Globus Introduction

Steve Tuecke
tuecke@globus.org
Thank you to our sponsors!
**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. 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.

**Additional Information**

- Only a Web browser required
- Use storage system of your choice
- Access using your campus credentials
Benefits of SaaS

• Users do not need to deploy software
  – Easy to access via Web browser
  – Command line, REST interfaces for flexible automation and integration

• New features automatically available

• Reduced IT operational costs
  – Small local footprint (Globus Connect)
  – Consolidated support and troubleshooting
<table>
<thead>
<tr>
<th><strong>Globus by the numbers</strong></th>
<th><strong>4 major services</strong></th>
<th><strong>180 PB transferred</strong></th>
<th><strong>20 billion files processed</strong></th>
<th><strong>31,000 registered users</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>13 national labs</strong></td>
<td><strong>3,000 active endpoints per month</strong></td>
<td><strong>~450 active daily users</strong></td>
<td><strong>99.9% uptime</strong></td>
<td><strong>35+ institutional subscribers</strong></td>
</tr>
<tr>
<td><strong>1 PB largest single transfer to date</strong></td>
<td><strong>3 months longest continuously managed transfer</strong></td>
<td><strong>130 federated campus identities</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Demo

- Login using existing identity
- Identity linking
- Endpoint search
- Transfer
- Globus Connect Personal
- Bookmarks
- Sharing
- Groups
- Management console
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
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
Data Publication and Discovery

Materials Data Facility Community home page

The Materials Data Facility (MDF) is a scalable repository where materials scientists can publish, preserve, and share research data. The repository provides a focal point for the materials community, enabling publication and discovery of materials data of all sizes.

MDF is a pilot project funded by NIST, and serves as the first pilot community of the National Data Service.

Contact Ben Blaiszik (blaiszik@uchicago.edu) to begin publishing your data
Peer reviewed paper data

(Re)format…
- PDF/A
- HDF
- …

Fully described…
- Dublin core metadata
- Domain metadata
- Provenance info

Replicated, public repositories

Formal, multi-step review
- Review → Update → Resubmit cycle

Persistent identifier
- DOI

(Re)format…

Replicated, public repositories
Globus collection policies

**Identifier**
- URL
- Handle
- DOI

**Description**
- None
- Standard
- Domain-specific
- Custom

**Curation**
- None
- Acceptance
- Human-validated
- Machine-validated

**Access**
- Anonymous
- Public
- Embargoed
- Collaborators

**Preservation**
- Transient
- Project Lifetime
- Archive
- “forever”
Raw NGS output

- Automated dataset acceptance
- Source environment
  - Instrument, timestamp,…
  - Unique ID
- Handle

Minimal metadata…

No curation

High durability, low cost store

Glacier
Analysis results

Optional metadata...
- “Implicit” metadata
- Description through organization

Team review
- Any collaborator may approve

Identify...
- Globus share

Widely accessible stores
Demo

• Publication

• Discovery
"A single global information space"
Making research data storage a first-class entity on the web
HTTPS access to endpoints
(coming soon)

• Enhanced use of research storage
  – Asynchronous, bulk transfer via GridFTP
  – (soon) Synchronous remote access via HTTPS

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

• Integration with clients and web apps
Cloud has transformed how software and platforms are delivered.

**Software as a service: SaaS**
(web & mobile apps)

**Platform as a service: PaaS**

**Infrastructure as a service: IaaS**

PaaS enables more rapid, cheap, and scalable delivery of powerful (SaaS) apps.
Platform 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?
# NCEP Climate Forecast System Version 2 (CFSv2) Monthly Products

**ds094.2**

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

<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><strong>Union of Available Products</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diurnal monthly means</td>
<td>Web File Listing</td>
<td>Request Globus Invitation</td>
<td>Get a Subset</td>
<td>GLADE File Listing</td>
</tr>
<tr>
<td>Regular monthly means</td>
<td>Web File Listing</td>
<td></td>
<td>Get a Subset</td>
<td>GLADE File Listing</td>
</tr>
<tr>
<td>Selected Parameter/Level Time Series</td>
<td>Web File Listing</td>
<td></td>
<td>Get a Subset</td>
<td>GLADE File Listing</td>
</tr>
</tbody>
</table>

For dataset access, please use the Globus Transfer Service (GridFTP).
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.

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.

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.
Globus Transfer API

Nearly all Globus Web App functionality implemented via public Transfer API

This API provides a REST-style interface to the Globus reliable file transfer service. The Transfer API supports monitoring the progress of a user’s file transfer tasks, managing file transfer endpoints, listing remote directories, and submitting new transfer and delete tasks. The API is ideal for integration into Portals or Gateways to provide complex reliable file transfer capabilities without having to develop and support these features on your own. It is also easy to use for scripting, using any standard HTTPS or REST client library in scripting languages like Python and Ruby.

Contents

- **API Overview** - overview of API with authentication instructions and examples
- **Endpoint Activation** - associate user credentials with an endpoint
- **Task Submission** - submit transfer and delete tasks
- **Task Management** - monitor and cancel background transfer and delete tasks
- **File Operations** - foreground filesystem operations, including directory listing (ls), creating directories (mkdir), and renaming files (rename)
- **Endpoint Management** - create, update, and delete endpoint definitions and servers
- **Endpoint Search** - find endpoints matching a search query, by display_name and other endpoint fields

[https://docs.globus.org/api/transfer/](https://docs.globus.org/api/transfer/)
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:

```bash
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:

```bash
git checkout https://github.com/globus/globus-sdk-python.git
cd globus-sdk-python
python setup.py install
```

Basic Usage

```bash
pip install globus-sdk-python
```

https://github.com/globus/globus-sdk-python
Prototypical research data portal
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...

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.

*If you would like to host a workshop at your institution please contact us.*

**Motivation:** New high-speed networks make it possible, in principle, to transfer

**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

[Workshops are free to attend and open to all](https://www.globusworld.org/tour/)
Globus Subscriptions

- **Globus Subscription**
  - Shared endpoints
  - Data publication
  - Management console
  - Usage reporting
  - Priority support
  - Application integration
  - HTTPS endpoints

- **Branded Web Site**

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

- **Premium Storage Connectors**

  [globus.org/subscriptions]
Summary

• Globus SaaS delivers advanced capabilities for researchers and research computing centers

• Globus PaaS opens new opportunities for developers creating applications for researchers