (eventually) be done

```
3/9/2022, 08:33 PM    endpoint too busy    View details ^

Error (transfer)
Endpoint: Vas Google Drive Collection (d6d62391-fdda-4ba5-ac78-6523f806ea79)
Server: m-422a8b.d8b83.36fe.data.globus.org:443
File: /My%20Drive/migration/uchicago-perftest/cc32-16p32-16/test1185
Command: STOR /My Drive/migration/uchicago-perftest/cc32-16p32-16/test1185
Message: Fatal FTP response
---
Details: 451-GlobusError: v=1 c=TOO_BUSY\r\n451-GlobusError: v=1 c=INTERNAL_ERROR\r\n451-\r\n451-GD-Method: "PA
"https://www.googleapis.com/upload/drive/v3/files/11h9NBppR7qG7YaIZDqyWi-8w9rzW9gGj"\r\n451-GD-Response-Code: "
exceeded."\r\n451 End.\r\n
```





# More cool stuff...



Accessing protected data

Accessing cloud and archival storage

Managing access using Globus groups





# On sustainability...



Development is partly funded by...



U.S. DEPARTMENT OF  
**ENERGY**



THE UNIVERSITY OF  
**CHICAGO**



**NIST**  
National Institute of  
Standards and Technology  
U.S. Department of Commerce

Argonne   
NATIONAL LABORATORY

# ...Operations are funded by subscribers





# Our sustainability model

- **File transfer for non-profit research is free to all**
  - Subscription required if transferring to/from a commercial entity
- **Subscriptions enable multiple enhanced features**
  - Researchers: Data sharing, transfer/sharing to/from personal endpoints, HTTPS access, Globus Flows for automation
  - Sysadmins: Management console, usage reporting
  - Developers: Globus Search, app integration support
  - Additional security/logging for protected data management
  - Priority support
- **Subscriptions are required for access to connectors**

# Resources

- Documentation: [docs.globus.org](https://docs.globus.org)
- Community discussion: [globus.org/mailing-lists](https://globus.org/mailing-lists)
- Support: [support@globus.org](mailto:support@globus.org)