The globusworld TOUR

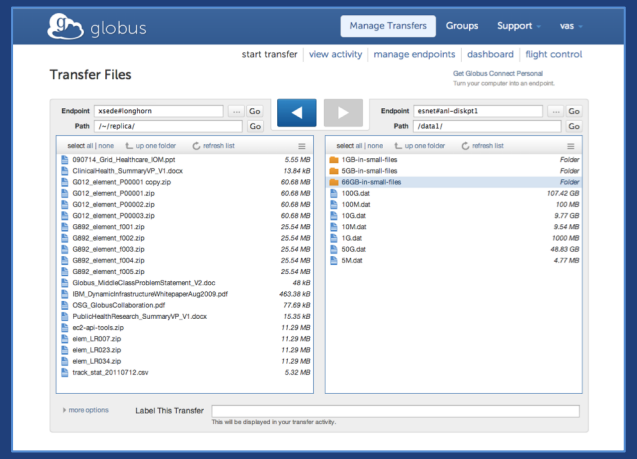
Introduction to the Globus Platform
Automating Research with Globus

Vas Vasiliadis
vas@uchicago.edu

NC State - March 27, 2018

globus

Use(r)-appropriate interfaces



Web App

Globus service

```
(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

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**

```
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, …]?

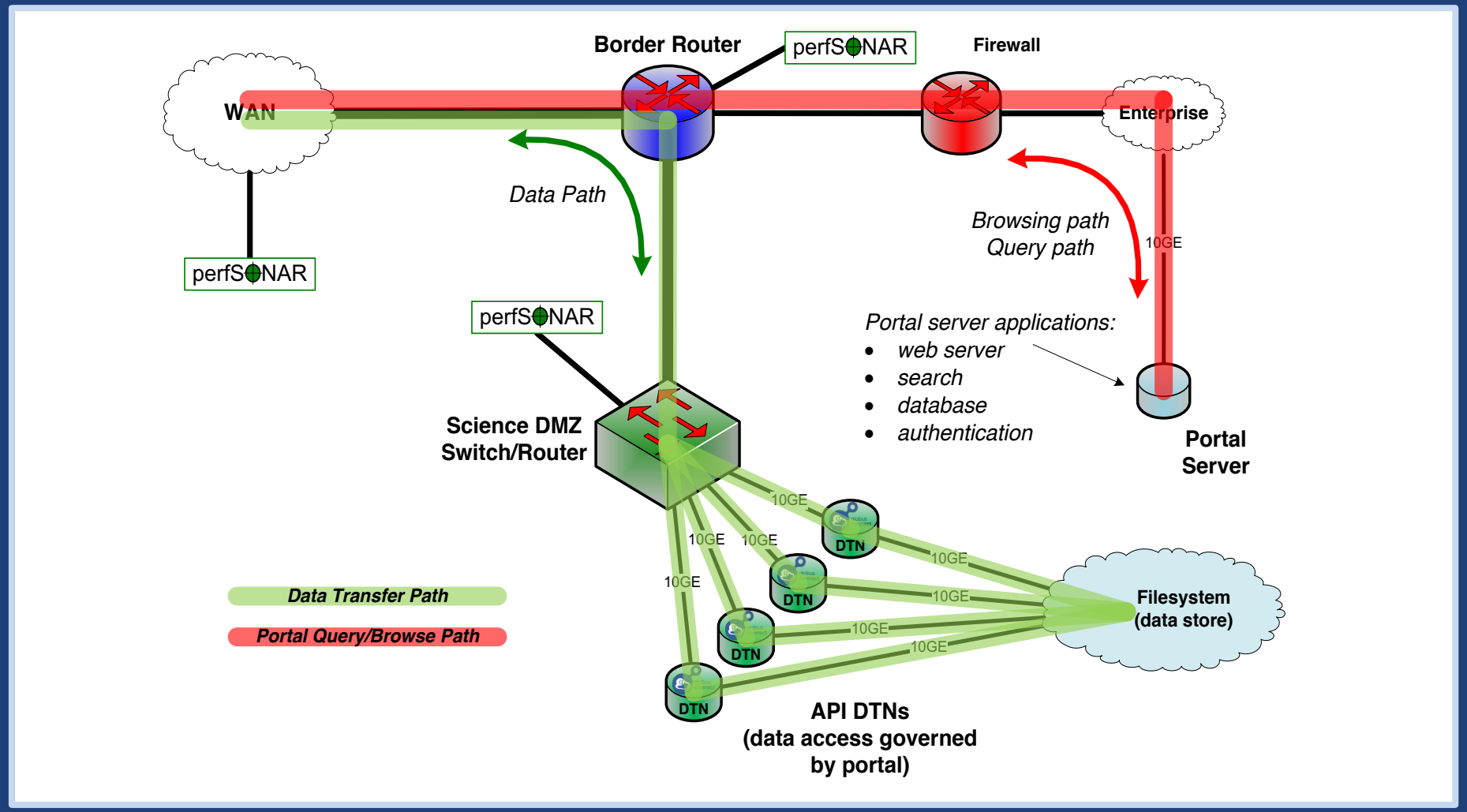
Demonstration

Modern Research Data Portal

mrdp.globus.org docs.globus.org/mrdp

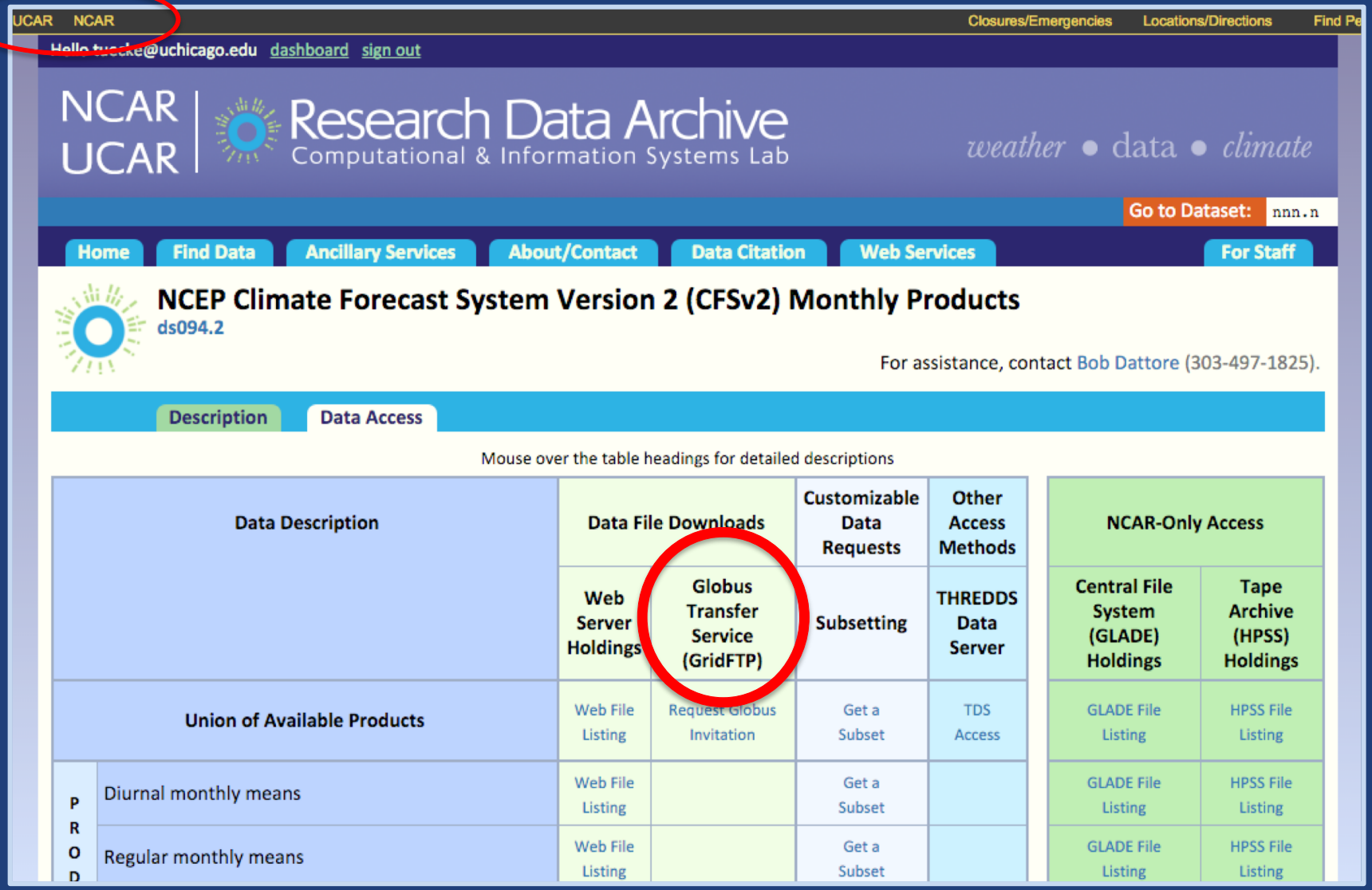
Modern Research Data Portal helps maximize the value of the Science DMZ

