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.
…or “We enable researchers to outsource the mundane—yet critical—data management tasks that impede their work.”
Development is partly funded by...
Operations are funded by subscribers
Globus and the 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 Globus Command Line Interface, API sets, Python SDK and Action Providers give you a platform…**

7. … for building science gateways, portals and automations.

8. **Streamlining research workflows and ensuring those that need access to the data have it.**

---

- Use a Web browser, CLI or platform services
- Access any storage
- Use an existing identity
Globus is SaaS and PaaS

Web app simplifies *ad hoc* data management

Platform services simplify creation of automated data management apps for diverse instrument types
Conceptual Architecture: Hybrid SaaS

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.

- Control channel: External user or application communicates with Globus services for file movement via communication with Globus Connect.

- Subscriber Security Domain: User identity mapped to local account.

- Globus Security Domain: Globus service orchestrates file movement via communication with Globus Connect.
Globus Connect Server Architecture

- MAPPED
- MAPPED
- GUEST
- MAPPED
- GUEST

STORAGE GATEWAY
iRODS

STORAGE GATEWAY
POSIX 2
High Assurance

STORAGE GATEWAY
Amazon Web
Services S3

globus
cnect server
GCS MANAGER
GridFTP SERVER
WEB SERVER
OIDC SERVER

DATA TRANSFER
NODE 1

encrypted with admin
generated key

DATA TRANSFER
NODE 2

back-up and sync
Globus supports a broad set of storage systems
Globus is fast … and, more importantly, reliable.

Activity List

RDA to ALCF noverify
transfer completed

Overview

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@globus.org)
Condition: SUCCEEDED
Requested: 2021-04-23 02:50 pm
Completed: 2021-04-23 02:53 pm
Duration: 2 minutes 47 seconds
Transfer Settings: 
- transfer is not encrypted
- overwriting all files on destination

72.8 Gbps

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
Globus is secure

- **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, not stored by Globus
- Integrity checks of transferred data
- High availability and redundancy
- Encryption of user files and Globus control data
Managing protected data with Globus

Security controls
→ NIST 800-53
→ 800-171 Low+

Restricted data handling
→ PHI, PII, CUI
→ Compliant data sharing

BAA w/Uchicago
→ UChicago BAA with Amazon
A quick peek...
Federated login (via Globus Auth: OAuth2/OIDC)

1,400+ IdPs
...via released R&S attributes or custom configuration
Transfer options and Globus “cron”
Sharing: Globus maintains a virtual ACL

<table>
<thead>
<tr>
<th>USER OR GROUP</th>
<th>CREATED</th>
<th>READ</th>
<th>WRITE</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="mailto:3ac0357c-025a-4580-914e-3328627e9691@clients.auth.globus.org">3ac0357c-025a-4580-914e-3328627e9691@clients.auth.globus.org</a></td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Vas Vasiliadis (<a href="mailto:vas@globusid.org">vas@globusid.org</a>)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Demo Documentation (<a href="mailto:demodoc@globusid.org">demodoc@globusid.org</a>)</td>
<td>1/6/2021, 05:39 PM</td>
<td>-</td>
<td>✔</td>
</tr>
</tbody>
</table>

**Share with users, groups and apps/services**
One service, many interfaces

Globus service

CLI

Web

Flows

Rest API

GET /endpoint/go%23ep1
PUT /endpoint/vas#my_endpt
200 OK
X-Transfer-API-Version: 0.10
Content-Type: application/json
...
What’s driving the development agenda?

Compliance with diverse data privacy and other legal protections
“I love scp/sftp/...; it works great in my perl scripts!”

If you're doing something the same way you have been doing it for ten years, the chances are you are doing it wrong.

—— Charles Kettering ——
Instrument innovations, research processes are evolving at breakneck speed.

How can we adapt our data management to keep pace?
Current Focus: Automating data management at scale
3 Degrees of Automation

**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
Understanding SARS-CoV-2 Protein Structure

"These data services have taken the time to solve a structure from weeks to days and now to hours"
Darren Sherrell, SBC beamline scientist APS Sector 19

Argonne researchers use Theta for real-time analysis of COVID-19 proteins

Argonne's User Facilities Continue to Enable Critical Work Combating and Addressing the Impacts of the COVID-19 Epidemic
June 12, 2020
Portals for data publication and faceted search

ALCF Community Data Co-Op

APS 2D XRF
X-Ray fluorescence imaging for a variety of applications

PyTorch BraggNN for CPU
name: aristana_uchicago/BraggNNPT_CPU

Authors
Nikil Ravj; Pranesh Chaturvedi; E.A. Huerta; Zhichen Liu; Aristana Scourtas; KJ Schmidt; Ryan Chard; Ben Blaiszik

Visibility Version Shim
public 0.10.0 python:PythonStaticMethodServable

PyTorch BraggNN for CPU
name: aristana_uchicago/BraggNNPT_CPU

Authors
Nikil Ravj; Pranesh Chaturvedi; E.A. Huerta; Zhichen Liu; Aristana Scourtas; KJ Schmidt; Ryan Chard; Ben Blaiszik

Visibility Version Shim
VSC iRODS and Globus use cases
iRODS and Globus?

But:

• A lot of windows users: WSL/iCommands? PRC?...
• Reliability: network issues, server problems .....  
• Repetitive transfers: scripting for everyone?
• Sharing with externals: create externals accounts, itickets?
iRODS and Globus!

- **iRODS**
  - Store
  - Describe
  - Discover
  - Automate

- **Globus**
  - Move
  - Share

WHAT IF I TOLD YOU
THEY ARE COMPLEMENTARY
VSC use cases

- Transfer
- Ingest
- Share
Transfer
Ingest (PoC): milk quality control at dairy farms

Livestock Technology Research Group
Dept Animal & Human Health Engineering
www.livestocktechnology.be
An upload directory per Farm is created in iRODS

The iRODS directory is shared with the GlobusID user with write permissions
Configuration: farm (1/2)

Local directory - farm data updates

iRODS dir shared with GlobusID Farm1
Configuration: farm (2/2)

Select sync option

Transfer is forced to be encrypted (defined by the sharing config)

A sync process is started every 6 hours

Schedule Start: 07/02/2022 06:00 AM

Repeat: 6 hours
Configuration: department

Transfer is forced to be encrypted

Select sync option

A sync process is started 30m after farm dta sync to keep nearly real time processing of the data
Monitoring

List of scheduled timers

Number of executions done
Share (soon PoC)

Authentication institutional account

VSC Tier-1 Cloud

Qupath

KU LEUVEN
Questions?