Globus in a nutshell

ATPESC August 5, 2022

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









Globus is ...

a non-profit service developed and operated by





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



Development is funded by...



















Operations are funded by subscribers









































CGIAR







































































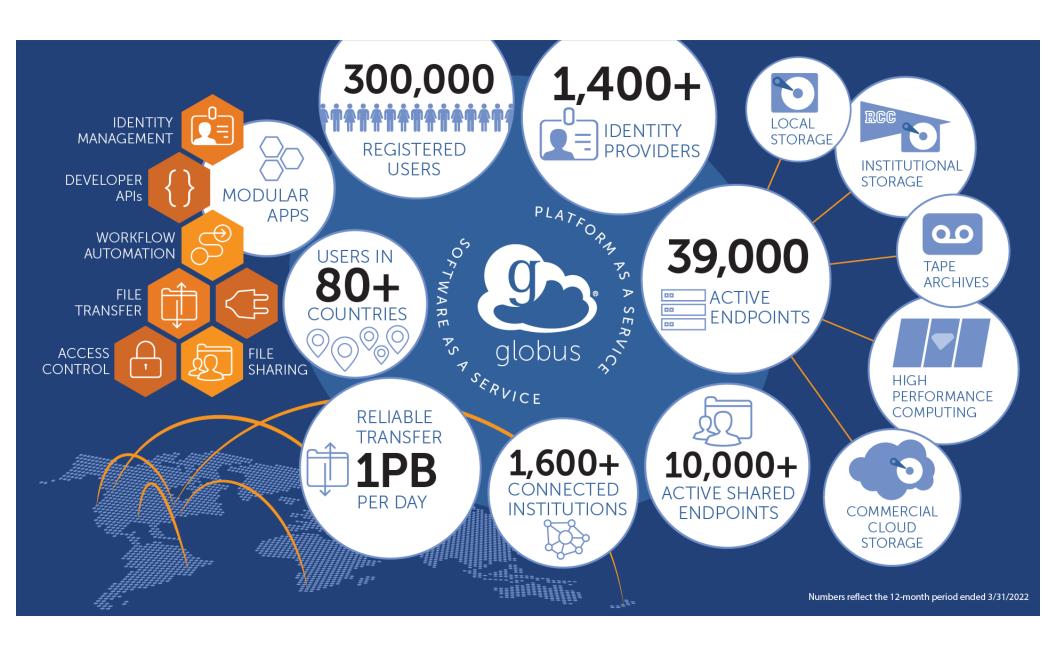
THE ROCKEFELLER UNIVERSITY



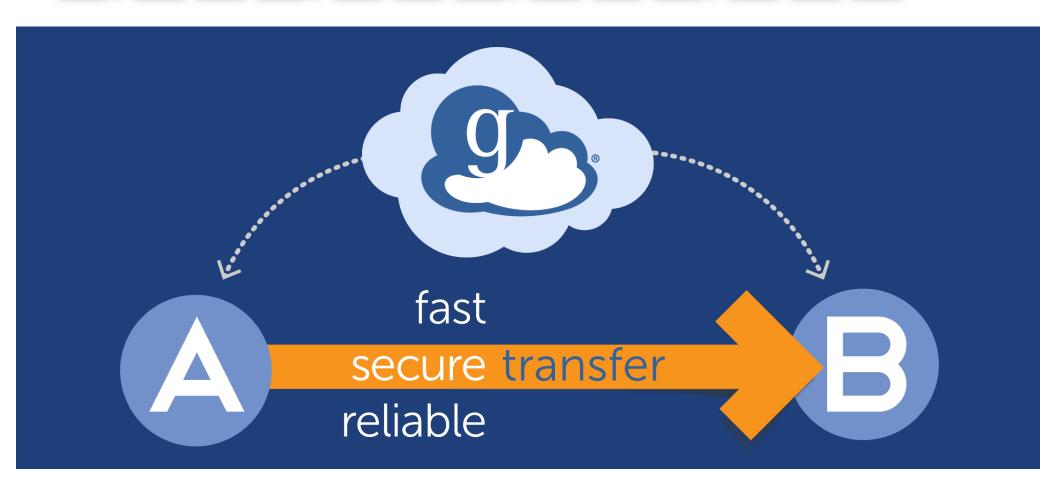








1,654,555,083,356MB





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





PREMIUM globus connectors

















Microsoft OneDrive







IBM Spectrum Scale











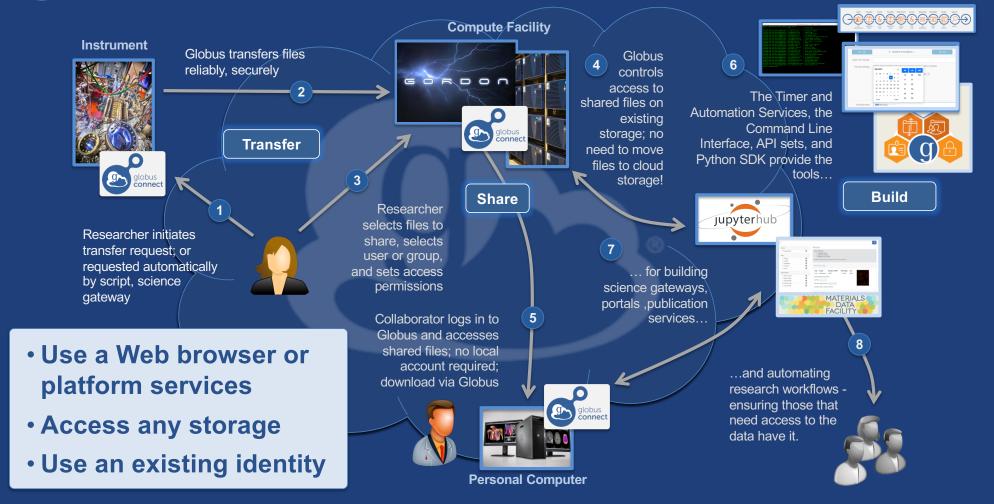






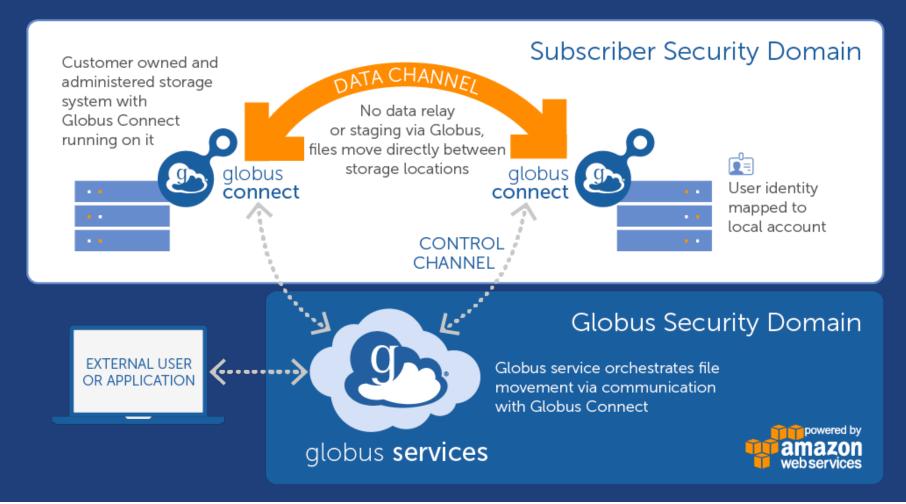


Globus SaaS / PaaS: Research data lifecycle



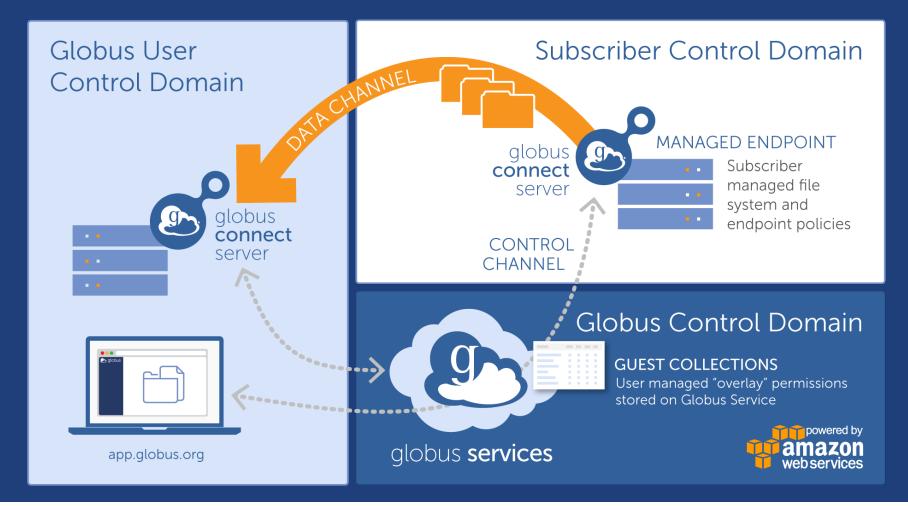


Hybrid SaaS – Transfer – Mapped Collections





Hybrid SaaS – Sharing – Guest Collections





