Globus is ... a non-profit service developed and operated by THE UNIVERSITY OF CHICAGO
Our Mission

Increase the efficiency and effectiveness of researchers engaged in data-driven science and scholarship through **sustainable** software.
Operations are funded by subscribers
300,000 REGISTERED USERS
1,400+ IDENTITY PROVIDERS
39,000 ACTIVE ENDPOINTS
1,600+ CONNECTED INSTITUTIONS
10,000+ ACTIVE SHARED ENDPOINTS
RELIABLE TRANSFER 1PB PER DAY

Numbers reflect the 12-month period ended 3/31/2022
Transfer and share data

1,654,555,083,356 MB transferred

A ➔ fast secure transfer reliable ➔ B
Globus delivers...

Fast and reliable big data transfer, sharing, and platform services...

...directly from your own storage systems...

...via software-as-a-service using existing identities with the overarching goal of...
...unifying access to data across tiers
Globus SaaS / PaaS: Research data lifecycle

1. Researcher initiates transfer request; or requested automatically by script, science gateway.

2. Globus transfers files reliably, securely.

3. Researcher selects files to share, selects user or group, and sets access permissions.

4. Globus controls access to shared files on existing storage; no need to move files to cloud storage!

5. Collaborator logs in to Globus and accesses shared files; no local account required; download via Globus.

6. The Timer and Automation Services, the Command Line Interface, API sets, and Python SDK provide the tools...

7. ... for building science gateways, portals, publication services...

8. ...and automating research workflows - ensuring those that need access to the data have it.

- Use a Web browser or platform services
- Access any storage
- Use an existing identity
Hybrid SaaS – Transfer – Mapped Collections

Customer owned and administered storage system with Globus Connect running on it

No data relay or staging via Globus, files move directly between storage locations

User identity mapped to local account

Globus service orchestrates file movement via communication with Globus Connect

Globus Security Domain

Subscriber Security Domain

DATA CHANNEL

CONTROL CHANNEL
Hybrid SaaS – Sharing – Guest Collections

Globus User Control Domain

Subscriber Control Domain

Globus Control Domain

GUEST COLLECTIONS
User managed "overlay" permissions stored on Globus Service

MANAGED ENDPOINT
Subscriber managed file system and endpoint policies

DATA CHANNEL

CONTROL CHANNEL
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
  - Fine grained access control on the collections

- **Data remain at institutions, not stored by Globus**

- **Data does not flow through the Globus Service but directly between Endpoints and their Collections**

- **Integrity checks of transferred data**

- **High availability and redundancy**

- **Encryption of user files and Globus control data**
Collections and Globus Connect

- **Globus Connect Server**
  - Multi user Linux Systems
  - [https://docs.globus.org/globus-connect-server/](https://docs.globus.org/globus-connect-server/)

- **Globus Connect Personal**
  - Personal Workstations and Laptops
  - [https://www.globus.org/globus-connect-personal](https://www.globus.org/globus-connect-personal)
  - OS specific instructions
    - [https://docs.globus.org/how-to/](https://docs.globus.org/how-to/)
Demo time!

- Identities and Accounts
- Transfer Details
- The Activity Monitor
- The Console
- Sharing
- Transfer
- Roles
- Bookmarks
- Groups
- Responsive Interface
- The Hamburger Menu
Globus Sharing at the ALCF

• Eagle

• Sharing on Eagle Using Globus

• Just Google “ALCF Globus Eagle”
Manage Protected Data

Higher assurance levels for HIPAA and other regulated data

- Support for protected data such as health related information
- Share data with collaborators while meeting compliance requirements
- Includes BAA option
Globus for high assurance data management

• **Restricted data handling**
  – PII ( Personally identifiable information )
  – CUI ( Controlled Unclassified Information )
  – PHI ( Protected Health Information )

• **University of Chicago security controls**
  – NIST SP 800-53
  – Superset of NIST SP 800-171

• **Business Associate Agreements (BAA) will be between University of Chicago and our subscribers**
  – University of Chicago has a BAA with Amazon
High Assurance features

• **Additional authentication assurance**
  – Per storage gateway policy on frequency of authentication with specific identity for access to data (timeout)
  – Ensure that user authenticates with the specific identity that gives them access within session (decoupling linked identities)

• **Session/device isolation**
  – Authentication context is per application, per session (~browser session)

• **Enforces encryption of all user data in transit**

• **Audit logging**
One service, many interfaces

Globus service

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

...
Globus Automation Capabilities

**Timer Service**
Scheduled and recurring transfers (a.k.a. Globus cron)

**Command Line Interface**
Ad hoc scripting and integration

**Globus Flows service**
Comprehensive task (data and compute) orchestration with human in the loop interactions
Custom portals? Science Gateways? Unique workflows? Our open REST APIs and Python SDK empower you to create an integrated ecosystem of research data services and applications.
Globus APIs

- Auth
- Groups
- **Transfer**
- Search
- Timer
- Flows
- GCS Manager

- Globus Web App consumes public Transfer API
- Resource named by URL (standard REST approach)
- Globus APIs use JSON for documents

[docs.globus.org/api/transfer]
Globus Python SDK

- Python client library for the Globus REST APIs
- Largely direct mapping to REST API
- `globus_sdk.TransferClient` class handles connection management, security, framing, marshaling

[Links]
globus.github.io/globus-sdk-python
Support Resources

• Globus Documentation: docs.globus.org
• Globus Timer: globus.org/blog/scheduled-and-recurring-transfers-now-available-globus-web-app
• Globus CLI: docs.globus.org/cli/
• Globus Automation Services and Flows: docs.globus.org/globus-automation-services/
• YouTube Channel: youtube.com/user/GlobusOnline
Developer References

• Globus API / SDK Documentation
  – Transfer API: docs.globus.org/api/transfer/

• Globus GitHub: github.com/globus/
  – Jupyter Notebooks
    o Stand alone notebooks and hub integrations that walk through much of the functionality of our SDK
    o https://github.com/globus/globus-jupyter-notebooks
  – Automation Examples
    o Shell scripted CLI and Python module examples of common research data management use cases
    o https://github.com/globus/automation-examples
Globus in a nutshell

ATPESC
August 5, 2022

Greg Nawrocki
greg@globus.org
nawrocki@uchicago.edu
nawrocki@anl.gov