Globus Introduction

Steve Tuecke
tuecke@globus.org

Thank you to our sponsors!

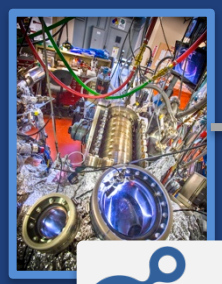
U.S. DEPARTMENT OF
ENERGY

THE UNIVERSITY OF
CHICAGO

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

Argonne
NATIONAL LABORATORY

Globus and the research data lifecycle



Globus transfers files reliably, securely

2

Transfer

4 Globus controls access to shared files on existing storage; no need to move files to cloud storage!

7 Curator reviews and approves; data set published on campus or other system

1

Researcher initiates transfer request; or requested automatically by script, science gateway

3

Share

Researcher selects files to share, selects user or group, and sets access permissions

6

Researcher assembles data set; describes it using metadata (Dublin core and domain-specific)

Publication Repository

5

Collaborator logs in to Globus and accesses shared files; no local account required; download via Globus

Personal Computer

Publish

6

8

Peers, collaborators search and discover datasets; transfer and share using Globus

Discover

- **Only a Web browser required**
- **Use storage system of your choice**
- **Access using your campus credentials**

Benefits of SaaS

- **Users do not need to deploy software**
  - Easy to access via Web browser
  - Command line, REST interfaces for flexible automation and integration
- **New features automatically available**
- **Reduced IT operational costs**
  - Small local footprint (Globus Connect)
  - Consolidated support and troubleshooting

Globus by the numbers

4 major services	**180 PB** transferred	**20 billion** files processed	**31,000** registered users
13 national labs use Globus	**3,000** active endpoints per month	**~450** active daily users	**99.9%** uptime
35+ institutional subscribers	**1 PB** largest single transfer to date	**3 months** longest continuously managed transfer	**130** federated campus identities

Demo

- **Login using existing identity**
- **Identity linking**
- **Endpoint search**
- **Transfer**
- **Globus Connect Personal**
- **Bookmarks**
- **Sharing**
- **Groups**
- **Management console**

Storage connectors

- **Standard storage connectors (Posix)**
  - Linux, Windows, MacOS
  - Lustre, GPFS, OrangeFS, etc.
- **Premium storage connectors**
  - HPSS
  - HDFS
  - S3
  - Ceph RadosGW (S3 API)
  - Spectra Logic BlackPearl
  - Google Drive

New Globus CLI
(coming soon)

