Globus: Recent Enhancements and Future Plans

Vas Vasiliadis
The University of Chicago

vas@uchicago.edu
Thank you to our sponsors!
...and Thank YOU!

<table>
<thead>
<tr>
<th><strong>5</strong></th>
<th><strong>190 PB</strong></th>
<th><strong>30 Bn</strong></th>
<th><strong>45,000</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>major services</td>
<td>transferred</td>
<td>files processed</td>
<td>registered users</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>13</strong></th>
<th><strong>10,000</strong></th>
<th><strong>10,000</strong></th>
<th><strong>99.9%</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>national labs use Globus</td>
<td>active endpoints</td>
<td>active users/year</td>
<td>uptime</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>60+</strong></th>
<th><strong>1 PB</strong></th>
<th><strong>3 months</strong></th>
<th><strong>130</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>institutional subscribers</td>
<td>largest single transfer to date</td>
<td>longest continuously managed transfer</td>
<td>federated campus identities</td>
</tr>
</tbody>
</table>
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. Peers, collaborators search and discover datasets; transfer and share using Globus.
7. Curator reviews and approves; data set published on campus or other system.
8. Publication Repository.

- Only a Web browser required
- Use storage system of your choice
- Access using your campus credentials
No Globus usernames required

- No longer require a Globus username/password
- Existing Globus usernames moved to separate, optional “Globus ID” IdP
- Any identity recognized by Globus is now sufficient to access Globus
- Globus Account = primary identity + [linked identities]
- Globus sharing and groups use any identity
Used by XSEDE and others

Jetstream Web App would like to:

☑ Access all Jetstream resources

By clicking “Allow”, you allow Jetstream Web App to have access to basic information about you, including your name and email address. You can rescind this consent at any time.

Allow Deny

Compute Canada has partnered with Globus to offer this high-performance file transfer service. Calcul Canada s’est associé à Globus pour vous offrir ce service de transfert de fichier à haute performance.

Log in to use Compute Canada Globus Web App

Use your existing organizational login

e.g. university, national lab, facility, project, Google or Globus ID

(Your Globus username and password used prior to February 13, 2018 is now Globus ID)

WestGrid

Continue

Didn’t find your organization? Then use Globus ID to sign up.
Adding your campus identity provider to Globus

• InCommon identity providers that release research & scholarship attributes to CILogon (free)

• OpenID Connect identity provider supported by Globus Auth (subscription)
Demonstration

- New authentication service
- Endpoint search and bookmarks
- Management console
- Data publication and discovery
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)
Filling the missing storage tier

On-line/High performance storage system

Science DMZ

Data Transfer Node

BlackPearl Gateway

ArcticBlue Nearline

Data

Archive/Tape System

Control

Data Transfer Node

BlackPearl Gateway

ArcticBlue Nearline

Data

Archive/Tape System

Control

Data Transfer Node

BlackPearl Gateway

ArcticBlue Nearline

Data

Archive/Tape System

Control

Data Transfer Node

BlackPearl Gateway

ArcticBlue Nearline

Data

Archive/Tape System

Control

Data Transfer Node

BlackPearl Gateway

ArcticBlue Nearline

Data

Archive/Tape System

Control

Data Transfer Node

BlackPearl Gateway

ArcticBlue Nearline

Data

Archive/Tape System

Control

Data Transfer Node

BlackPearl Gateway

ArcticBlue Nearline

Data

Archive/Tape System

Control

Data Transfer Node

BlackPearl Gateway

ArcticBlue Nearline

Data

Archive/Tape System

Control

Data Transfer Node

BlackPearl Gateway

ArcticBlue Nearline

Data

Archive/Tape System

Control
New Globus CLI (coming soon)

$ globus transfer

Usage: globus transfer [OPTIONS] COMMAND [ARGS]...

Interact with Globus Transfer API. Transfer, Delete, List, and Rename files on Endpoints, manage your Endpoints and Shares, and monitor your ongoing Transfer Tasks

Options:
-F, --format [json|text]  Output format for stdout. Defaults to text
-h, --help  Show this message and exit.
--version  Show the version and exit.

Commands:
  acl  Manage Endpoint Access Control Lists
  async-delete  Submit a Delete Task
  async-transfer  Submit a Transfer Task
  bookmark  Manage Endpoint Bookmarks
  endpoint  Manage Globus Endpoint definitions
  ls  List Endpoint directory contents
  mkdir  Make a directory on an Endpoint
  rename  Rename a file or directory on an Endpoint
  task  Manage asynchronous Tasks
Globus Platform-as-a-Service

Globus APIs
- Data Publication & Discovery
- File Sharing
- File Transfer & Replication
- Auth & Groups

Globus Connect

Globus Toolkit
Motivating Questions

• How do you leverage Globus services in your own applications?
• How do you extend Globus with your own services?
• How do we empower the research community to create an integrated ecosystem of services and applications?
Example: Modern Research Data Portal

- **Browser**
  - Desktop
  - User’s Endpoint (optional)
- **Portal Web Server (Client)**
- **Portal Endpoint**
  - **Science DMZ**
  - **Firewall**
  - **Globus Cloud**
  - **Globus Auth**
  - **Globus Web Helper Pages**
  - **Globus Transfer**
  - **Other Services**
  - **Other Endpoints**

- **HTTPS**
- **REST**
- **GridFTP**

- **Identity Provider**
Example: Modern Research Data Portal

- Move portal storage into Science DMZ, with Globus endpoint
- Leave Portal Web server behind firewall
- Globus handles the security and data heavy lifting
Modern Research Portals Leverage the Science DMZ

Portal server applications:
- web server
- search
- database
- authentication

fasterdata.es.net/science-dmz
# Research data portal

## NCEP Climate Forecast System Version 2 (CFSv2) Monthly Products

**ds094.2**

For assistance, contact Bob Dattore (303-497-1825).