Modern data apps leverage the Science DMZ

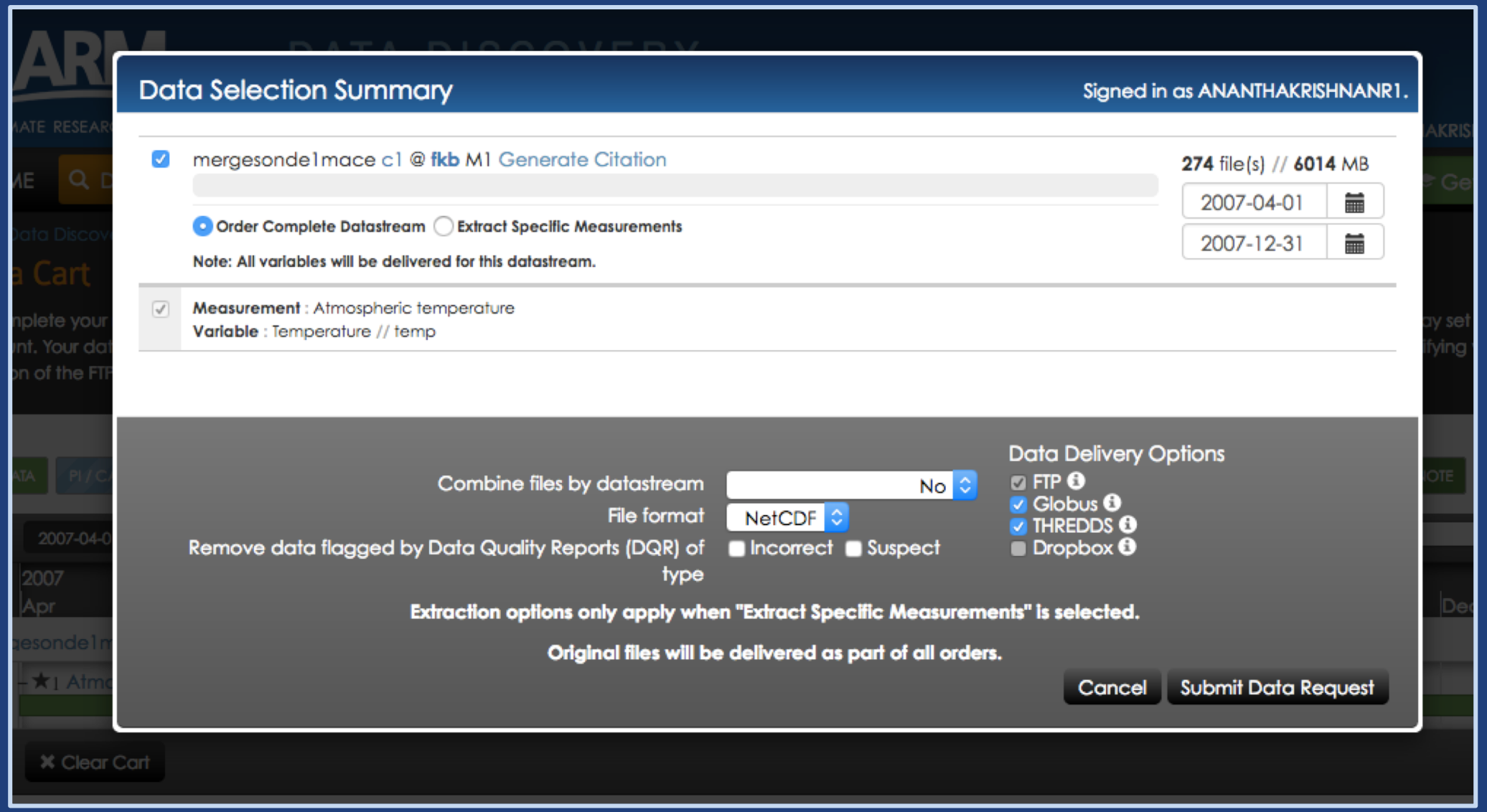


fasterdata.es.net/

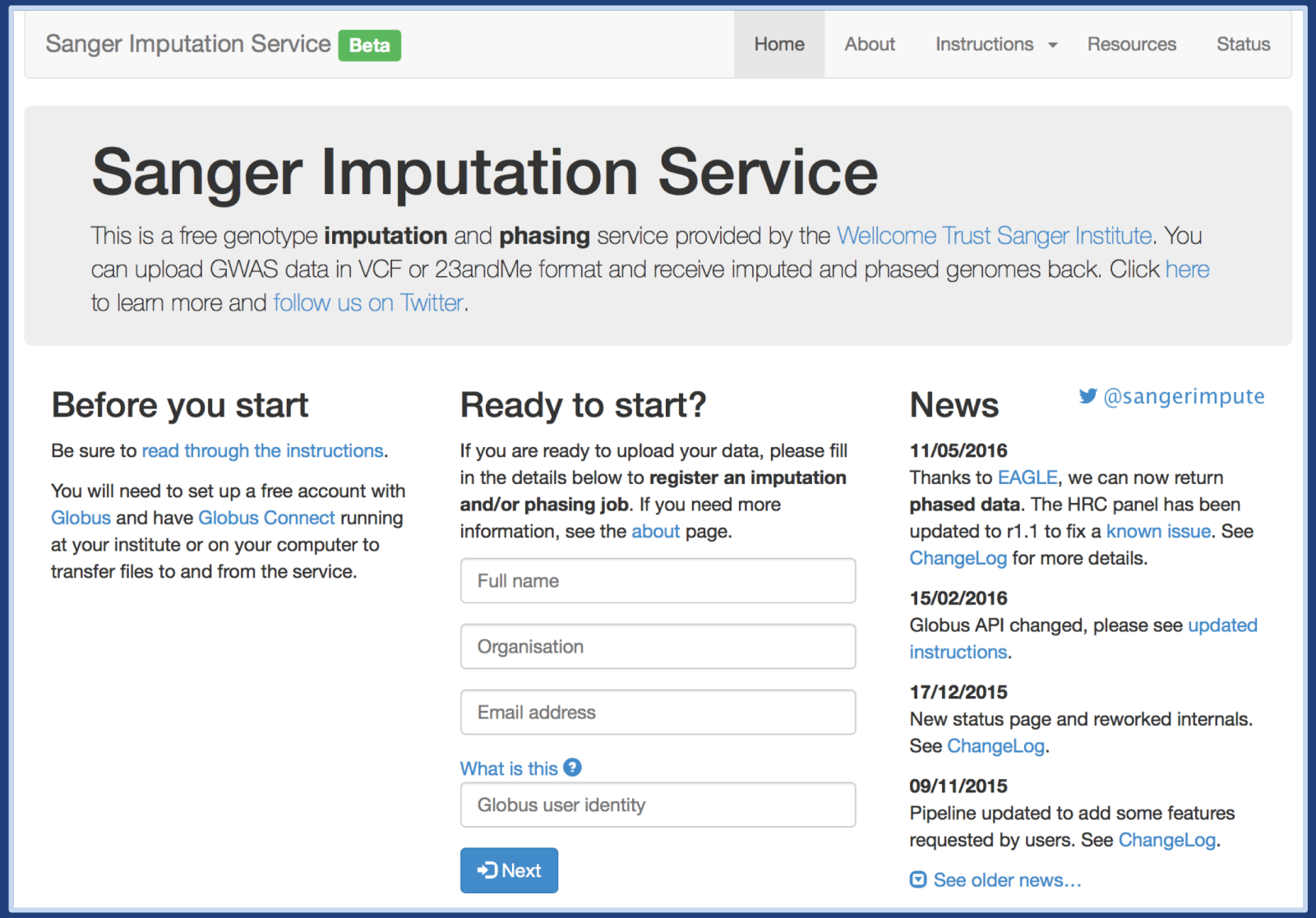
Data access: NCAR RDA



Data distribution: ARM Climate Research Facility



Analysis app: Wellcome Sanger



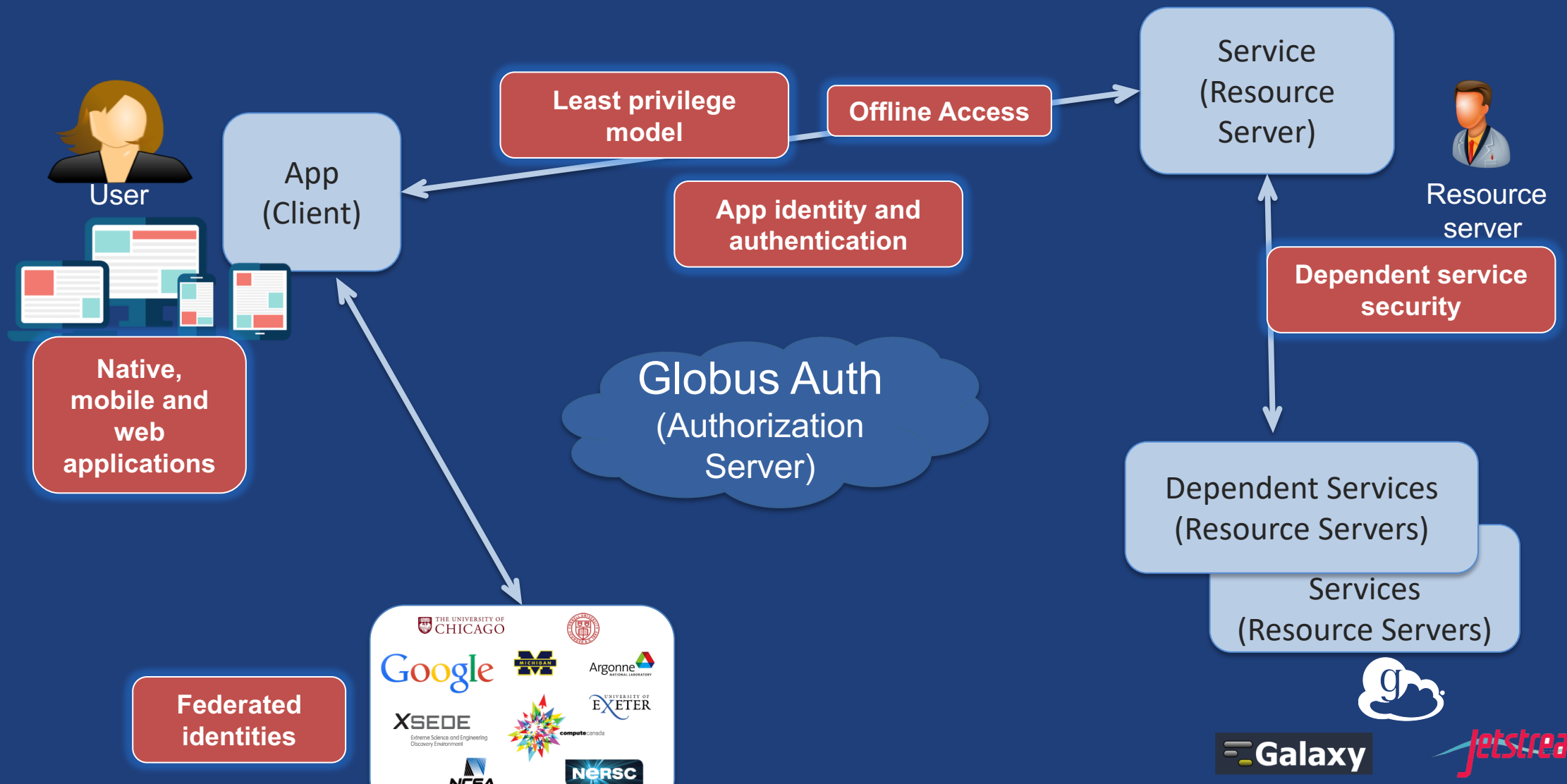
Globus Platform