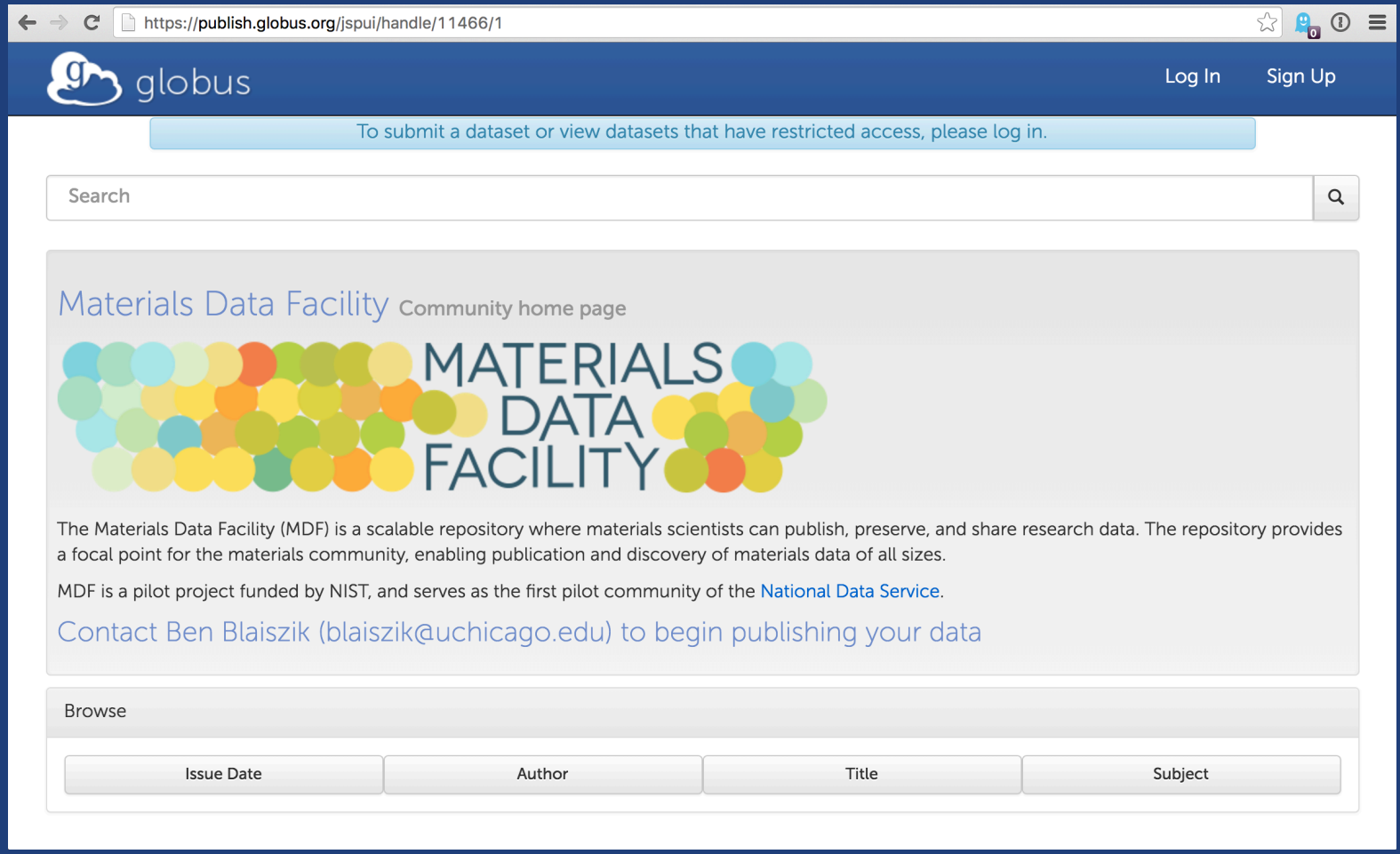
```
$ globus transfer
Usage: globus transfer [OPTIONS] COMMAND [ARGS]...

  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
```

