



Introduction to the Globus Platform

Automating Research with Globus

Vas Vasiliadis
vas@uchicago.edu

NC State - March 27, 2018

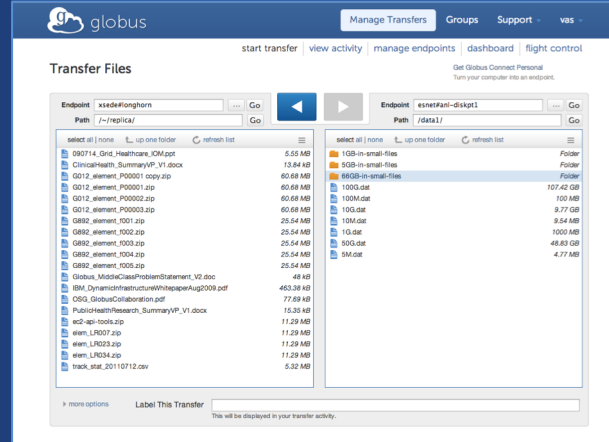




Use(r)-appropriate interfaces



Globus service



Web App

```
(globus-cli) jupiter:~ vas$ globus
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:
  bookmark      Manage Endpoint Bookmarks
  config        Modify, view, and manage your Globus CLI config.
```

CLI

```
GET /endpoint/go%23ep1
PUT /endpoint/vas#my_endpt
200 OK
X-Transfer-API-Version: 0.10
Content-Type: application/json
...
```