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

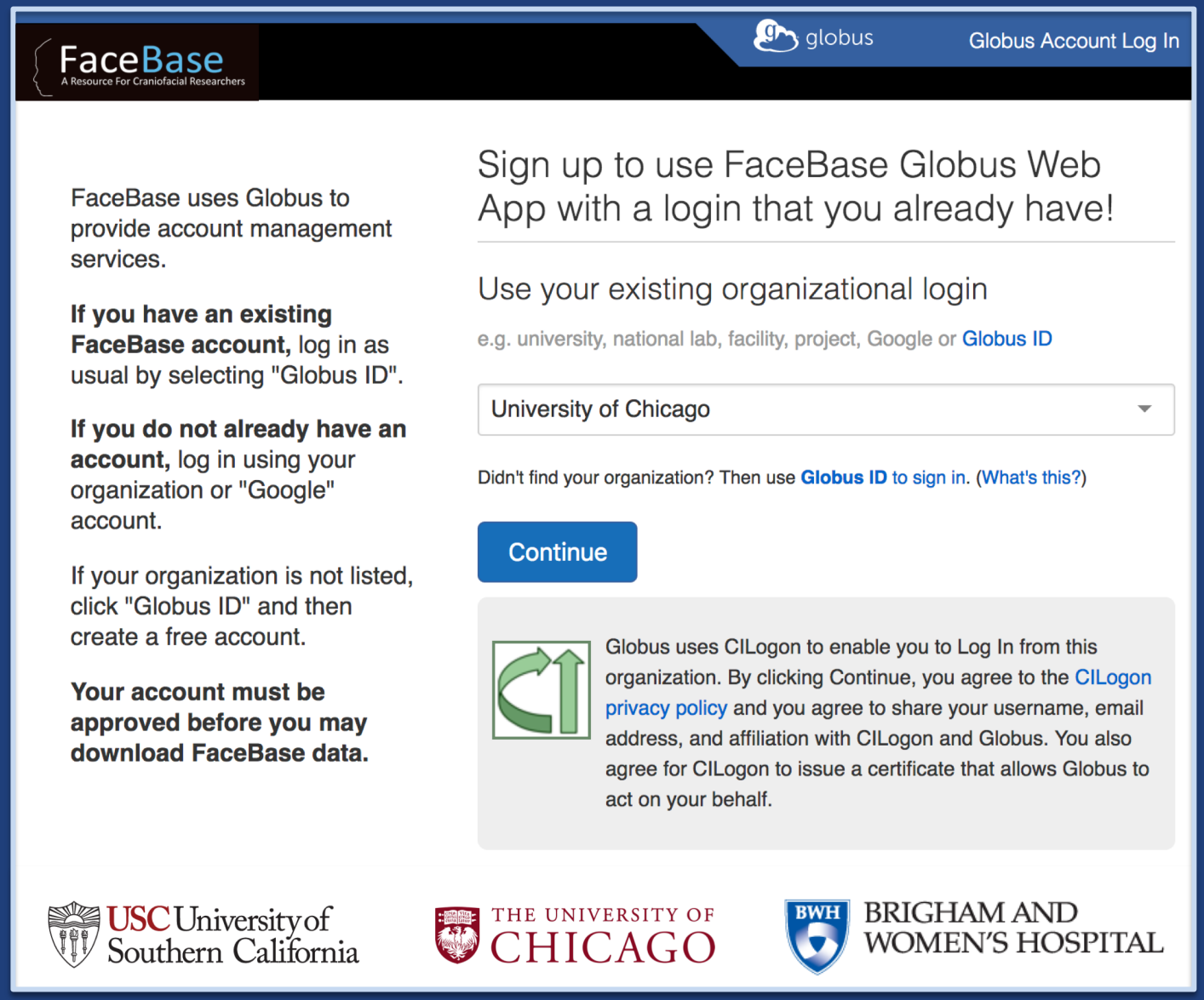
- **Client_id and client_secret for service**
- **App display name**
- **Declare required scopes**
  - Need long-term, offline refresh tokens?
  - May require authorization from scope admin
- **OAuth2 redirect URIs**
- **Links for terms of service & privacy policy**