Data Publication and Discovery



Peer reviewed paper data

(Re)format…

- PDF/A
- HDF
- …

Fully described…

- Dublin core metadata
- Domain metadata
- Provenance info

Replicated, public repositories

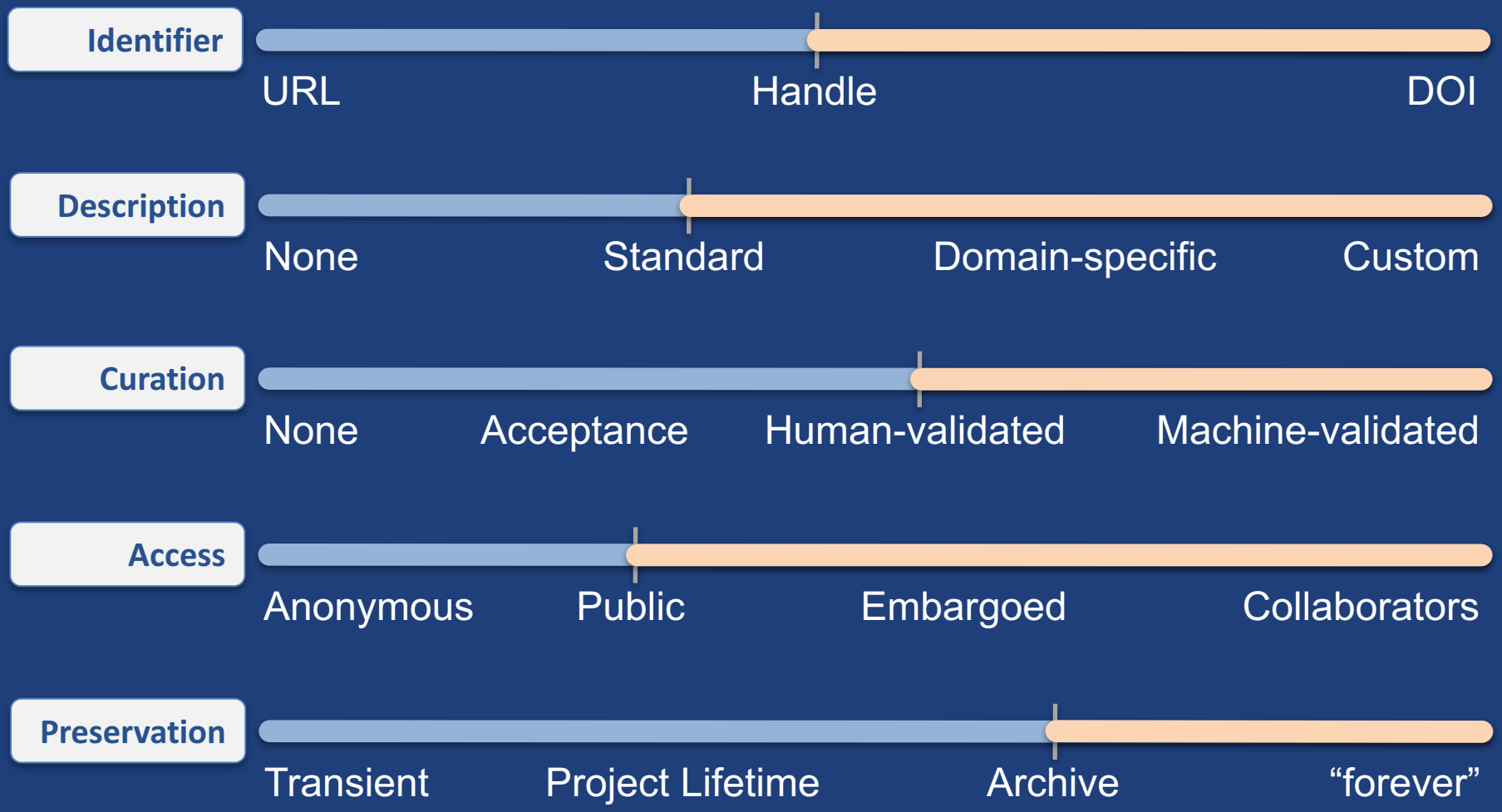
Persistent identifier

- DOI

Formal, multi-step review

- Review → Update → Resubmit cycle

Globus collection policies



Raw NGS output

Minimal metadata…

- Source environment
  - Instrument, timestamp,…
- Unique ID

No curation

- Automated dataset acceptance

Identify…

- Handle

High durability, low cost store

Glacier

Analysis results

XSEDE
Extreme Science and Engineering Discovery Environment
EC2
Jetstream
NERSC

Optional metadata…

- "Implicit" metadata
- Description through organization

Team review

- Any collaborator may approve

Identify…

- Globus share

Widely accessible stores

amazon web services S3
red cloud
Jetstream

Demo

- **Publication**
- **Discovery**

“A single global information space”

HTTPS access to endpoints

Making
research data storage
a first-class entity on the web

HTTPS access to endpoints

(coming soon)

- **Enhanced use of research storage**
  - Asynchronous, bulk transfer via GridFTP
  - (soon) Synchronous remote access via HTTPS
- **Enhanced Globus web app**
  - Browser-based upload/download
  - Inline file viewer
- **Integration with clients and web apps**

Cloud has transformed how software and platforms are delivered

Software as a service: **SaaS**

(web & mobile apps)