Rest API



Globus Command Line Interface

```
(globus-cli) jupiter:~ vas$ globus
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:
  bookmark      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](https://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 asdfkqhafsdafeawk

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**

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

globus.github.io/globus-sdk-python

SDK provides wrapper around REST API

- **TransferClient low-level calls**

- `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

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

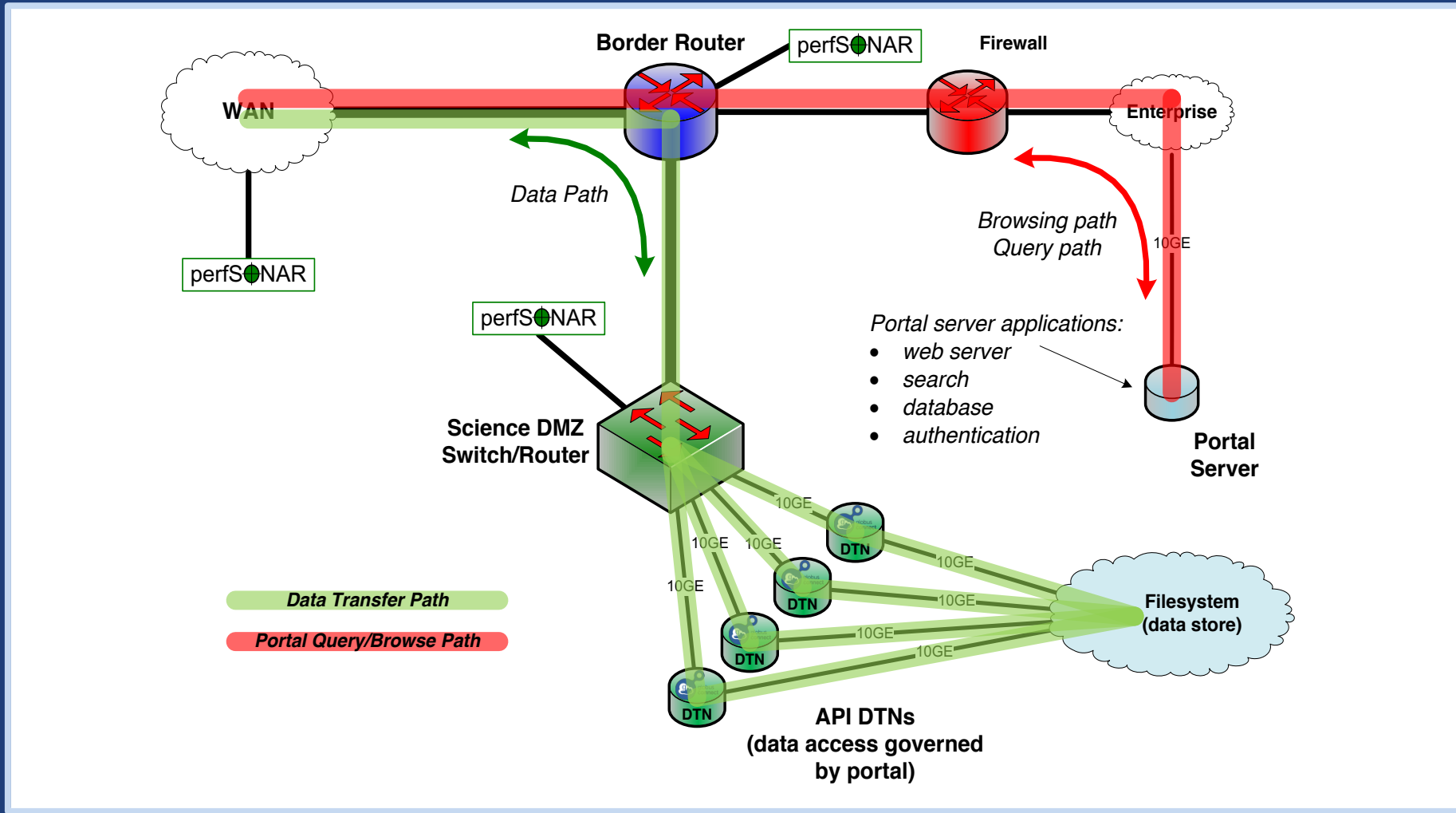
docs.globus.org/mrdp



**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

UCAR NCAR Closures/Emergencies Locations/Directions Find Pe

Hello [twacke@uchicago.edu](#) [dashboard](#) [sign out](#)

NCAR UCAR Research Data Archive Computational & Information Systems Lab *weather • data • climate*

Go to Dataset:

[Home](#) [Find Data](#) [Ancillary Services](#) [About/Contact](#) [Data Citation](#) [Web Services](#) [For Staff](#)

NCEP Climate Forecast System Version 2 (CFSv2) Monthly Products

ds094.2

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

[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
P R O D	Diurnal monthly means	Web File Listing	Get a Subset		GLADE File Listing	HPSS File Listing	
	Regular monthly means	Web File Listing	Get a Subset		GLADE File Listing	HPSS File Listing	



Data distribution: ARM Climate Research Facility

Data Selection Summary

Signed in as ANANTHAKRISHNANR1.

mergesonde1mace c1 @ fkb M1 [Generate Citation](#) 274 file(s) // 6014 MB

Order Complete Datastream Extract Specific Measurements

Note: All variables will be delivered for this datastream.

Measurement : Atmospheric temperature