**Data Description**

<table>
<thead>
<tr>
<th>Data Description</th>
<th>Data File Downloads</th>
<th>Customizable Data Requests</th>
<th>Other Access Methods</th>
<th>NCAR-Only Access</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Web Server Holdings</td>
<td>Globus Transfer Service (GridFTP)</td>
<td>Subsetting</td>
<td>THREDDS Data Server</td>
</tr>
<tr>
<td>Union of Available Products</td>
<td>Web File Listing</td>
<td>Request Globus Invitation</td>
<td>Get a Subset</td>
<td>TDS Access</td>
</tr>
</tbody>
</table>

**Products**

- Diurnal monthly means
- Regular monthly means
- Selected Parameter/Level Time Series

**Central File System (GLADE) Holdings**

- GLADE File Listing
- HPSS File Listing

**Tape Archive (HPSS) Holdings**

- GLADE File Listing
- HPSS File Listing

Mouse over the table headings for detailed descriptions.
Globus Transfer API

Nearly all Globus Web App functionality implemented via public Transfer API

/docs.globus.org/api/transfer
Challenge

• **How to provide:**
  – Login to apps
    o Web, mobile, desktop, command line
  – Protect all REST API communications
    o App → Globus service (e.g., Transfer API)
    o App → non-Globus service
    o Service → service

• **While:**
  – Providing seamless integration with apps and services
  – Not introducing even more identities
  – Providing least privileges security model
  – Being agnostic to programming language and framework
  – Being web friendly
  – Making it easy for users and developers
Globus Auth

- **Foundational identity and access management (IAM) platform service**
- **Simplify creation and integration of advanced apps and services**
- **Brokers authentication and authorization interactions between:**
  - end-users
  - identity providers: InCommon, XSEDE, Google, portals
  - services: resource servers with REST APIs
  - apps: web, mobile, desktop, command line clients
  - services acting as clients to other services
- **Based on OAuth2 and OpenID Connect**
- **IAM platform for XSEDE (e.g., XUP, Jetstream, Jira)**
Learn more tomorrow...

“Globus Auth”

Tuesday, September 27, 8:00am
Room: Bayfront B
HTTPS Access to Endpoints (coming soon)

• **Enhanced use of research storage:**
  – asynchronous, bulk transfer via GridFTP
  – synchronous remote access via HTTPS

• **Enhanced Globus web app**
  – Browser-based upload/download
  – Inline file viewer

• **Integration with clients and web apps**
Globus SDK for Python (Beta)

This SDK provides a convenient Pythonic interface to Globus REST APIs, including the Transfer API and the Globus Auth API. Documentation for the REST APIs is available at https://docs.globus.org.

Two interfaces are provided - a low level interface, supporting only GET, PUT, POST, and DELETE operations, and a high level interface providing helper methods for common API resources.

Source code is available at https://github.com/globus/globus-sdk-python.

Installation

The Globus SDK requires Python 2.6+ or 3.2+. If a supported version of Python is not already installed on your system, see this Python installation guide.

The simplest way to install the Globus SDK is using the pip package manager (https://pypi.python.org/pypi/pip), which is included in most Python installations:

```
pip install globus-sdk
```

This will install the Globus SDK and its dependencies.

Bleeding edge versions of the Globus SDK can be installed by checking out the git repository and installing it manually:

```
git checkout https://github.com/globus/globus-sdk-python.git
cd globus-sdk-python
python setup.py install
```
Sample data portal (Python)

Modern Research Data Portal

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...

github.com/globus/globus-sample-data-portal
Demonstration
HTTPS access
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
Harness the power of the Globus research data management cloud.

Transfer API
The Transfer API provides a REST-style interface to the Globus reliable file transfer service. The API can be used to monitor the progress of file transfers, manage file transfer endpoints, list remote directories, and submit new transfer and delete tasks.
LEARN MORE

Resource Providers
Globus allows you, as a resource provider, to easily offer reliable, secure, high-performance research data management capabilities to your users and their collaborators, directly from your own storage infrastructure.
LEARN MORE

Toolkit
The open source Globus Toolkit is a fundamental enabling technology for the “Grid,” allowing users to access high-performance computing resources securely across corporate, institutional, and geographic boundaries without sacrificing local autonomy.
LEARN MORE
New subscription pricing model

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 (click (\textsuperscript{\text{1}}) for description)</th>
<th>Basic - Free</th>
<th>Starter Subscription</th>
<th>Standard Subscription</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transfer level (\textsuperscript{\text{1}})</td>
<td>Unlimited</td>
<td>Unlimited</td>
<td>Unlimited</td>
</tr>
<tr>
<td>User level (\textsuperscript{\text{1}})</td>
<td>Unlimited</td>
<td>Unlimited</td>
<td>Unlimited</td>
</tr>
<tr>
<td>Managed endpoints (\textsuperscript{\text{1}})</td>
<td>None</td>
<td>1</td>
<td>Unlimited</td>
</tr>
<tr>
<td>Management console (\textsuperscript{\text{1}})</td>
<td>—</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Usage reports (\textsuperscript{\text{1}})</td>
<td>—</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Support for Globus Connect, Web, CLI (\textsuperscript{\text{1}})</td>
<td>—</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Shared endpoints (\textsuperscript{\text{1}})</td>
<td>—</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>Globus Plus users (\textsuperscript{\text{1}})</td>
<td>—</td>
<td>—</td>
<td>✔</td>
</tr>
</tbody>
</table>

globus.org/subscriptions
Summary

• Globus SaaS continues to improve its capabilities for researchers and research computing centers

• Globus PaaS opens new opportunities for developers creating web and mobile applications for researchers