NETFLIX

Platform as a service: **PaaS**

Infrastructure as a service: **IaaS**

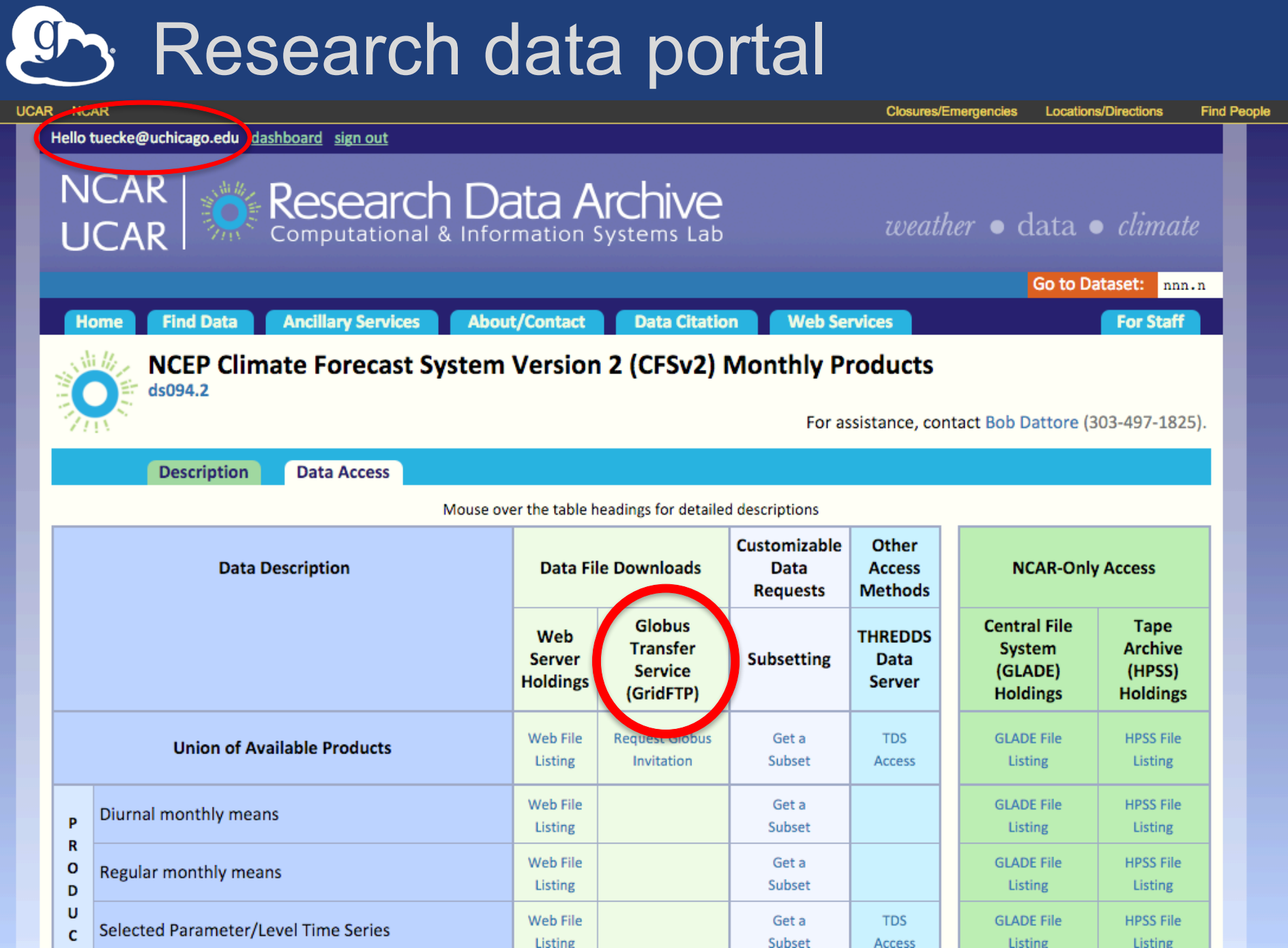
PaaS enables more rapid, cheap, and scalable delivery of powerful (SaaS) apps

