Welcome to the SGCI Webinar!

• We will be starting shortly.

• Your audio has been muted, and you are encouraged to turn off your video during the presentation.
  • Controls for these are near the bottom of the right-side control panel for BlueJeans.

• You may submit questions at any time using Chat, and the moderator will share them with the presenter when appropriate.

• This presentation will be recorded and slides will be posted.
Building the Modern Research Data Portal

Introduction to the Globus Platform

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

U.S. DEPARTMENT OF ENERGY

THE UNIVERSITY OF CHICAGO

NATIONAL INSTITUTES OF HEALTH

Argonne National Laboratory

powered by Amazon Web Services

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

NSF
<table>
<thead>
<tr>
<th>Statistic</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Major services</td>
<td>5</td>
</tr>
<tr>
<td>National labs using Globus</td>
<td>13</td>
</tr>
<tr>
<td>Registered users</td>
<td>45,000</td>
</tr>
<tr>
<td>Transferred PBs</td>
<td>245 PB</td>
</tr>
<tr>
<td>Files processed</td>
<td>30 Bn files</td>
</tr>
<tr>
<td>Active endpoints</td>
<td>10,000 endpoints</td>
</tr>
<tr>
<td>Active users/year</td>
<td>10,000 users</td>
</tr>
<tr>
<td>Uptime</td>
<td>99.9%</td>
</tr>
<tr>
<td>Institutional subscribers</td>
<td>60+</td>
</tr>
<tr>
<td>Largest single transfer</td>
<td>1 PB</td>
</tr>
<tr>
<td>Longest continuously managed transfer</td>
<td>3 months</td>
</tr>
<tr>
<td>Federated campus identities</td>
<td>130</td>
</tr>
</tbody>
</table>
...and thank YOU, our subscribers!
Welcome to GlobusWorld 2017!

Join us for the biggest annual gathering of all things Globus. Learn how to manage research data at scale, and connect with other research computing providers, HPC systems administrators, researchers, and developers who are building and using cutting-edge research data management solutions.

Our theme for 2017 is **Building the Research Applications and Services Ecosystem**. Campus IT professionals are challenged by the growing data management needs of researchers, as discovery in virtually all disciplines becomes data-driven. Our mission is to make it easy for developers to deliver powerful applications and services that support their researcher communities; and for research computing organizations to deliver services for managing data at every stage of the research lifecycle.

The GlobusWorld Tour...

Coming to a campus near you!

By popular demand, we're taking GlobusWorld on the road. Our tour includes multiple stops across the US where we will be presenting tutorials for developers and Globus administrators.

learn more...

https://www.globusworld.org/
Cloud has transformed how software and platforms are delivered

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

![TripIt](https://example.com/tripit.png)

**NETFLIX**

Platform as a service: **PaaS**

![Amazon](https://example.com/amazon.png)
![Microsoft](https://example.com/microsoft.png)

**force.com**

Infrastructure as a service: **IaaS**

![Amazon](https://example.com/amazon.png)
![Microsoft](https://example.com/microsoft.png)

**EC2**

**S3**

**Google Compute Engine**

PaaS enables more rapid, cheap, and scalable delivery of powerful (SaaS) apps
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

- Only a Web browser required
- Use storage system of your choice
- Access using your campus credentials
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?
**Example: NCAR RDA**

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

*ds094.2*

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

### 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>Subsetting</td>
<td>THREDDS Data Server</td>
<td>Central File System (GLADE) Holdings</td>
</tr>
<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>

**Globus Transfer Service (GridFTP)**
Globus PaaS

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

Sample Research Data Portal

github.com/globus/globus-sample-data-portal
Prototypical research data portal

- Identity Provider
- Globus Web Helper Pages
- Globus Auth
- Portal Web Server (Client)
- Portal Endpoint
- User’s Endpoint (optional)
- Global Cloud
- Other Services
- Other Endpoints
- Browser
- Applications
- Desktop
- Firewall
- GridFTP
- REST
- HTTPS
Prototypical research data portal