- **Effective identity policy (optional)**

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

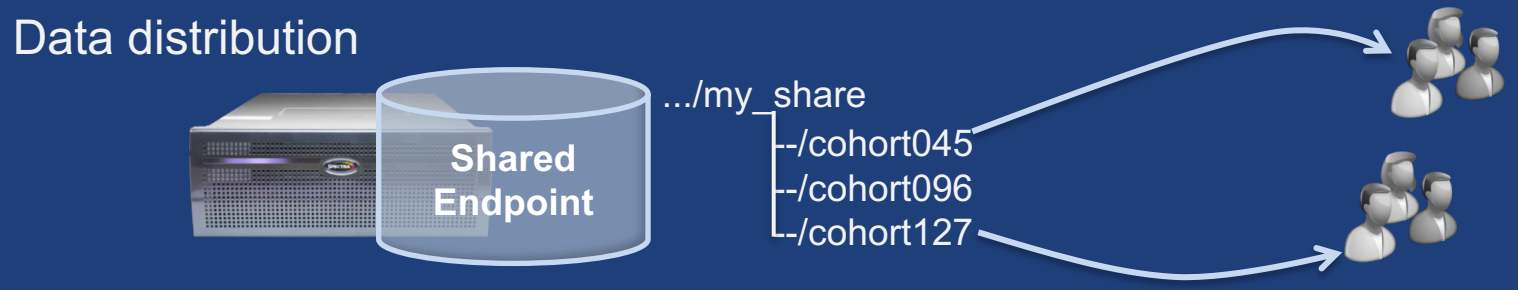


- **Globus Auth for single sign on and federated login**
- **Globus groups for access control on resources**
- **Globus branded site for consistent look and feel**

