Globus: Making It Easy to Move, Share and Publish Your Data

Vas Vasiliadis
vas@uchicago.edu
Thank you to our sponsors!

U.S. DEPARTMENT OF ENERGY

THE UNIVERSITY OF CHICAGO

NATIONAL INSTITUTES OF HEALTH

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

Argonne
NATIONAL LABORATORY

powered by
amazon web services
<table>
<thead>
<tr>
<th>Major Services</th>
<th>190 PB Transferred</th>
<th>30 Bn Files Processed</th>
<th>45,000 Registered Users</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13 National Labs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Use Globus</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10,000 Registered Users</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Active Endpoints</td>
<td>10,000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10,000 Active Users/Year</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>99.9% Uptime</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>60+ Institutional Subscribers</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 PB Largest Single Transfer to Date</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 Months Longest Continuously Managed Transfer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>130 Federated Campus Identities</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Agenda

• Overview for those new to Globus
• Survey of “recent” enhancements
• Brief demonstration
• Questions
Globus delivers...
Big data transfer, sharing, publication, and discovery...
...directly from your own storage systems...
...via software-as-a-service
Data management today

• Move?
• Share?
• Discover?
• Reproduce?
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. Researchers assembles data set; describes it using metadata (Datacite & domain-specific).

7. Curator reviews and approves; data set published on campus or other system.

8. Peers, collaborators search and discover datasets; transfer and share using Globus.

- Only a Web browser required
- Use storage system of your choice
- Access using your campus credentials
Another perspective…

Campus Bridging

…within and beyond campus boundaries
Bridge to national cyberinfrastructure

Move datasets to supercomputer, national facility

Move results to campus (…)

Edison
Bridge to instruments

- Clear staging store
- Data cleansing
- Pre-processing
- ...

Amazon Glacier
High durability, low cost store

Raw Source Data

Analysis store
Bridge to your collaborators

- Group management
- Share creation
- Permissions configuration

Local/Campus stores

Cloud stores

• EC2
• XSEDE
• Jetstream
• NERSC
Bridge to community/public

Multi-step curation workflow
- Review → Update → Resubmit cycle

(Re)format...
- PDF/A
- HDF
- ...

Describe...
- Datacite metadata
- Domain metadata
- Provenance info
- ...

Persistent identifier
- DOI

Persistent identifier

Replicated, public repositories

(Re)format...
Why use Globus?

- Ease of use: consistent UI across systems
- “Fire-and-forget” file transfer
- Maximized WAN throughput
- Secure access w/multi-tier security model
- No/low-overhead external collaboration
- Rapid deployment w/standard packages
- Highly automatable: CLI, RESTful API
What’s new?

• New authentication service
• Endpoint search and bookmarks
• Management console
• Data publication and discovery
• Storage connectors
• Globus Platform-as-a-Service
Demonstration
Globus Connect…
Makes your storage system a Globus endpoint
Globus Connect Personal

- Installers do not require admin access
- Zero configuration; auto updating
- Handles NATs
Globus Connect Server

- Enable access for all users with local accounts
- Create endpoint on practically any filesystem
- Native packages: RPMs and DEBs
Storage connectors

• **Standard storage connectors (Posix)**
  – Linux, Windows, MacOS
  – Lustre, GPFS, OrangeFS, etc.

• **Premium storage connectors**
  – HPSS
  – HDFS
  – S3
  – Ceph RadosGW (S3 API)
  – Spectra Logic BlackPearl
  – Google Drive (coming soon)
• Move portal storage into Science DMZ, with Globus endpoint
• Leave Portal Web server behind firewall
• Globus handles the security and data heavy lifting
Sample python data portal

Modern Research Data Portal
It's how research data management is done!

Globus Transfer API
API reference for transfer and sharing functions.

Globus Auth API
API reference for authentication and authorization.

Frequently Asked Questions
When all else fails...

Coming Soon

- **HTTPS support**
  - Synchronous alternative GridFTP
  - Implements same fine-grained access control model
  - Greatly simplifies sharing/transfer for “small” datasets
  - Supports standard browser behaviors such as viewing known file types in the browser
  - Allows integration with clients and web apps to further leverage existing research storage systems

- **New command line interface**
  - Locally installed Python client
  - Similar to AWS command line tools
Globus Subscriptions

• **Standard Subscription**
  – Shared endpoints
  – Data publication
  – HTTPS support*
  – Management console
  – Usage reporting
  – Priority support
  – Application integration

• **Branded Web Site**

• **Premium Storage Connectors**
  – Amazon S3, Ceph, HPSS, Spectra, Google Drive*, …

• **Alternate Identity Provider (InCommon is standard)**

*Available in early 2017
Globus Subscriptions

Globus file transfer is free to use for non-profit research and educational purposes. Subscriptions offer researchers access to enhanced features such as file sharing and data publication, and provide system administrators with a comprehensive view into how storage systems are being utilized. Your subscription helps us make Globus self-sustaining so we can continue to offer high-quality data management infrastructure to all researchers.

Click here to start your 90-day free trial of Globus premium features

Subscriptions for Non-Profit Research and Education

If you wish to use Globus in a commercial setting, you must have a commercial subscription.

<table>
<thead>
<tr>
<th>Features</th>
<th>Basic - Free</th>
<th>Starter Subscription</th>
<th>Standard Subscription</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transfer level</td>
<td>Unlimited</td>
<td>Unlimited</td>
<td>Unlimited</td>
</tr>
<tr>
<td>User level</td>
<td>Unlimited</td>
<td>Unlimited</td>
<td>Unlimited</td>
</tr>
<tr>
<td>Managed endpoints</td>
<td>None</td>
<td>1</td>
<td>Unlimited</td>
</tr>
<tr>
<td>Management console</td>
<td>—</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Usage reports</td>
<td>—</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Support for Globus Connect, Web, CLI</td>
<td>—</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Shared endpoints</td>
<td>—</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Globus Plus users</td>
<td>—</td>
<td>—</td>
<td>✓</td>
</tr>
</tbody>
</table>
Representative subscribers
Learn more on the GlobusWorld Tour!

Welcome to the GlobusWorld Tour!

We're presenting a series of Globus tutorials and developer workshops across the US, building on the success of the workshop held at GlobusWorld 2016. These workshops are made possible by the various hosting institutions that generously provide meeting space and other financial support.

The following workshops are currently scheduled:

- **September 13-14, 2016** - LBNL, Berkeley, CA
- **October 12-13, 2016** - Yale University, New Haven, CT
- **October 25-26, 2016** - NCAR, Boulder, CO
- **Dates TBD** - NIH, Washington D.C.

**Why Attend?**

- Learn how the Globus platform simplifies development of web applications for researchers
- Experiment with new Globus services and APIs
- Exchange ideas with peers on ways to apply Globus technologies
- Expand your knowledge of Globus administration features

- [www.globusworld.org/tour/](http://www.globusworld.org/tour/)
- **Webinar: October 6th, 1pm CDT**
- **SC16: Full-day developer tutorial (Booth #436)**
Join the Globus community

• Access the service: globus.org/login
• Create a personal endpoint: globus.org/app/endpoints/create-gcp
• Create a server endpoint: docs.globus.org/resource-provider-guide
• Need help? support@globus.org
• Subscribe: globus.org/subscriptions
• Follow us: @globusonline