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

Connections:
- HTTPS
- REST
- GridFTP

Applications
Browser
Login

User's
Endpoint
(optional)
Globus Transfer API

• Nearly all Globus Web App functionality implemented via public Transfer API

docs.globus.org/api/transfer
Globus Python SDK

• Python client library for the Globus Auth and Transfer REST APIs

globus.github.io/globus-sdk-python

• Open source
Endpoint Search

• **Plain text search for endpoint**
  – Searches owner, display name, keywords, description, organization, department
  – Full word and prefix match

• **Limit search to pre-defined scopes**
  – all, my-endpoints, recently-used, in-use, shared-by-me, shared-with-me

• Returns: List of endpoint documents
Endpoint Management

- Get endpoint (by id)
- Update endpoint
- Create & delete (shared) endpoints
- Manage endpoint servers
Endpoint Activation

• Activating endpoint means binding a credential to an endpoint for login

• Globus Connect Server endpoint that have MyProxy or MyProxy OAuth identity provider require login via web

• Auto-activate
  – Globus Connect Personal and shared endpoints use Globus-provided credential
  – An endpoint that shares an identity provider with another activated endpoint will use credential

• Must auto-activate before any API calls to endpoints
File operations

- List directory contents (ls)
- Make directory (mkdir)
- Rename

**Note:**
- Path encoding & UTF gotchas
- Don’t forget to auto-activate first
Task submission

• Asynchronous operations
  – Transfer
    o Sync level option
  – Delete

• **Get submission_id, followed by submit**
  – Once and only once submission
Task management

- Get task by id
- Get task_list
- Update task by id (label, deadline)
- Cancel task by id
- Get event list for task
- Get task pause info
Bookmarks

- Get list of bookmarks
- Create bookmark
- Get bookmark by id
- Update bookmark
- Delete bookmark by id

- Cannot perform other operations directly on bookmarks
  - Requires client-side resolution
Shared endpoint access rules (ACLs)

- Access manager role required to manage permission/ACLs

- Operations:
  - Get list of access rules
  - Get access rule by id
  - Create access rule
  - Update access rule
  - Delete access rule
Management API

• Allow endpoint administrators to monitor and manage all tasks with endpoint
  – Task API is essentially the same as for users
  – Information limited to what they could see locally

• Cancel tasks

• Pause rules
Python SDK Jupyter notebook

• Jupyter (iPython) notebook demonstrating use of Python SDK

github.com/globus/globus-jupyter-notebooks

• Open source
Prototypical research data portal

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

**Login**
Next-Generation Portal Leverages Science DMZ

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

Portal Query/Browse Path

Data Transfer Path

https://fasterdata.es.net/
Prototypical research data portal

- Identity Provider
- Globus Web Helper Pages
- Globus Auth
- Portal Web Server (Client)
- Globus Transfer
- Other Services
- Other Endpoints
- Portal Endpoint

- Science DMZ
- Firewall
- Desktop
- User’s Endpoint (optional)
- Applications
- Browser

- HTTPS
- REST
- GridFTP
HTTPS to Endpoints

- Each endpoint HTTPS server is a Globus Auth service (resource server)
- Web page can link to file on server
  - Browser GET will cause HTTPS server to authorize request via Globus Auth (note SSO)
- Portal (client) can request scope for endpoint resource server
  - Use access token in requests
“A single global information space”
Prototypical research data portal

- Identity Provider
- Globus Web Helper Pages
- Globus Auth
- Portal Web Server (Client)
- Portal Endpoint
- Other Endpoints
- Other Services
- GridFTP
- REST
- HTTPS

Browser
Applications
Desktop
User’s Endpoint (optional)

Firewall
Globus Cloud
Science DMZ
Challenge

• **How to provide:**
  – Login to apps
    o Web, mobile, desktop, command line
  – Protect all REST API communications
    o App ➔ Globus service
    o App ➔ non-Globus service
    o Service ➔ service

• **While:**
  – 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
Globus Auth

- Identity and access management PaaS

[docs.globus.org/api/auth](docs.globus.org/api/auth)

- Introduction
- Developer Guide
- Reference
Based on widely used web standards

- **OAuth 2.0 Authorization Framework**
  - aka OAuth2

- **OpenID Connect Core 1.0**
  - aka OIDC

- **Use various OAuth2 and OIDC libraries**
  - Google OAuth Client Libraries (Java, Python, etc.), Apache mod_auth_openidc, etc.
  - Globus Python SDK
Globus account