Variable : Temperature // temp

2007-04-01

2007-12-31

Combine files by datastream

File format

Remove data flagged by Data Quality Reports (DQR) of type Incorrect Suspect

Data Delivery Options

- FTP
- Globus
- THREDDS
- Dropbox

Extraction options only apply when "Extract Specific Measurements" is selected.

Original files will be delivered as part of all orders.



Analysis app: Wellcome Sanger

Sanger Imputation Service **Beta**

Home About Instructions Resources Status

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

[Next](#)

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](#).

[See older news...](#)



Globus Platform

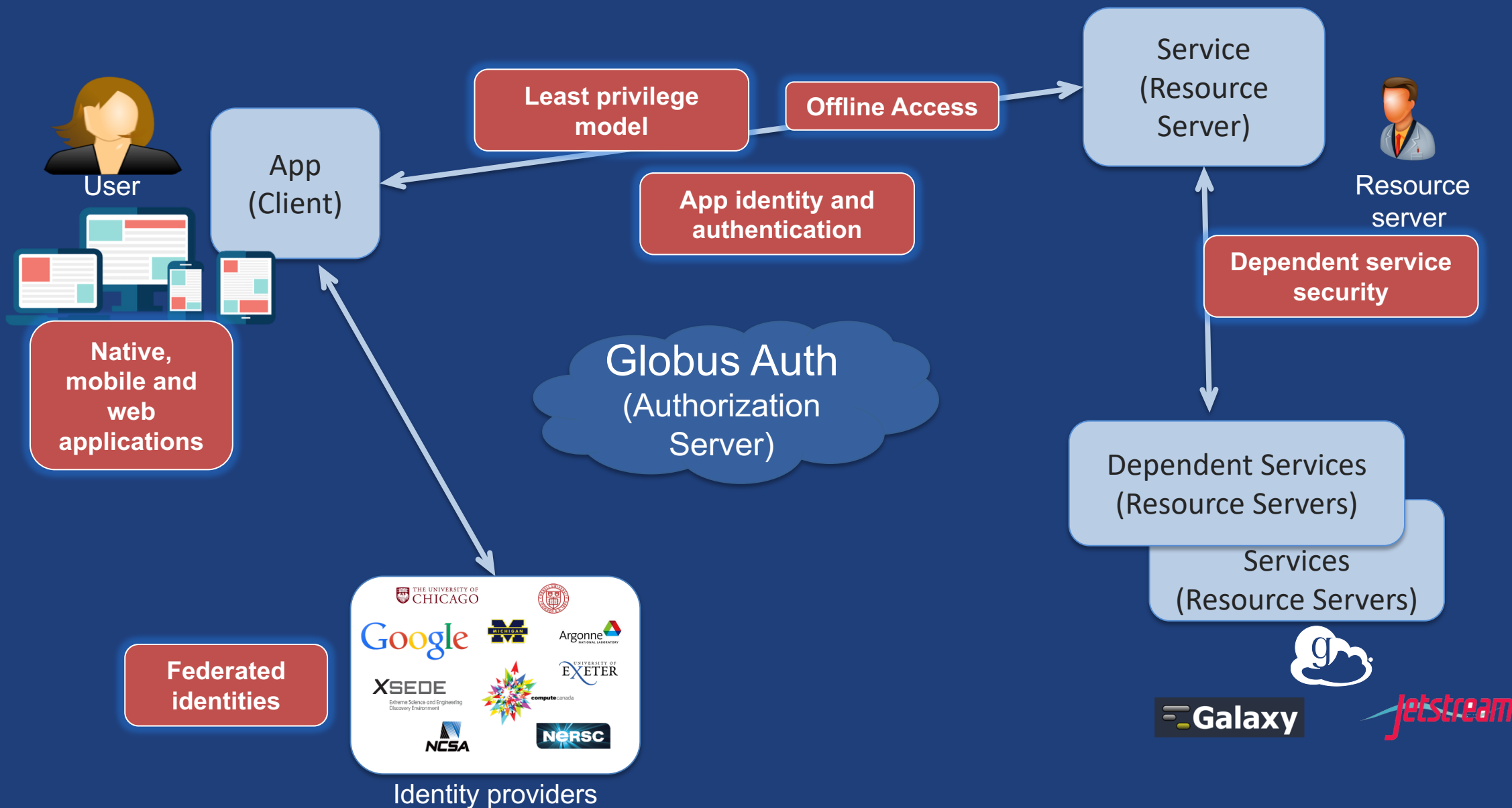
Auth API

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





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

FaceBase
A Resource For Craniofacial Researchers

globus Globus Account Log In

FaceBase uses Globus to provide account management services.

If you have an existing FaceBase account, log in as usual by selecting "Globus ID".

If you do not already have an account, log in using your organization or "Google" account.

If your organization is not listed, click "Globus ID" and then create a free account.

Your account must be approved before you may download FaceBase data.


Sign up to use FaceBase Globus Web App with a login that you already have!




Use your existing organizational login
e.g. university, national lab, facility, project, Google or [Globus ID](#)

University of Chicago

Didn't find your organization? Then use [Globus ID to sign in](#). ([What's this?](#))

Continue

 Globus uses CI Logon to enable you to Log In from this organization. By clicking Continue, you agree to the [CI Logon privacy policy](#) and you agree to share your username, email address, and affiliation with CI Logon and Globus. You also agree for CI Logon to issue a certificate that allows Globus to act on your behalf.

