

Introduction to the Globus Platform and Data Portals

Steve Turoscy – sturoscy@globus.org

May 24, 2023



THE UNIVERSITY OF
CHICAGO





Globus Platform Services

...Integrate Globus capabilities into your data portal, science gateway, data commons or other research app

Globus PaaS accelerates development

- **Auth**
 - **Groups**
 - **Transfer**
 - Compute
 - **Search**
 - Timer
 - Flows
 - GCS Manager
- Globus web app consumes the same public APIs
 - Resources named by URL (standard REST approach)
 - Request/response body is JSON

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

Globus Python SDK

- **Python client library for the Globus REST APIs**
- **Largely direct mapping to REST API**
- **`globus_sdk.SearchClient` class handles index creation and deletion, metadata ingest, and querying**
- **`globus_sdk.TransferClient` class handles connection management, security, framing, marshaling**

globus-sdk-python.readthedocs.io
github.com/globus/globus-sdk-python



App registration





Get app credentials at

<https://app.globus.org/settings/developers>

FILE MANAGER

BOOKMARKS

ACTIVITY

COLLECTIONS

GROUPS

CONSOLE

FLOW

COMPUTE

SETTINGS

LOGOUT

HELP & SITEMAP

Settings

Identities

Consents

Globus Plus

Developers

Register an...

App

Register a service account or application credential for autom...

Applications that authenticate and act as the application itself. These appli... used for automation and as service or community accounts, and do NOT a... other users.

Register a portal, science gateway, or other application you h...

Applications that are OAuth confidential clients, and can manage a client s... requests for user consent.

Register a thick client or script that will be installed and run by... on their devices

Applications that are OAuth public clients or native applications that are di... users, and thus cannot manage a client secret.

Register a service API that will use Globus for authentication

A RESTful API (OAuth resource server) that uses Globus Auth tokens for au... Users accessing the service login via Globus and consent for the client to c...

Advanced registration

Register any client with Globus Auth - an application (confidential or publi... service account, or service API.

Django ESGF2 Portal

Contact Email: sturoscy@globus.org

Apps

Policies

Roles

[BACK TO REGISTERED APPS](#)

Django ESGF2 Portal

Client UUID

b19fbd77-a940-47bb-98d3-0fdec16c0bc6

Identity

b19fbd77-a940-47bb-98d3-0fdec16c0bc6@clients.auth.globus.org

App Type

Globus Connect Server

Client Secrets

Search and auth clients for Globus SDK

Created 234 days ago

UUID: bc7ad45d-4bdd-4854-8b5d-52b4b38583b0

Edit App Details

Add Client Secret

Delete App Registration

Globus Auth: Foundational IAM service

<https://globus-sdk-python.readthedocs.io/en/stable/services/auth.html>

- **Integral in making your portal widely and easily accessible**
- **Brokers authentication and authorization among...**
 - End-users
 - Identity providers: enterprise, external (federated identities)
 - Services: resource servers with REST APIs
 - Apps: web, mobile, desktop, command line clients
 - Services acting as clients to other services
- **OAuth 2.0 Authorization Framework (a.k.a. OAuth2)**
- **OpenID Connect Core 1.0 (a.k.a. OIDC)**

Fundamental Concepts

- **Scopes**
 - APIs that client is requesting access to
 - Service and resources within that service
- **Consents**
 - Authorizes a client to access a service, within limited scope, on the resource owner's behalf
- **Multiple methods for user to grant consent depending on the type of application**

Several authentication modes supported

- **A portal/science/gateway or other application you host**
 - Authorization code grant: authenticate as **user identity**
 - Browser redirect; auth code returned automatically; tokens stored securely
- **A command-line/thick client or script that is installed and run on the user's device**
 - Native app grant: authenticate as **user identity**
 - Auth code returned automatically; tokens stored per installation
- **Service account or application credentials for automation**
 - Client credentials grant: authenticate as **application identity**
 - Client ID and Secret stored securely
- **Application able to manage tokens for offline/long lived tasks**
 - Request refresh tokens in addition to access tokens