Platform Questions

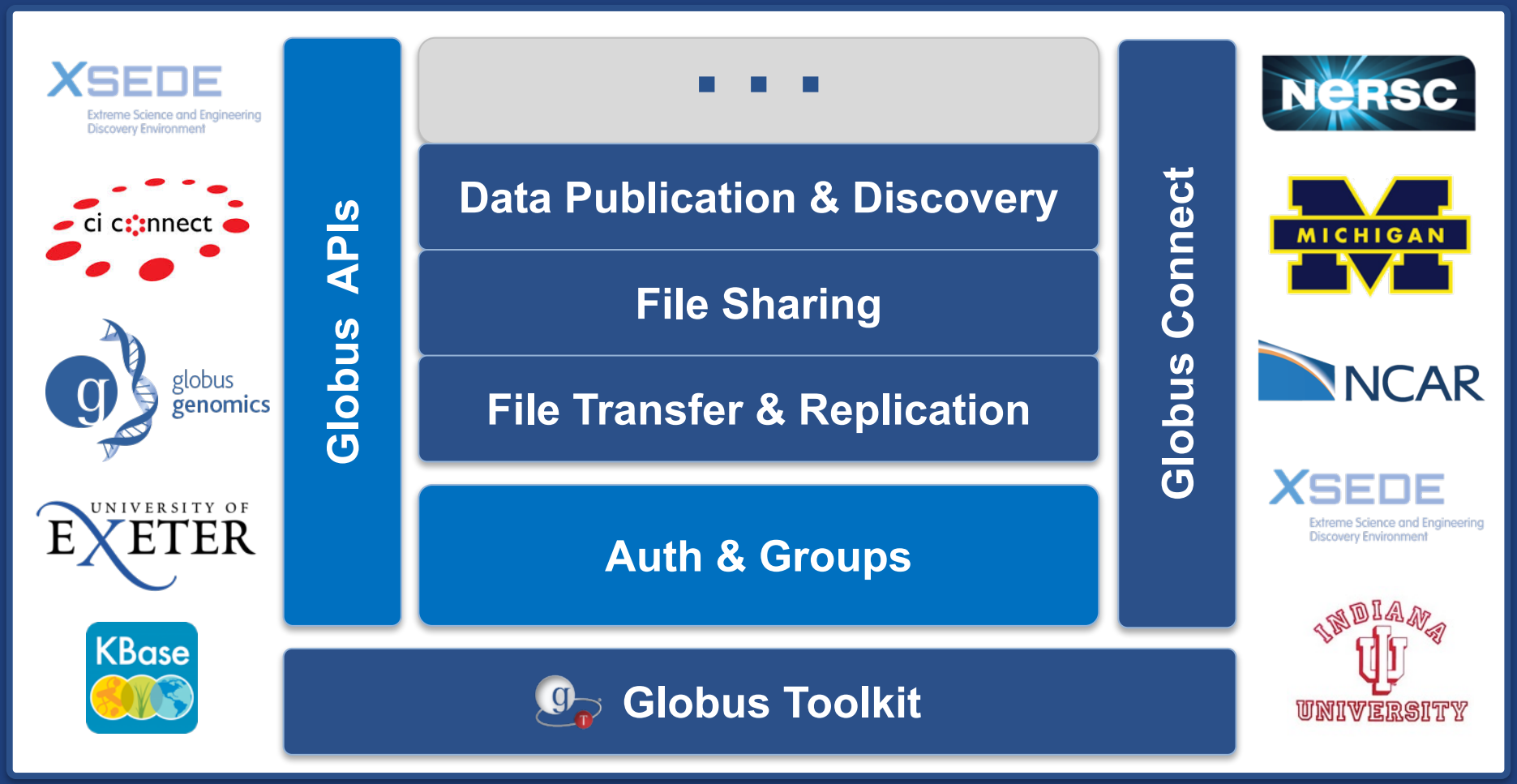
- **How do you leverage Globus services in your own applications?**
- **How do you extend Globus with your own services?**

- **How do we empower the research community to create an integrated ecosystem of services and applications?**

Research data portal



Globus PaaS


XSEDE
Extreme Science and Engineering Discovery Environment
ci connect
globus genomics
UNIVERSITY OF EXETER
KBase
Globus APIs
. . .
Data Publication & Discovery
File Sharing
File Transfer & Replication
Auth & Groups
Globus Toolkit
Globus Connect
NERSC
MICHIGAN
NCAR
XSEDE
Extreme Science and Engineering Discovery Environment
INDIANA UNIVERSITY

docs.globus.org

https://docs.globus.org

globus docs

APIs How To Guides Support

Harness the power of the Globus research data management cloud.

