Introduction to the Globus Platform
Automating Research with Globus

Vas Vasiliadis
vas@uchicago.edu

NC State - March 27, 2018
Use(r)-appropriate interfaces

Globus service

Web App

CLI

Rest API

GET /endpoint/go%23ep1
PUT /endpoint/vas#my_endpt
200 OK

X-Transfer-API-Version: 0.10
Content-Type: application/json

...
Globus Command Line Interface

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

Options:
- v, --verbose Control level of output
- h, --help Show this message and exit.
- F, --format [json|text] Output format for stdout. Defaults to text
--map-http-status TEXT Map HTTP statuses to any of these exit codes:
0,1,50-99. e.g. "404=50,403=51"

Commands:
bookmarks Manage Endpoint Bookmarks
config Modify, view, and manage your Globus CLI config.
delete Submit a Delete Task
endpoint Manage Globus Endpoint definitions
get-identities Lookup Globus Auth Identities
list-commands List all CLI Commands
login Login to Globus to get credentials for the Globus CLI
logout Logout of the Globus CLI
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
transfer Submit a Transfer Task
version Show the version and exit
whoami Show the currently logged-in identity.

Open source, uses Python SDK

docs.globus.org/cli

github.com/globus/globus-cli
Demonstration
Globus CLI
How do I go beyond simple scripts?
Globus Platform
Transfer API
Useful developer links

> github.com/globus
> docs.globus.org
Globus Transfer API

- Globus Web App consumes public Transfer API
- Globus APIs use JSON for documents and resource representations
- Resource named by URL (standard REST approach)
  - Query params allow refinement (e.g., subset of fields)
- Requests authorized via OAuth2 access token
  - Authorization: Bearer asdflkqhafsdafeawk

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

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

• `globus_sdk.TransferClient` class handles connection management, security, framing, marshaling

```python
from globus_sdk import TransferClient
tc = TransferClient()
```

globus.github.io/globus-sdk-python
• **TransferClient low-level calls**
  – post(), get(), update(), delete()

get(path, params=None, headers=None, auth=None, response_class=None)
  
  o path – path for the request, with or without leading slash
  o params – dict to be encoded as a query string
  o headers – dict of HTTP headers to add to the request
  o response_class – class response object, overrides the client’s default_response_class
  o Returns: GlobusHTTPResponse object
Anything you can do using Globus webapp…

- Endpoint search
- Endpoint management
- File operations (listing, rename …)
- Task submission (transfer, sync …)
- Task management
- Bookmarks
- Sharing and permission management
- Management Console
Demonstration
Jupyter Notebook

github.com/globus/globus-jupyter-notebooks
How can I do this in my [science gateway, data portal, web app, ...]?
Demonstration
Modern Research Data Portal

mrdp.globus.org  docs.globus.org/mrdp
Modern Research Data Portal helps maximize the value of the Science DMZ
Modern data apps leverage the Science DMZ

fasterdata.es.net/
Data access: NCAR RDA
Data distribution: ARM Climate Research Facility

![Data Selection Summary](image)

- **Data Selection Summary**
  - **Signed in as ANANTHAKRISHNANR1.**
  - **mergesonde1.mace@fkb M1 Generate Citation**
  - 274 file(s) // 6014 MB
  - **Order Complete Datastream** or **Extract Specific Measurements**
  - **Note:** All variables will be delivered for this datastream.
  - **Measurement:** Atmospheric temperature
  - **Variable:** Temperature // temp

- **Data Delivery Options**
  - **Combine files by datastream:** No
  - **File format:** NetCDF
  - **Remove data flagged by Data Quality Reports (DQR) of type**
    - Incorrect
    - Suspect

- **Extraction options only apply when "Extract Specific Measurements" is selected.**
- **Original files will be delivered as part of all orders.**

- **Data Delivery Options**:
  - FTP
  - Globus
  - THREDDS
  - Dropbox

- **Date Range**:
  - 2007-04-01
  - 2007-12-31
Sanger Imputation Service

This is a free genotype imputation and phasing service provided by the Wellcome Trust Sanger Institute. You can upload GWAS data in VCF or 23andMe format and receive imputed and phased genomes back. Click here to learn more and follow us on Twitter.