Use case: Native apps

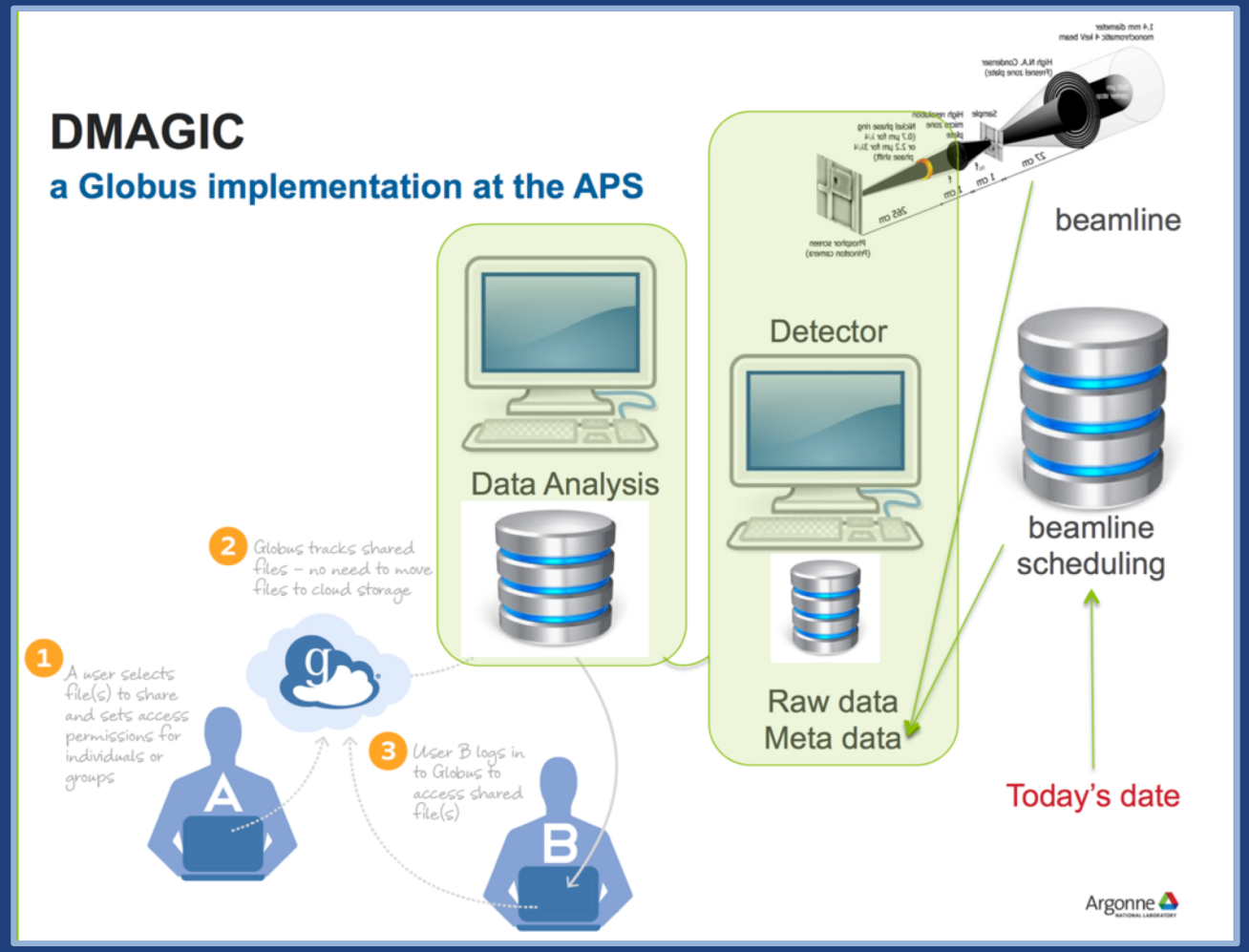
- **When to use: any client that cannot keep a secret**
  - Command line, desktop apps, mobile apps. Jupyter notebooks
- **Native app is registered with Globus Auth**
  - Not a confidential client
- **Native App Grant is used**
  - Variation on the Authorization Code Grant
- **Globus SDK:**
  - To get tokens: NativeAppAuthClient
  - To use tokens: AccessTokenAuthorizer

Automation Example: Data distribution

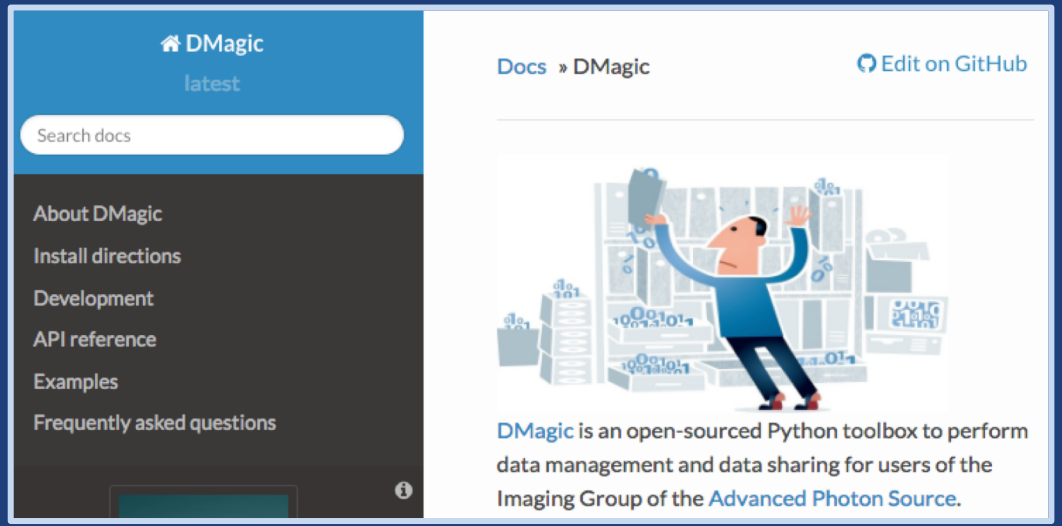


- **Uses Auth and Transfer API via SDK**
- **Native app grant**
- **Client credential grant**
  - Portal or service
  - Permission for the client id

