What is Globus?

Rick Wagner
rick@globus.org

LUG18 – April 26, 2018
Research data management today

How do we...
...move?
...share?
...discover?
...reproduce?

Index?
Globus delivers…
Secure, reliable, data transfer, sharing, publication, and discovery…
…directly from your own storage systems…
…via software-as-a-service
Globus enables...

Campus Bridging

...within and beyond campus boundaries
Bridge to campus HPC

Move datasets to campus research computing center

Move results to laptop, department, lab, ...
Bridge to national cyberinfrastructure

Move datasets to supercomputer, national facility

Move results to campus (….)
Bridge to instruments

Pre-processed Data

High durability, low cost store

Amazon Glacier

Analysis store

Raw Source Data

Raw Data
Bridge to collaborators

External Campus Storage

Public/Private Cloud stores

EC2

XSEDE
Extreme Science and Engineering Discovery Environment

Jetstream

NERSC

NC STATE UNIVERSITY
Bridge to community/public

Project Repositories, Replication Stores

Public Repositories
Globus SaaS: 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. Researcher assembles data set; describes it using metadata (Dublin core and 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

- Use a Web browser
- Access any storage
- Use an existing identity
Conceptual architecture: Hybrid SaaS

**Subscriber Control Domain**
- Subscriber owned and administered storage system
- Globus "client" software

**Source Endpoint**

**Destination Endpoint**

**DATA Channel**
- No data relay or staging via Globus

**CONTROL Channel**
- Single, globally accessible multi-tenant service
Conceptual architecture: Sharing

External User Control Domain

Global Control Domain

DATA Channel

CONTROL Channel

Managed Endpoint

Administrator managed filesystem permissions

Shared Endpoint

User managed "overlay" permissions

Subscriber Control Domain
Why use Globus?

- **Simplicity**
  - Consistent UI across systems
  - Easy access to collaborators

- **Reliability and performance**
  - “Fire-and-forget” file transfer
  - Maximized WAN throughput

- **Operational efficiency**
  - Low overhead SaaS model
  - Highly automatable: CLI, RESTful API

- **Access to a large and growing community**
How can I use Globus on my computer?
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

• Makes your storage accessible via Globus
• Multi-user server, installed and managed by sysadmin
• Default access for all local accounts
• Native packaging Linux: DEB, RPM

docs.globus.org/globus-connect-server-installation-guide/
Globus Connect Server

Non-POSIX Connectors
- Google Drive
- WD
- HPSS
- Spectra
- Amazon S3
- Hadoop HDFS
- Ceph

POSIX-compliant Connector

Local system users

Globus Connect Server
- MyProxy CA
- OAuth Server
- GridFTP Server

Local Storage System
- HPC cluster
- NAS

DTN
How can I integrate Globus into my research workflows?
Globus serves as...

...a platform for building science gateways, portals, and other web applications in support of research and education.
Use(r)-appropriate interfaces

Globus service

Web

CLI

Rest API

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

...(more content)
Globus as PaaS

Integrate file transfer and sharing capabilities into scientific web apps, portals, gateways, etc.

Use existing institutional ID systems in external web applications
Globus PaaS developer resources

Python SDK

Sample Application

Jupyter Notebook

```python
from __future__ import print_

tutorial_endpoint_1 = "db59af0-6d04-11e5-ba64-22000b92c66c" # endpoint "Globus"
tutorial_endpoint_2 = "506a29c-63ac-11e4-8062-22000ab68755" # endpoint "Globus"
tutorial_users_group = "506a29c-63ac-11e4-8062-22000ab68755" # group "Tutori

Configuration

First you will need to configure the client with an OAuth2 access token. For the purpose of this tutorial, you can use a
```
```
```
Globus sustainability model

- **Standard Subscription**
  - Shared endpoints
  - Data publication
  - Management console
  - Usage reporting
  - Priority support
  - Application integration
  - HTTPS support (coming soon)

- **Branded Web Site**

- **Premium Storage Connectors**

- **Alternate Identity Provider** (InCommon is standard)
Thank you to our users…

<table>
<thead>
<tr>
<th>Most server endpoints at a single organization</th>
<th>48</th>
</tr>
</thead>
<tbody>
<tr>
<td>100TB+ users</td>
<td>500</td>
</tr>
<tr>
<td>Active users</td>
<td>14,000</td>
</tr>
<tr>
<td>Longest running managed transfer</td>
<td>3 months</td>
</tr>
<tr>
<td>Largest single transfer to date</td>
<td>1 PB</td>
</tr>
<tr>
<td>Active shared endpoints</td>
<td>5,000</td>
</tr>
<tr>
<td>Uptime</td>
<td>99.5%</td>
</tr>
<tr>
<td>Transferred</td>
<td>384 PB</td>
</tr>
<tr>
<td>Tasks processed</td>
<td>64 billion</td>
</tr>
<tr>
<td>Registered users</td>
<td>76,000</td>
</tr>
</tbody>
</table>
Our supporters

- NERSC
- NIH
- Johns Hopkins University
- NCAR
- Yale
- University of Michigan
- Wellcome Trust Sanger Institute
- UC
- Virginia Tech
- UF
- University of Florida
- Cornell University
- NCSA
- NASA JPL
- University of Chicago
- Indiana University
- Michigan State University
- Stanford University
- New York Genome Center
- Oak Ridge National Laboratory
- Dartmouth
- SIMONS FOUNDATION
- Los Alamos National Laboratory
- Argonne National Laboratory
Join the Globus community

• Access the service: globus.org/login
• Create a personal endpoint: globus.org/app/endpoints/create-gcp
• Documentation: docs.globus.org
• Engage: globus.org/mailing-lists
• Subscribe: globus.org/subscriptions
• Need help? support@globus.org
• Follow us: @globusonline