Before you start
Be sure to read through the instructions.
You will need to set up a free account with Globus and have Globus Connect running at your institute or on your computer to transfer files to and from the service.

Ready to start?
If you are ready to upload your data, please fill in the details below to register an imputation and/or phasing job. If you need more information, see the about page.

- Full name
- Organisation
- Email address

What is this
Globus user identity

News

11/05/2016
Thanks to EAGLE, we can now return phased data. The HRC panel has been updated to r1.1 to fix a known issue. See ChangeLog for more details.

15/02/2016
Globus API changed, please see updated instructions.

17/12/2015
New status page and reworked internals. See ChangeLog.

09/11/2015
Pipeline updated to add some features requested by users. See ChangeLog.
Globus Platform
Auth API
Globus Auth: Foundational IAM service

- Enables login for diverse app ecosystem
- Simplifies creation/integration of apps, services
- Outsources mundane feature development
- Brokers authentication and authorization interactions
- Protects REST API communications
- No new identity required
- Employs least privileges security model
- Programming language and framework agnostic
Globus Auth Features

- User
- App (Client)
- Native, mobile and web applications
- Federated identities
- Identity providers
- Identity providers
- Least privilege model
- App identity and authentication
- Offline Access
- Service (Resource Server)
- Dependent service security
- Dependent Services (Resource Servers)
- Services (Resource Servers)

- Resource server
- Globus Auth (Authorization Server)
- Identity providers
- Google
- NERSC
- XSEDE
- Argonne
- Exeter
- NCSA
- Galaxy
- jetstream

App (Client) to Service (Resource Server) via Globus Auth (Authorization Server)
Based on widely used web standards

• OAuth 2.0 Authorization Framework (a.k.a. OAuth2)
• OpenID Connect Core 1.0 (a.k.a. OIDC)
• Access via OAuth2 and OIDC libraries of your choice
  – Google OAuth Client Libraries (Java, Python, etc.), Apache mod_auth_openidc, etc.
  – Globus Python SDK

docs.globus.org/api/auth
Fundamental Concepts

• **Scopes:** APIs that client is requesting access to
  – Scope syntax: OpenID Connect: openid, email, profile
  – urn:globus:auth:scope:<service-name>::<scope-name>

• **Consents:** authorization client to access a service, within limited scope, on the resource owner's behalf
Globus account

• **Globus Account = Primary identity + Linked Identities**
  – An identity can be primary on only one account
  – Identities can be linked to only one account

• **Account does not have own identifier**
  – An account is uniquely identified using its primary identity
App registration

• Client_id and client_secret for service
• App display name
• Declare required scopes
  – Need long-term, offline refresh tokens?
  – May require authorization from scope admin
• OAuth2 redirect URIs
• Links for terms of service & privacy policy
• Effective identity policy (optional)

developers.globus.org
Use case: Calling services on user’s behalf

• **When to use:** web apps able to store secrets
  – e.g. portal initiating transfer, adding user(s) to shared endpoints

• **Confidential client**
  – Secret stored on web app server

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

• **Globus SDK:**
  – To get tokens: ConfidentialAppAuthClient
  – To use tokens: AccessTokenAuthorizer
Portal: FaceBase – Craniofacial data/resources

- **Globus Auth** for single sign on and federated login
- **Globus groups** for access control on resources
- **Globus branded site** for consistent look and feel
Use case: Native apps

• **When to use:** any client that cannot keep a secret
  – Command line, desktop apps, mobile apps. Jupyter notebooks

• **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
Automation Example: Data distribution

- Uses Auth and Transfer API via SDK
- Native app grant
- Client credential grant
  - Portal or service
  - Permission for the client id
Data Distribution: APS - DMagic

dmagic.readthedocs.io

Courtesy of Francesco De Carlo, Argonne National Laboratory (2016)
Demo: Simple automation (REST API & CLI)