Globus core security features



- 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/

Globus Connect Personal

- Personal Workstations and Laptops
- https://www.globus.org/globus-connect-personal
- OS specific instructions
 - o https://docs.globus.org/how-to/











Demo time!

Sharing

Responsive Interface

Identities and Accounts

Transfer

Bookmarks

Groups

Roles

The Activity Monitor

The Hamburger Menu

The Console

Transfer Details



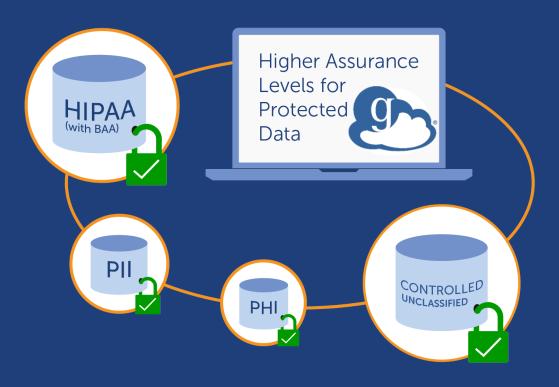
Globus Sharing at the ALCF

- Eagle
- Sharing on Eagle Using Globus
- https://www.alcf.anl.gov/support-center/theta-andthetagpu/eagle-datasharing#Accessing%20your%20Eagle%20Project%20 **Directory**
- 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

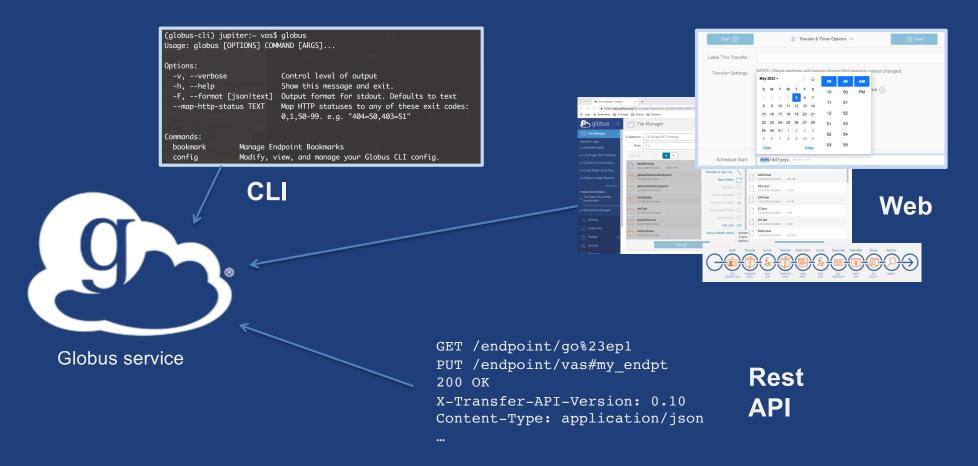
- Per storage gateway policy on frequency of authentication with specific identity for access to data (timeout)
- Ensure that user authenticates with the specific identity that gives them access within session (decoupling linked identities)

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 Automation Capabilities



Command Line Interface

Ad hoc scripting and integration



Timer Service

Scheduled and recurring transfers (a.k.a. Globus cron)



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

globus-sdk-python.readthedocs.io/en/stable/ globus.github.io/globus-sdk-python



Support Resources

- Globus Documentation: docs.globus.org
- Globus Timer: globus.org/blog/scheduled-andrecurring-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/
 - SDK: globus-sdk-python.readthedocs.io/en/stable/
- Globus GitHub: github.com/globus/
 - Jupyter Notebooks
 - o Stand alone notebooks and hub integrations that walk through much of the functionality of our SDK
 - https://github.com/globus/globus-jupyter-notebooks
 - Automation Examples
 - o Shell scripted CLI and Python module examples of common research data management use cases
 - 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