Managing service accounts/app credentials

- **Application Identity:**
appclientid@clients.auth.globus.org
- **These are confidential apps with client id and secret**
- **Ensure application is on a secure device**
- **Set up policy for rotation of secret**
- **Assign project admins to manage the registration**



Globus Groups: Use groups for authorization

- **Groups allow your portal to automatically manage access/permissions**
 - Grant membership manager role to your application
 - Your web app can add/remove members to grant/remove access
- **Use groups for your application's permissions**
 - Instead of managing a bunch of ACLs in your application, use group membership
 - Lookup membership
 - Check membership to determine permissions
 - Add/remove members
 - Configure policy settings
 - Create/delete groups

globus-sdk-python.readthedocs.io/en/stable/services/groups.html



Guest collections simplify data access, transfer

- **Users don't need local accounts to access data and the portal can act on the collections using its service account**
- **Grant the application Access Manager role**
 - Allows the application to manage permissions on the collection
 - Set for application identity:
`appclientid@clients.auth.globus.org`
- **Grant roles for management of endpoint and tasks**



TransferClient low-level calls

- **Thin wrapper around REST API**

- `post()`, `get()`, `update()`, `delete()`

`get(path, params=None, headers=None, auth=None, response_class=None)`

- `path` – path for the request, with or without leading slash
- `params` – dict to be encoded as a query string
- `headers` – dict of HTTP headers to add to the request
- `response_class` – class response object, overrides the client's `default_response_class`
- Returns: `GlobusHTTPResponse` object

globus-sdk-python.readthedocs.io/en/stable/services/transfer.html

TransferClient higher-level calls

- **One method for each API resource and HTTP verb**
- **Largely direct mapping to REST API**

```
endpoint_search(filter_fulltext=None,  
                filter_scope=None,  
                num_results=25,  
                **params)
```

 Getting tokens: jupyter.demo.globus.org

- Sign in with Globus and verify the consents
- Start My Server and open `globus-jupyter-notebooks`
- Run `Platform_Introduction_Native_App_Auth`
- Alternatively, use notebooks on your own machine:
github.com/globus/globus-jupyter-notebooks



Globus Search and the Django Globus Portal Framework

<https://github.com/globus/django-globus-portal-framework>





Making Data Findable with Globus Search



Data description and discovery

- Metadata store with fine-grained visibility controls
- Schema agnostic dynamic schemas
- Simple search using URL query parameters
- Complex search using search request document



docs.globus.org/api/search



Data ingest with Globus Search

POST /index/{index_id}/ingest'

```
{
  "ingest_type": "GMetaList",
  "ingest_data": {
    "gmeta": [
      {
        "id": "filetype",
        "subject": "https://search.api.globus.org/abc.txt",
        "visible_to": ["public"],
        "content": {
          "metadata-schema/file#type": "file"
        }
      },
      ...
    ]
  }
}
```



- Bulk create and update
- Task model for ingest at scale



Data ingest with Globus Search

POST /index/{index_id}/ingest'

```
{
  "ingest_type": "GMetaList",
  "ingest_data": {
    "gmeta": [
      {
        "id": "weight",
        "subject": "https://search.api.globus.org/abc.txt",
        "visible_to": ["urn:globus:auth:identity:46bd0f56-
                      e24f-11e5-a510-131bef46955c"],
        "content": {
          "metadata-schema/file#size": "37.6",
          "metadata-schema/file#size_human": "<501b"
        }
      },
      ...
    ]
  }
}
```