[github.com/globus/automation-examples]

- See README for installation
- Using the REST API: share_data.py
  - Stage data to a folder and set sharing ACL (user and/or group)
- Using the CLI: share_data.sh
- Monitor and cleanup: cleanup_cache.py
Use case: Apps requiring long-lived access

• **When to use: “offline services”**
  – e.g. transfer working on your behalf even when you are offline
  – e.g. running command line app from script
  – e.g. checking transfer status when user is not logged in

• **App requests refresh tokens**

• **Globus SDK:**
  – To get token: ConfidentialAppClient or NativeAppClient
  – To use tokens: RefreshTokenAuthorizer
Automation Example: Repeated replication

- Using Globus CLI or SDK
- Meant to be run via cron or other task manager
- Native app grant
Demo: Native App/Refresh Tokens

github.com/globus/native-app-examples

• See README for installation

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

• ./example_copy_paste_refresh_token.py
  – Stores refresh token locally, uses it to get new access tokens
Analysis App: Wellcome Sanger

Sanger Imputation Service

This is a free genotype imputation and phasing service provided by the Wellcome Trust Sanger Institute. You can upload GWAS data in VCF or 23andMe format and receive imputed and phased genomes back. Click here to learn more and follow us on Twitter.

Before you start

Be sure to read through the instructions.

You will need to set up a free account with Globus and have Globus Connect running at your institute or on your computer to transfer files to and from the service.

Ready to start?

If you are ready to upload your data, please fill in the details below to register an imputation and/or phasing job. If you need more information, see the about page.

- Full name
- Organisation
- Email address
- What is this
  - Globus user identity

News

@sangerimpute

11/05/2016
Thanks to EAGLE, we can now return phased data. The HRC panel has been updated to r1.1 to fix a known issue. See ChangeLog for more details.

15/02/2016
Globus API changed, please see updated instructions.

17/12/2015
New status page and reworked internals. See ChangeLog.

09/11/2015
Pipeline updated to add some features requested by users. See ChangeLog.
Globus Helper Pages

- Globus pages designed for use by your web apps
  - Browse Endpoint
  - Activate Endpoint
  - Select Group
  - Manage Identities
  - Manage Consents
  - Logout

docs.globus.org/api/helper-pages
Globus PaaS developer resources

Python SDK

Installation

The Globus SDK requires Python 2.6+ or 3.2+. In a terminal:

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

This will install the Globus SDK and its dependencies. For Mac users:

- `brew install python-core`
- `pip install globus-sdk`

Bleeding-edge versions of the Globus SDK can be installed from source:

- `git clone https://github.com/globus/globus-sdk.git` (then `cd global-sdk-python` and `python set-up.py install`)

**Sample Application**

**Requirements**

- You need to be in the tutorial users group for sharing: [https://www.globus.org/app/groups/50b6a29c-63f](https://www.globus.org/app/groups/50b6a29c-63f)
- Installed Globus Python SDK

**Jupyter Notebook**

```python
from future import print_

tutorial_endpoint_1 = "dd59f0-6d04-11e5-ba46-22000b92c6ec" # endpoint "Globus CLI"
tutorial_endpoint_2 = "dd59f0-6d04-11e5-ba46-22000b92c6ec" # endpoint "Globus CLI"
tutorial_users_group = "50b6a29c-63ac-11e4-8062-22000ab68755" # group "Tutor"
```

**Configuration**

First you will need to configure the client with an OAuth2 access token. For the purpose of this tutorial, you can use a website. Click on the "Jupyter Notebook" option and copy the resulting text below, or click on "Globus CLI" and

```python
transfer_token = None # if None, tries to get token from ~/.globus.cfg file
```
Support resources

- **Globus documentation**: docs.globus.org
- **Helpdesk and issue escalation**: support@globus.org
- **Customer engagement team**
- **Globus professional services team**
  - Assist with portal/gateway/app architecture and design
  - Develop custom applications that leverage the Globus platform
  - Advise on customized deployment and integration scenarios
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