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

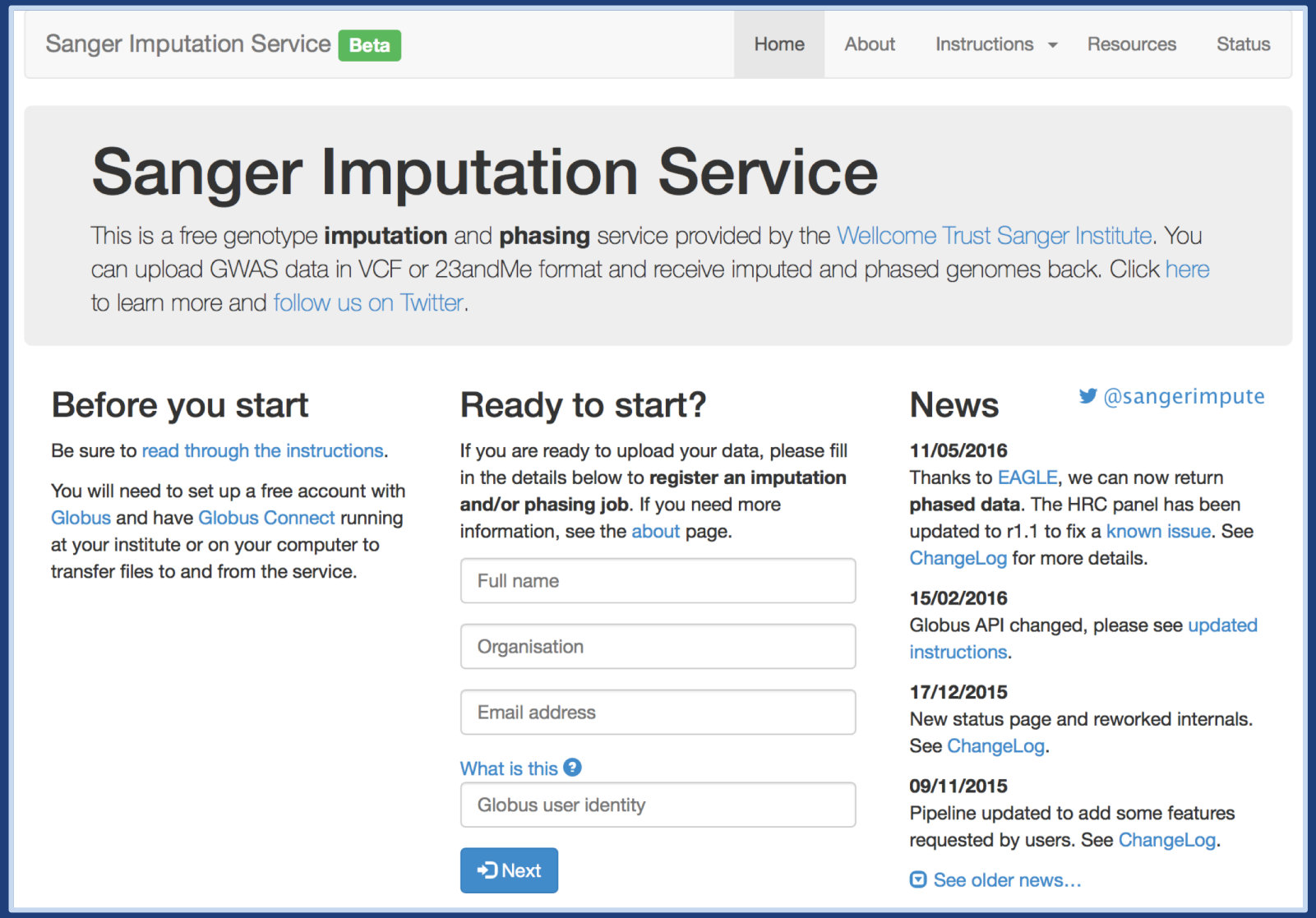
- **Using Globus CLI or SDK**
- **Meant to be run via cron or other task manager**
- **Native app grant**

Demo: Native App/Refresh Tokens

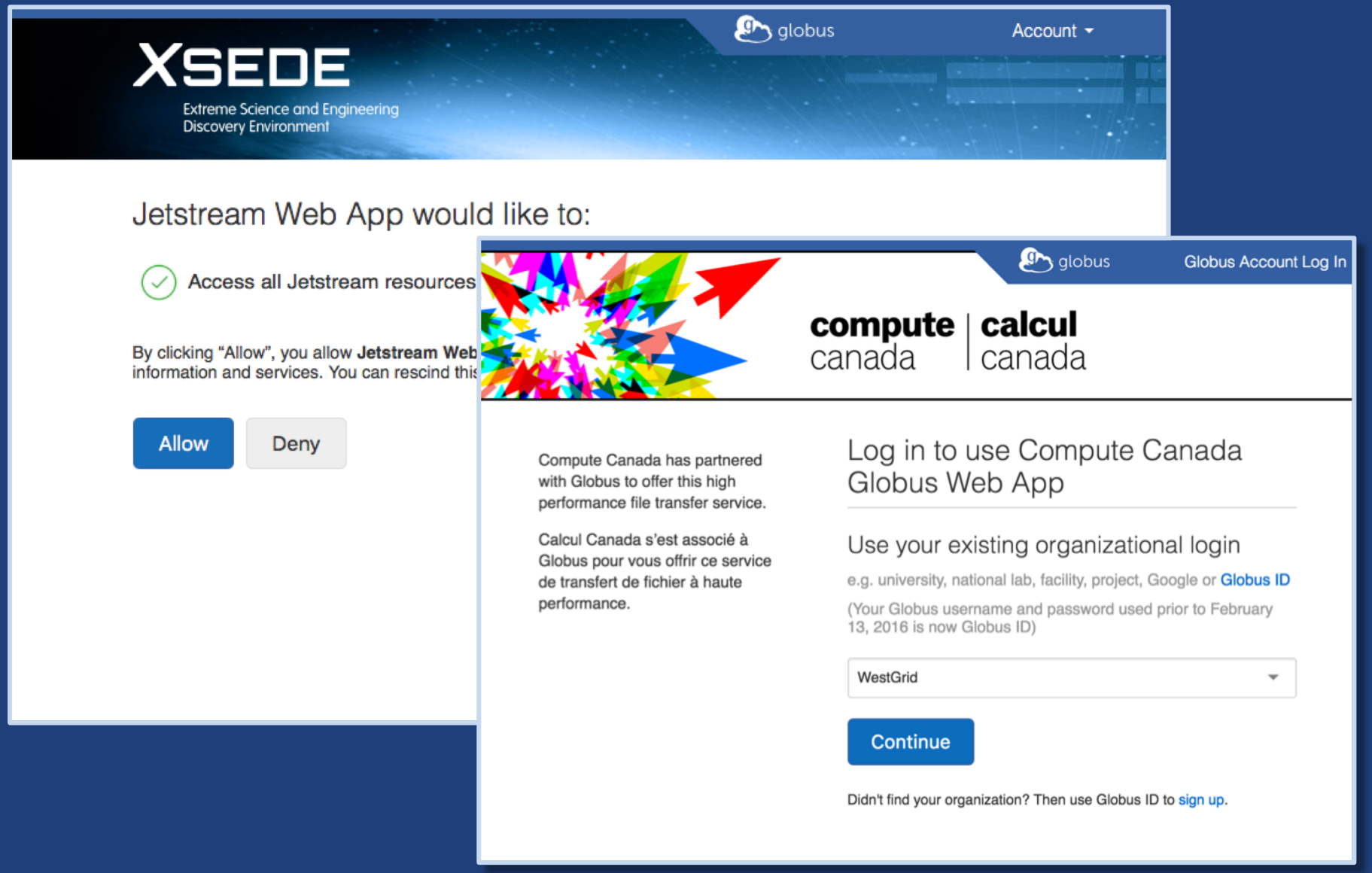
github.com/globus/native-app-examples

- **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



National Resource Access



Identity Management

KBase
PREDICTIVE BIOLOGY
DOE Systems Biology Knowledgebase

Search this site

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 ope source software framework and application system. KBase will offer free and open access to data, models and simulations, enabling scientists and researchers to buil new knowledge and share their findings.