Search
Index



Visibility limited to Globus Auth identity

- Single user
- Globus Group
- Registered client application



Data discovery with Globus Search

GET /index/{index_id}/search?q=type%3Ahdf5

Simple query

```
{
  "@datatype": "GSearchResult",
  "@version": "2017-09-01",
  "count": 1,
  "gmeta": [
    {
      "@datatype": "GMetaResult",
      "@version": "2019-08-27",
      "entries": [
        { ... }
      ],
      "subject": "https://..."
    }
  ],
  "offset": 0,
  "total": 1
}
```



Search
Index





Data discovery with Globus Search

POST `/index/{index_id}/search`

```
{
  "filters": [
    {
      "type": "range",
      "field_name": "pubdate",
      "values": [
        {
          "from": "*",
          "to": "2020-12-31"
        }
      ]
    }
  ],
  "facets": [
    {
      "name": "Publication Date",
      "field_name": "pubdate",
      ...
    }
  ]
}
```

Complex query

Filter
Facets
Boosts
Sort
Limit



Search
Index





Example portals/science gateways/apps

Sanger Imputation Service **Beta**

Home About Instructions Resources

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 data. To learn more and follow us on Twitter.

ARM DATA DISCOVERY SEARCH RESULTS

HOME DATA SEARCH DATASTREAM SEARCH ARM DATA CENTER HELP FEEDBACK Getting Started

SEARCH

Search Text: Enter search text and press return key...

Start Date: (Start Date) End Date: (End Date)

[Expand All] [Close All] Clear Apply

CATEGORIES

ROUTINE DATA PJ DATA EVALUATION DATA MODEL DATA

INCORRECT SUSPECT MISSING NOTE LIMITED ACCESS

UCAR NCAR

Hello tuecke@uchicago.edu dashboard sign out

NCAR UCAR Research Data Archive Computational & Information Systems Lab

Home Find Data Ancillary Services About/Contact Data Citation Web

NCEP Climate Forecast System Version 2 (CFSv2) Monthly ds094.2

Description Data Access

Mouse over the table headings for detailed descriptions

Data Description	Data File Downloads		Customizable Data Requests	Other Access Methods	NCAR-Only Access	
	Web Server Holdings	Globus Transfer Service (GridFTP)	Subsetting	THREDDS Data Server	Central File System (GLADE) Holdings	Tape Archive (HPSS) Holdings
Union of Available Products	Web File Listing	Request Globus Invitation	Get a Subset	TDS Access	GLADE File Listing	HPSS File Listing

DLHub

Data and Learning Hub for Science

A simple way to find, share, publish, and run machine learning models and discover training data for science

Documentation

Examples Read the Docs Python SDK CLI

Papers and Presentations

DLHub on ArXiv DLHub Slides





Evolving the MRDP design pattern

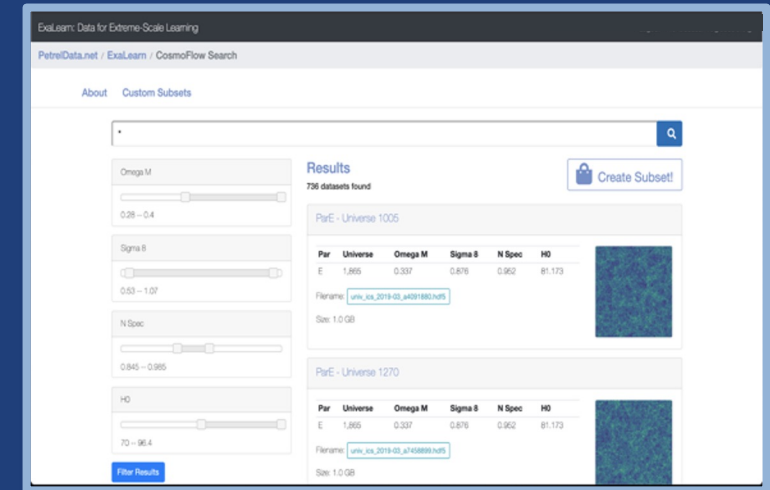


Globus Search



Ingest metadata, set
visibility policies

MRDP



Enabling discoverability:
MRDP + Faceted Search

Portal Core Functionality

- **User authentication**
- **Django-based framework**
 - Portal URL mappings
 - Token loading
- **Service calls to Globus Search**
- **Manage request lifecycle**
- **Post process search requests**