• **A Globus account is a set of identities**
  – A *primary identity*
    o Identity can be primary of only one account
  – One or more *linked identities*
    o Identity can (currently) be linked to only one account

• **Account does not have own identifier**
  – An account is uniquely identified using its primary identity
Globus Auth interactions

User (Resource Owner) -> Login

App (Client) -> HTTPS/REST call -> Service (Resource Server)

Globus Auth (Authorization Server)

Identity Provider
Globus Auth interactions

1. Request authorization

App (Client) → Globus Auth (Authorization Server) → Service (Resource Server)

- For a set of scopes
  - Login: openid, email, profile
  - HTTPS/REST APIs
- User selects identity provider
Globus Auth interactions

1. Request authorization
2. Authenticates a resource owner

- Using existing identities
  - XSEDE, University (via InCommon), Google, web app, etc.
- User can link multiple identities into a single Globus Account
- No Globus username (Globus ID) required
- Globus Auth handles naming details, e.g., ePPN vs ePTID
Globus Auth interactions

1. Request authorization
2. Authenticates a resource owner
3. Obtains authorization (consent) for a client to access a resource

- Resource is provided by a resource server
- Limited by a scope
Globus Auth interactions

1. Request authorization
2. Authenticates a resource owner
3. Obtains authorization (consent) for a client to access a resource
4. Issues OAuth2 access_token to client

- Some grant types issue authorization code, which client exchanges for access token
- Access token is opaque to client
- May include a refresh token, for offline access
Globus Auth interactions

1. Request authorization
2. Authenticates a resource owner
3. Obtains authorization (consent) for a client to access a resource
4. Issues OAuth2 access_token to client
5. May issue OIDC id_token to client with resource owner identity

JWT id_token:
- sub: Globus Auth identity id
- iss: https://auth.globus.org
- name: full name
- preferred_username:
  - e.g., tuecke@uchicago.edu
- email: email contact
- other standard OIDC claims
Globus Auth interactions

1. Request authorization
2. Authenticates a resource owner
3. Obtains authorization (consent) for a client to access a resource
4. Issues OAuth2 access_token to client
5. May issue OIDC id_token to client with resource owner identity
6. HTTPS/REST call with access_token
Globus Auth interactions

1. Request authorization
2. Authenticates a resource owner
3. Obtains authorization (consent) for a client to access a resource
4. Issues OAuth2 access_token to client
5. May issue OIDC id_token to client with resource owner identity
6. HTTPS/REST call with access_token
7. Validates access_token for resource server, gets additional info

RFC 7662: OAuth 2.0 Token Introspection response:
- active: true or false
- client_id
- scope
- sub: Globus Auth identity id
- username: user@myu.edu
- identity_set: linked identities
- email
- name
- other standard claims
1. Request authorization
2. Authentiate a resource owner
3. Obtains authorization (consent) for a client to access a resource
4. Issues OAuth2 access_token to client
5. May issue OIDC id_token to client with resource owner identity
6. HTTPS/REST call with access_token
7. Validates access_token for resource server, gets additional info
8. Issues dependent access tokens to resource server

- Allows resource server to act as client to other resource servers
- Service uses request access_token to get a dependent access_token for each dependent service
- Service acts as client to its dependent services
Use case: Log in with Globus

- Similar to: “Log in with Google”  “Log in with Facebook”
- Using existing identities
- Providing access to community services
Branding

- Can skin Globus Auth pages
Use case: Portal calling services on user’s behalf

• **Examples:**
  – Portal starting transfer for user

• **Authorization Code Grant**
  – With service scopes
  – Can also request OIDC scopes

• **Confidential client**

• **Globus SDK:**
  – To get tokens: ConfidentialAppAuthClient
  – To use tokens: AccessTokenAuthorizer
Prototypical research data portal

- Science DMZ
- Firewall
- Desktop
- Prototypical research data portal
- Identity Provider
- Globus Web Helper Pages
- Globus Auth
- Portal Web Server (Client)
- Portal Endpoint
- User’s Endpoint (optional)
- Firewall
- Globus Cloud
- Globus Transfer
- Other Services
- Science DMZ
- GridFTP
- Other Endpoints
- HTTPS
- REST
Use case: Native apps