Collaborate with us | Get Started | Develop with us

What can KBase do?

- Combine heterogenous 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 funct
- Enable sharing of data, hypotheses, and newly-generated knowledge

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

ci connect

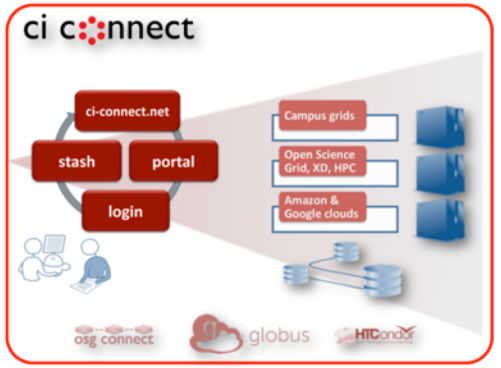
Accelerate research on campus by providing connective services for local, cloud and national cyberinfrastructure

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:

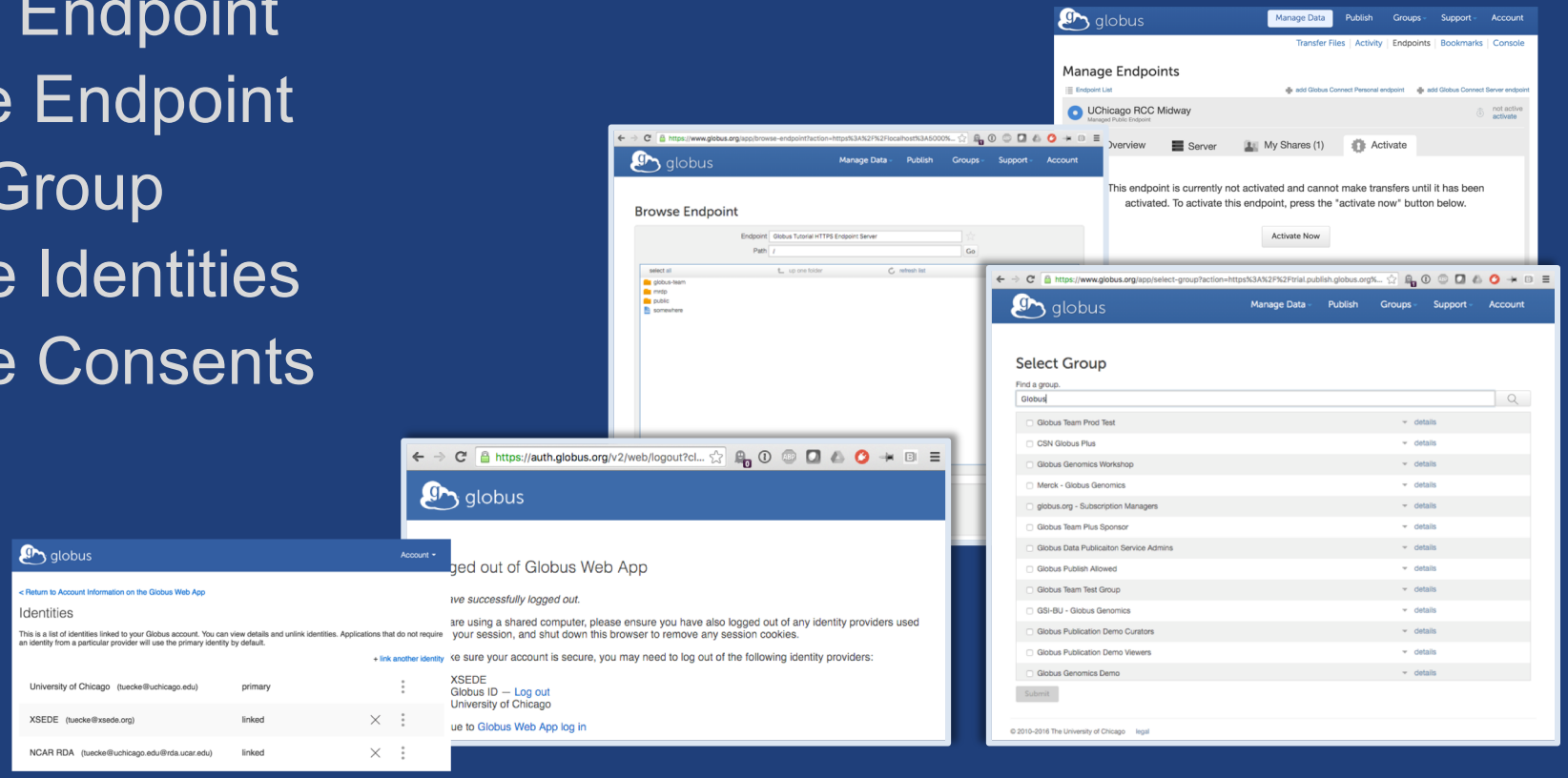


Connected environments from hosted services

Resources of a campus cluster (or campus grid) can be

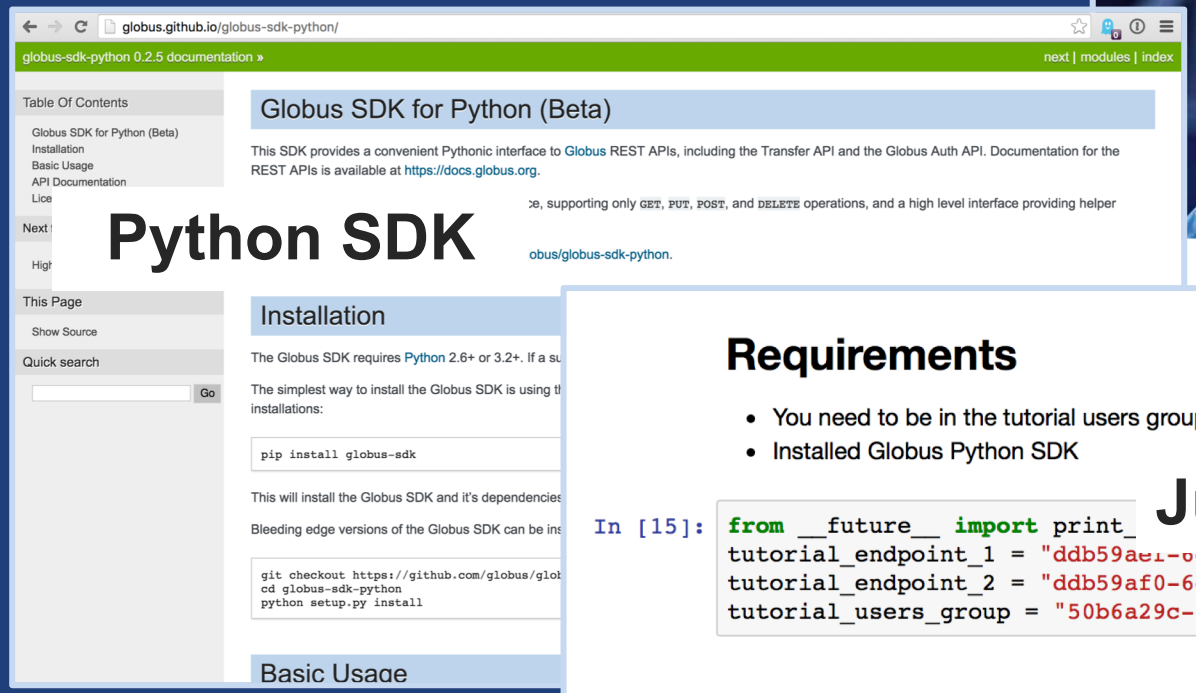
Globus Helper Pages

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

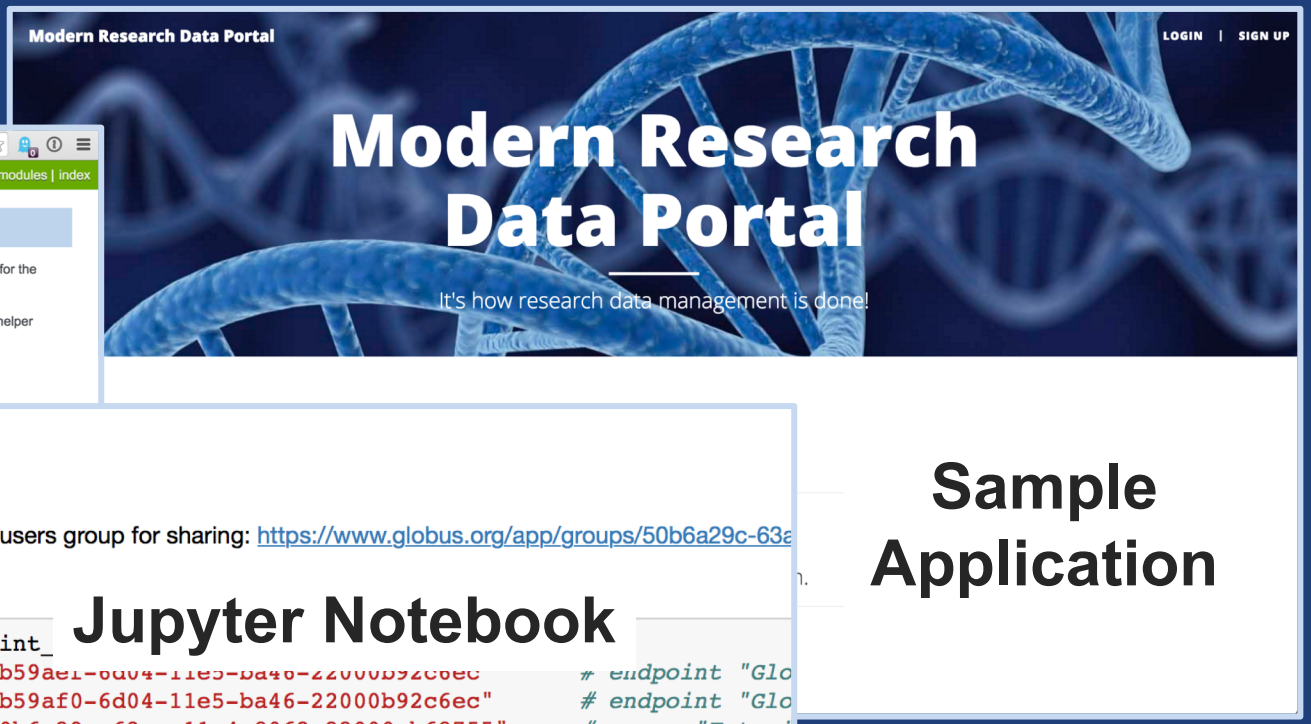


docs.globus.org/api/helper-pages

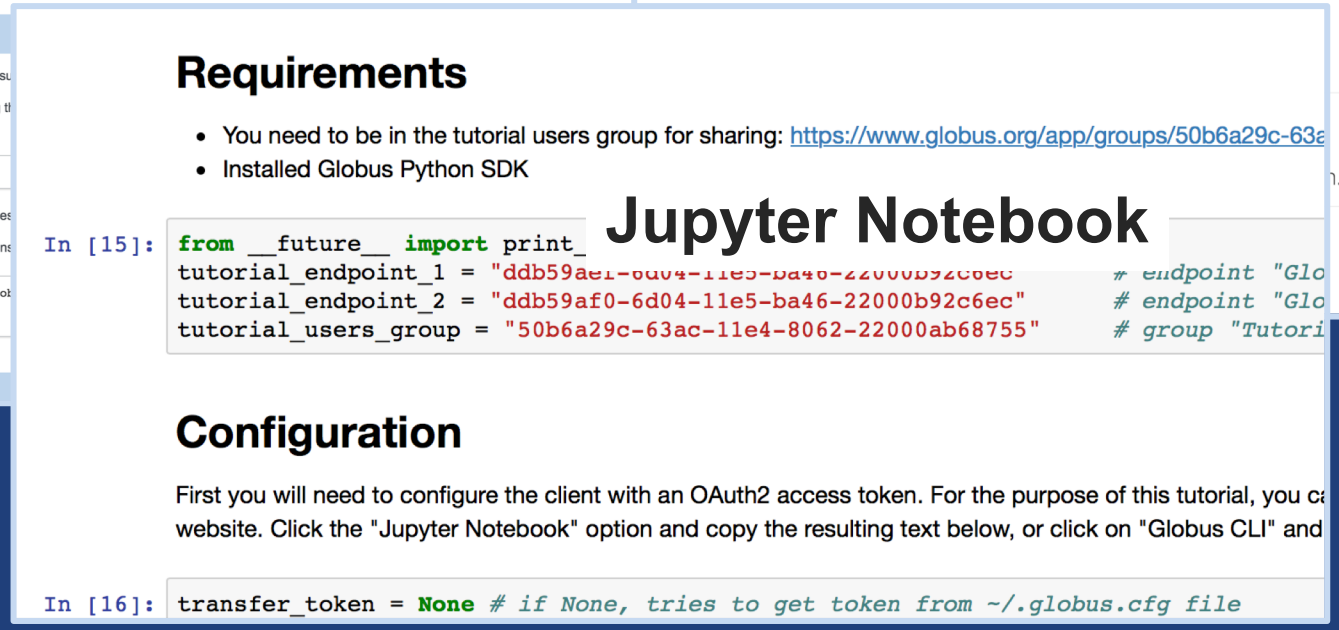
Globus PaaS developer resources



Python SDK



Sample Application



Jupyter Notebook

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**