User authentication

- **Scopes are configured in the portal**
- **Users authenticate with Globus using standard flow**
 - Python Social Auth used for Authentication backend
- **User tokens are saved in the database**
- **Future requests authorized with user access tokens**
 - Searches use Search bearer token

```
SOCIAL_AUTH_GLOBUS_SCOPE = [  
    'urn:globus:auth:scope:search.api.globus.org:all',  
    'urn:globus:auth:scope:transfer.api.globus.org:all',  
    'urn:globus:auth:scope:groups.api.globus.org:view_my_groups_and_memberships',
```

Portal service calls use the Globus SDK

- Globus portal framework loads tokens from database
- Globus service object instantiated with token
- Call to Globus service(s)
- Portal renders result in templates

```
def list_my_files(request):  
    tc = load_transfer_client(request.user)  
    path = tc.operation_ls(request.GET['endpoint'], path=request.GET['path'])  
    return render(request, 'my_view.html', {'dir': path.data})
```



An index is configuration driven

- A Search index is configured in portal settings
- Add Globus Search index UUID
- Add a name
- Add facets
- Add fields
- Start searching!

```
'facets': [  
  {  
    "name": "Dates",  
    "field_name": "dc.dates.date",  
    "type": "date_histogram",  
    "date_interval": "day",  
  },  
],
```

Dates	
<input type="checkbox"/> Dec 18 2019	1
<input type="checkbox"/> Apr 01 2020	40
<input type="checkbox"/> Apr 02 2020	39
<input type="checkbox"/> Apr 07 2020	2
<input type="checkbox"/> Apr 14 2020	12
<input type="checkbox"/> Apr 21 2020	87
<input type="checkbox"/> Apr 28 2020	22

Lifecycle of a request

- **User makes a query**
- **Portal sends request to Globus Search**
 - Request contains user bearer token
- **Portal receives response**
- **Portal does processing on response**
 - Parse Dates, build URL for Globus webapp, etc.
- **Portal renders data into templates**
- **User receives a search page**



Creating your science gateway using the Globus portal framework

<https://bit.ly/gw-tut>

Source: github.com/globus/django-globus-portal-framework

Docs: django-globus-portal-framework.readthedocs.io/en/stable/





Application registration

<https://app.globus.org/settings/developers>

- Set redirect URLs
- Get client ID and secret
- Consents implement least privileges principle

The screenshot displays the Globus application registration interface. The top section, titled 'My Data Portal', shows the following configuration details:

- Client ID:** 5a9e5bd2-18e0-4a66-8f38-0ec8ec82885a
- Redirect URLs:** <https://mydataportal.globus.org/auth/callback>
- Scopes:**
 - urn:globus:auth:scope:transfer.api.globus.org:all (Manage data using Globus Transfer)
 - urn:globus:auth:scope:search.api.globus.org:all (Full access to Globus Search)
 - urn:globus:auth:scope:groups.api.globus.org:view_my_groups_and_memberships (View Groups and Memberships)
- Required Identity:** no specific identity provider required
- Pre-select Identity Provider:** no specific identity provider pre-selected
- Privacy Policy URL:** <https://mydataportal.globus.org/privacy>
- Terms and Conditions URL:** <https://mydataportal.globus.org/terms>
- Client Identity Username:** 5a9e5bd2-18e0-4a66-8f38-0ec8ec82885a@clients.auth.globus.org
- Client Secrets:** [Generate New Client Secret](#)

Below the configuration details, a consent dialog is shown. It lists the permissions COSMIC2 would like to use, each with a green checkmark:

- Manage data using G
- View identity details
- View your email address
- View your identity

The dialog also includes a 'To work, the above' section with two items:

- View the identities in your Globus account
- Manage your Globus Groups

At the bottom, a message states: 'By clicking "Allow", you allow COSMIC2 (this client has not provided terms of service or a privacy policy to Globus) to use the above listed information and services. You can rescind this and other consents at any time.' Below this message are 'Allow' and 'Deny' buttons.