- **Examples**
  - Command line, desktop apps
  - Mobile apps
  - Jupyter notebooks
  - Any client that cannot keep a secret (downloaded)

- **Native app is registered with Globus Auth**
  - Not a confidential client

- **Native App Grant is used**
  - Variation on the Authorization Code Grant

- **Globus SDK:**
  - To get tokens: NativeAppAuthClient
  - To use tokens: AccessTokenAuthorizer
Mobile apps

- **Globus Auth supports mobile apps**
  - “Log in with Globus” in mobile apps
    - RFC 7636: Proof Key for Code Exchange by OAuth Public Clients (PKCE, pronounced “pixy”)
    - Extension to OAuth2 to allow OAuth2 Authorization Code Grant to work from mobile apps
  - Uses mobile browser for web-based login
  - Mobile apps can call any service REST APIs that use Globus Auth
  - iOS and Android
  - Same approach as used by Google, Facebook, etc.
Desktop & command line apps

- **Globus Auth “Native App” PKCE support**
- **Use browser if possible**
  - “OAuth 2.0 for Native Apps”
    - draft-ietf-oauth-native-apps-02
    - Use external browser if possible
    - Embed browser in app
    - Embed mini web server in app
- **Allows copy-n-paste of authorization code**
  - A little like app passwords, but OAuth2 compliant
- **Globus Python SDK and CLI support Native App login**
- **Limited support for username/password authentication**
  - Not recommended
Use case: Apps that need access tokens for long time

• **Examples:**
  – Portal checks for transfer status when user is not logged in
  – Run command line app from script

• **App requests refresh tokens**

• **Globus SDK:**
  – To get token: ConfidentialAppClient or NativeAppClient
  – To use tokens: RefreshTokenAuthorizer
Refresh tokens

- For “offline services”
  - E.g., Globus transfer service working on your behalf even when you are offline

- Refresh tokens issued to a particular client for use with a particular scope

- Client uses refresh token to get access token
  - Confidential client: client_id and client_secret required
  - Native app: client_secret not required

- Refresh token good for 6 months after last use

- Consent rescindment revokes resource token
Native App Examples

https://github.com/globus/native-app-examples

• README for install instructions

• ./example_copy_paste.py
  – Copy paste code to the app

• ./example_local_server.py
  – Local server to get the code

• ./example_copy_paste_refresh_token.py
  – Stores refresh token locally, uses it to get new access tokens
New Globus CLI (beta)

$ 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

https://github.com/globus/globus-cli
Use case: App invoking services as itself

• **Examples**
  – Sample portal invoking graph service and accessing endpoints as itself
  – Robots, agents, services

• **App registers with Globus to get client id and secret**
  – Native app cannot do this, because no client_secret

• **Client Credential Grant is used**

• **Globus SDK:**
  – To get tokens: ConfidentialAppAuthClient
  – To use tokens: AccessTokenAuthorizer
1. Authenticate with portal client id and secret

2. Access Tokens

3. Authenticate as portal with access tokens to invoke service
User identity vs. portal identity

• User logging into portal results in portal having user’s identity and access token
  – Used to make requests on the user’s behalf

• Portal may also need its own identity
  – Access and refresh tokens for this identity
  – Used to make requests on its own behalf
Prototypical research data portal

- Identity Provider
- Globus Web Helper Pages
- Globus Auth
- Portal Web Server (Client)
- Portal Endpoint
- Other Endpoints
- Other Services
- GridFTP
- REST
- HTTPS
- Desktop
- Applications
- Browser
- User’s Endpoint (optional)
- Firewall
- Science DMZ
- Globus Cloud
- Other
- Services

Applications

Login

User’s Endpoint (optional)
• Globus provided web pages designed for use by your web apps
  – Browse Endpoint
  – Activate Endpoint
  – Select Group
  – Manage Identities
  – Manage Consents
  – Logout

[Image of web pages and terminals showing Globus interfaces]

docs.globus.org/api/helper-pages
Prototypical research data portal

- Identity Provider
- Globus Web Helper Pages
- Globus Auth
- Portal Web Server (Client)
- Portal Endpoint
- Other Endpoints
- Global Transfer
- Other Services
- Global Cloud
- Science DMZ
- HTTPS
- REST
- GridFTP
- Global Transfer
- Other Services
- Global Cloud
- Science DMZ
- HTTPS
- REST
- GridFTP
Adding your identity provider