- **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`

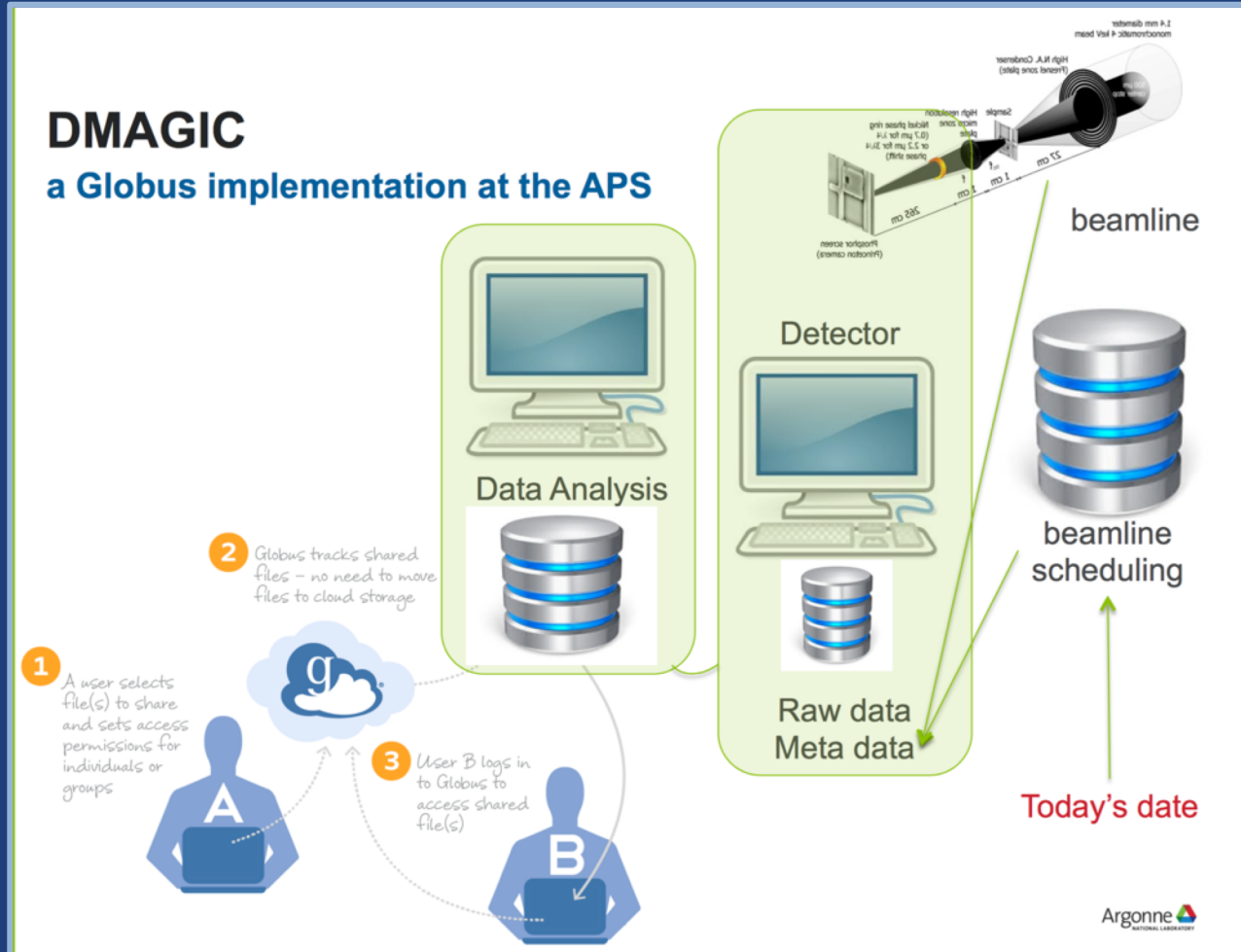
Automation Example: Data distribution



- **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

DMagic

latest

Search docs

About DMagic

Install directions

Development

API reference

Examples

Frequently asked questions

Docs » DMagic

Edit on GitHub

DMagic is an open-sourced Python toolbox to perform data management and data sharing for users of the Imaging Group of the Advanced Photon Source.

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

Recurring transfers
with sync option



Copy /ingest
Daily @ 3:30am



- **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 **Beta**

Home About Instructions Resources Status

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

[Next](#)

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](#).

[See older news...](#)



National Resource Access

XSEDE
Extreme Science and Engineering
Discovery Environment

globus Account ▾

Jetstream Web App would like to:

✓ Access all Jetstream resources

By clicking "Allow", you allow Jetstream Web information and services. You can rescind this

Allow **Deny**

globus Globus Account Log In

compute | **calcul**
canada | canada

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, 2016 is now Globus ID)

WestGrid ▾

Continue

Didn't find your organization? Then use Globus ID to [sign up](#).



Identity Management



- Home
- About ▾
- News ▾
- Developer Zone ▾
- KBase Labs