Redirect URLs

<https://tutN.globusdemo.org:8443/>

<https://tutN.globusdemo.org:8443/complete/globus/>



Portal deployment

- **Install dependent libraries**
 - For production use, add robust WSGI/ASGI server
- **Deploy a portal instance using** `cookiecutter`
- **Configure settings**
- **Run and use!**
- **Future: containers**



Exporting data via Globus from easy to custom





Where's the data?

- **Remember – we're using Globus Connect, so your datasets are in a Globus collection**
- **Three options for enabling transfers from your portal:**
 1. Link to the collection in the Globus web app (Easy! But not customizable.)
 2. Use the Globus Helper Page (Easy! A bit customizable.)
 3. Use a JavaScript interface (Less easy. Very customizable.)
- **Let's see an example of each...**



Easy: Link to the Globus web app





**Easy: Select destination
with Globus Helper Page**



Bulk Transfer via Globus Helper Pages

Even though this is a little more complicated and requires some amount of additional setup, you get the benefit of using the Globus Web App to handle authorization as well as being able to customize your portal a little further

docs.globus.org/api/helper-pages/browse-endpoint

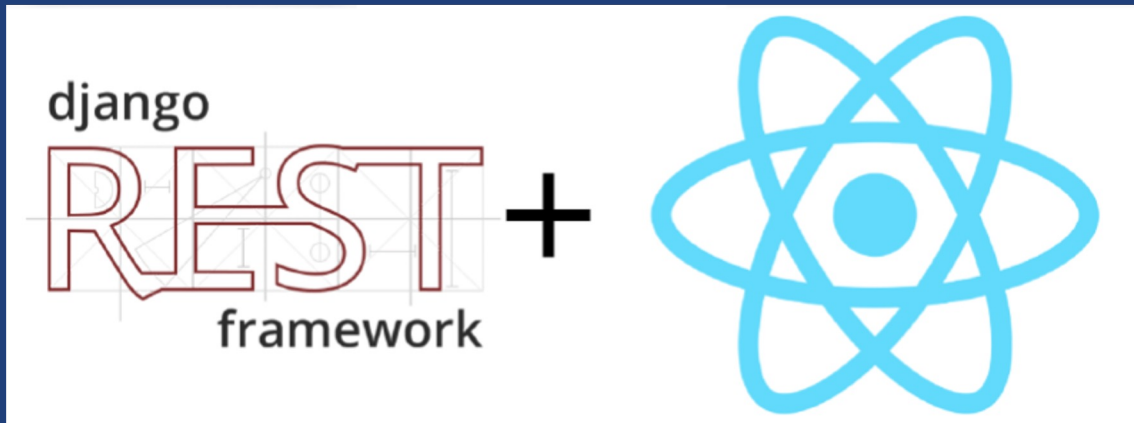


Advanced: Create a custom UI that uses Globus SDK



Advanced Customization via Django REST Framework and Javascript

More complicated!



- www.django-rest-framework.org
- reactjs.org
 - recoiljs.org/ (manage state)
 - reactrouter.com/en/main
- webpack.js.org

Why?

- Complete control over form and function
- Benefits of a Single Page Application (SPA)

Support resources

- Globus documentation: docs.globus.org
- GitHub: github.com/globus
- Migrating applications to use GCSsv5:
docs.globus.org/globus-connect-server/migrating-to-v5.4/application-migration/
- Sample code: github.com/globus/automation-examples
- YouTube channel: youtube.com/user/GlobusOnline