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

• Any other OpenID Connect identity provider (subscription)
Adding an identity provider

• If your portal has identities already:
  – Deploy OIDC server in front of it
    o Globus Python OIDC (coming soon)
    o Any standard OIDC server should work
    o Requires claim that can map to username
    o Optional claims: name, email, organization
  – Can register apps and services with an effective identity policy
    o Requires account to have identity from your identity provider when logging into your app
Portal accounts

- Your app portal can still have portal accounts for users
- Tie portal account to Globus account identity, rather than username/password
- Associate your profile with this account
- Globus Auth handles authentication of that identity, in order to log user into your portal account
Prototypical research data portal

- **Identity Provider**
- **Globus Web Helper Pages**
- **Globus Auth**
- **Portal Web Server (Client)**
- **Globus Transfer**
- **Other Services**

**Network Components**

- **Desktop**
  - **Login**
  - **Applications**
  - **User’s Endpoint (optional)**

- **Firewall**
  - **HTTPS**
  - **REST**

- **Science DMZ**
  - **Other Endpoints**

- **Globus Cloud**
  - **Globus**
  - **Auth**

- **Other**

**Protocols and Services**

- **HTTPS**
- **GridFTP**
- **REST**
- **Other Services**

**Applications**

- **Browser**
- **Portal Web Server (Client)**

**Endpoints**

- **Portal Endpoint**
- **Other Endpoints**
Why create your own services?

• **Front-end / back-end within your portal**
  – Remote backend for portal
  – Backend for pure Javascript browser apps

• **Extend your portal with a public REST API, so that other app and service developers can integrate with and extend your portal**
Why Globus Auth for your service?

• Outsource all identity management and authentication
  – Federated identity with InCommon, Google, etc.

• Outsource your REST API security
  – Consent, token issuance, validation, revocation
  – You provide service-specific authorization

• Apps use your service like all others
  – Its standard OAuth2 and OIDC

• Your service can seamlessly leverage other services

• Other services can leverage your service

• Implement your service using any language and framework

Add your service to the science cyberinfrastructure platform
Prototypical research data portal

- Identity Provider
- Globus Web Helper Pages
- Globus Auth
- Portal Web Server (Client)
- Portal Endpoint
- Other Endpoints
- Other Services
- GridFTP
- REST
- HTTPS

- Browser
- Applications
- Desktop
- User’s Endpoint (optional)
- Firewall
- Science DMZ
- Globus Cloud
- Other
- Services
- Globus Transfer
Dependent tokens

• Your service can act as client to other services (scopes)
  – Globus Transfer and Auth
  – XSEDE (e.g., Jetstream, XUP)
  – Other community services
  – Future: Commercial services (e.g., Google Drive)

• Entire service call tree consented by user and service owners
  – Rescinding consent revokes all dependent tokens

• Dependent tokens are restricted to a particular client, calling a particular scope, on behalf of a particular resource owner (e.g., user)
  – Restricted delegation!
Join the Globus developer community

• Join developer-discuss@globus.org mailing lists: globus.org/mailing-lists

• Python SDK is open source
  – github.com/globus/globus-sdk-python
  – Submit issues, pull requests
  – Discussions on developer-discuss@globus.org

• Jupyter notebook, sample data portal and native applications are open source on github

• Documentation: docs.globus.org
Welcome to GlobusWorld 2017!

Join us for the biggest annual gathering of all things Globus. Learn how to manage research data at scale, and connect with other research computing providers, HPC systems administrators, researchers, and developers who are building and using cutting-edge research data management solutions.

Our theme for 2017 is **Building the Research Applications and Services Ecosystem**. Campus IT professionals are challenged by the growing data management needs of researchers, as discovery in virtually all disciplines becomes data-driven. Our mission is to make it easy for developers to deliver powerful applications and services that support their researcher communities; and for research computing organizations to deliver services for managing data at every stage of the research lifecycle.

---

https://www.globusworld.org/
Thank you for participating!

• Please offer your feedback with our 30-second evaluation:  
  http://sciencegateways.org/webinareval

• Join us next month (April 12) for  
  Gateway Showcase featuring Greg Madey of the Ensayo project and another gateway TBD