The new **Systems Biology Knowledgebase (KBase)** is a collaborative effort designed to accelerate our understanding of microbes, microbial communities, and plants. It will be a community-driven, extensible and scalable open source software framework and application system. KBase will offer free and open access to data, models and simulations, enabling scientists and researchers to build new knowledge and share their findings.

- Collaborate with us
- Get Started
- Develop with us

What can KBase do?

- ✔ Combine heterogeneous data types
- ✔ Offer standardized access to bioinformatic and modeling analyses
- ✔ Use evidence-supported annotations of genome structure and genetic function
- ✔ Discover new associations and network structures in community and molecular networks
- ✔ Map genotype to complex organismal traits
- ✔ Design and refine experiments using models of metabolism, regulation and community function
- ✔ Enable sharing of data, hypotheses, and newly-generated knowledge



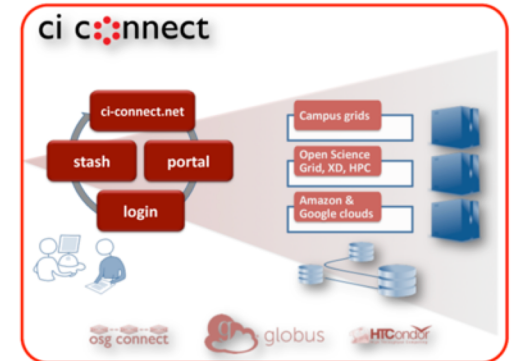
Accelerate research on campus by providing connective services for local, cloud and national cyberinfrastructure

Connecting Science to Cycles and Data

CI Connect offers campus IT organizations connective services for cyberinfrastructure. End-users quickly gain access to shared campus clusters while gaining access to distributed resources of national HPC or HTC facilities (such as the [Open Science Grid](#)) or cloud-based resources. CI Connect offers several capabilities based on [Globus](#), [HTCondor](#) and other HTC and data sharing technologies that augment your vision of providing scientific computational resources to research communities you support.

Reliable, High Performance CI Services

CI Connect provides services to CI providers at US universities and laboratories following a simple set of principles:



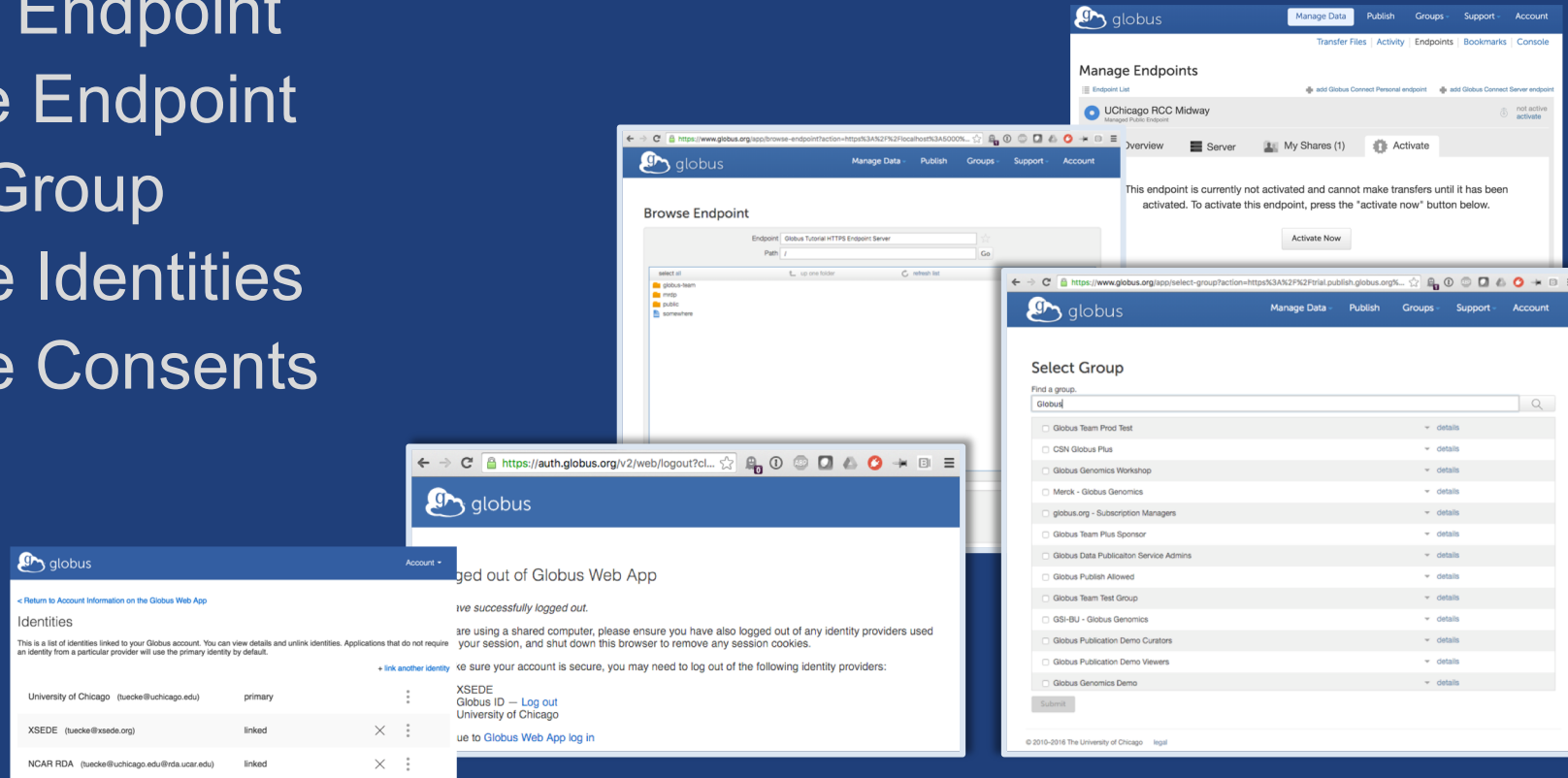
Connected environments from hosted services

Resources of a campus cluster (or [campus grid](#)) can be

2013-02-25
Proposed: Genomic Science
Contractors-Grantees Meeting

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

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

globus-sdk-python 0.2.5 documentation »

next | modules | index

Table Of Contents

- Globus SDK for Python (Beta)
- Installation
- Basic Usage
- API Documentation
- License

Python SDK

Installation

The Globus SDK requires Python 2.6+ or 3.2+. If a user does not have Python installed, they should first install Python.

The simplest way to install the Globus SDK is using the following instructions:

```
pip install globus-sdk
```

This will install the Globus SDK and its dependencies.

Bleeding edge versions of the Globus SDK can be installed using the following instructions:

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

Basic Usage

Modern Research Data Portal

LOGIN | SIGN UP

Modern Research Data Portal

It's how research data management is done!

Requirements

- You need to be in the tutorial users group for sharing: <https://www.globus.org/app/groups/50b6a29c-63ac-11e4-8062-22000ab68755>
- Installed Globus Python SDK

Jupyter Notebook

```
In [15]: from __future__ import print_function
tutorial_endpoint_1 = "ddb59ae1-0004-11e5-ba46-22000b92c6ec" # endpoint "Globus"
tutorial_endpoint_2 = "ddb59af0-6d04-11e5-ba46-22000b92c6ec" # endpoint "Globus"
tutorial_users_group = "50b6a29c-63ac-11e4-8062-22000ab68755" # group "TutorialUsers"
```

Configuration

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

```
In [16]: transfer_token = None # if None, tries to get token from ~/.globus.cfg file
```

Sample Application

docs.globus.org/api

github.com/globus

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](https://twitter.com/globusonline)