Transfer API

The Transfer API provides a REST-style interface to the Globus reliable file transfer service. The API can be used to monitor the progress of file transfers, manage file transfer endpoints, list remote directories, and submit new transfer and delete tasks.

LEARN MORE

Resource Providers

Globus allows you, as a resource provider, to easily offer reliable, secure, high-performance research data management capabilities to your users and their collaborators, directly from your own storage infrastructure.

LEARN MORE

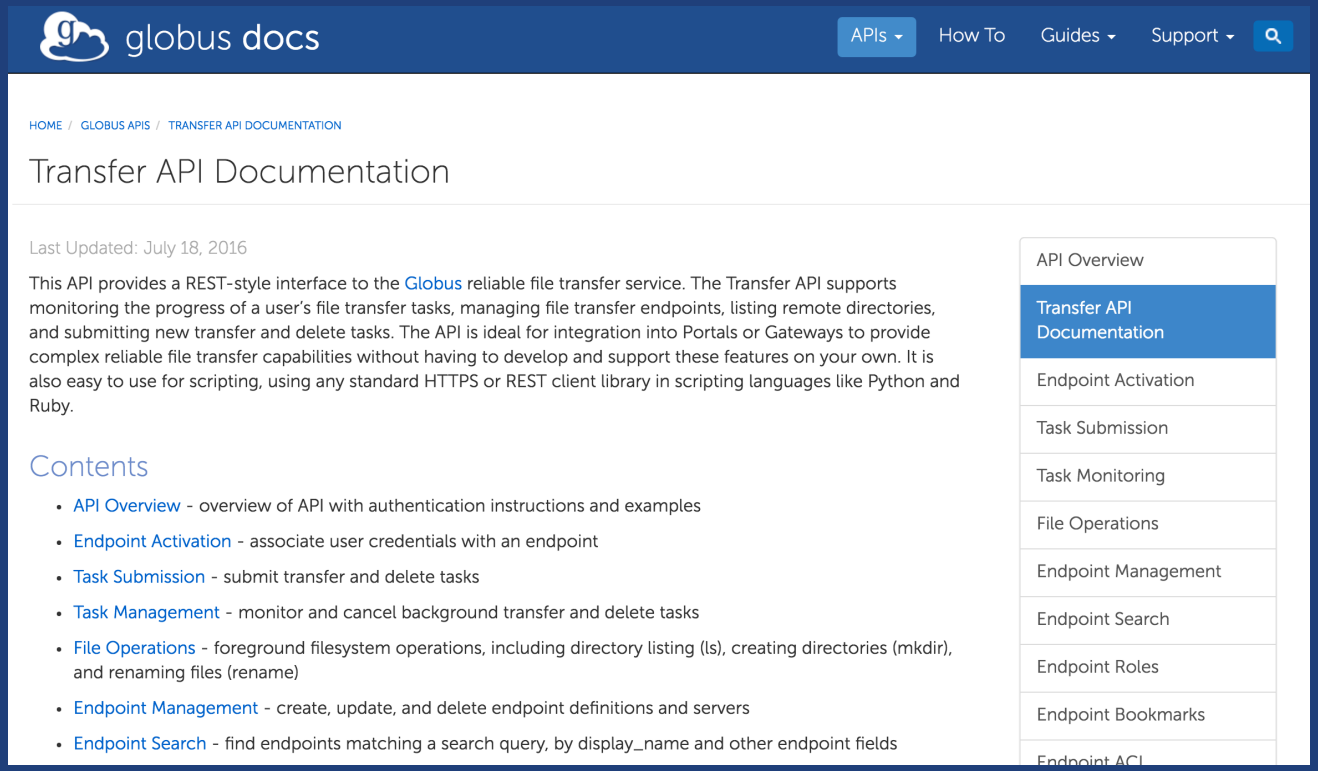
Toolkit

The open source Globus Toolkit is a fundamental enabling technology for the "Grid," allowing users to access high-performance computing resources securely across corporate, institutional, and geographic boundaries without sacrificing local autonomy.

LEARN MORE

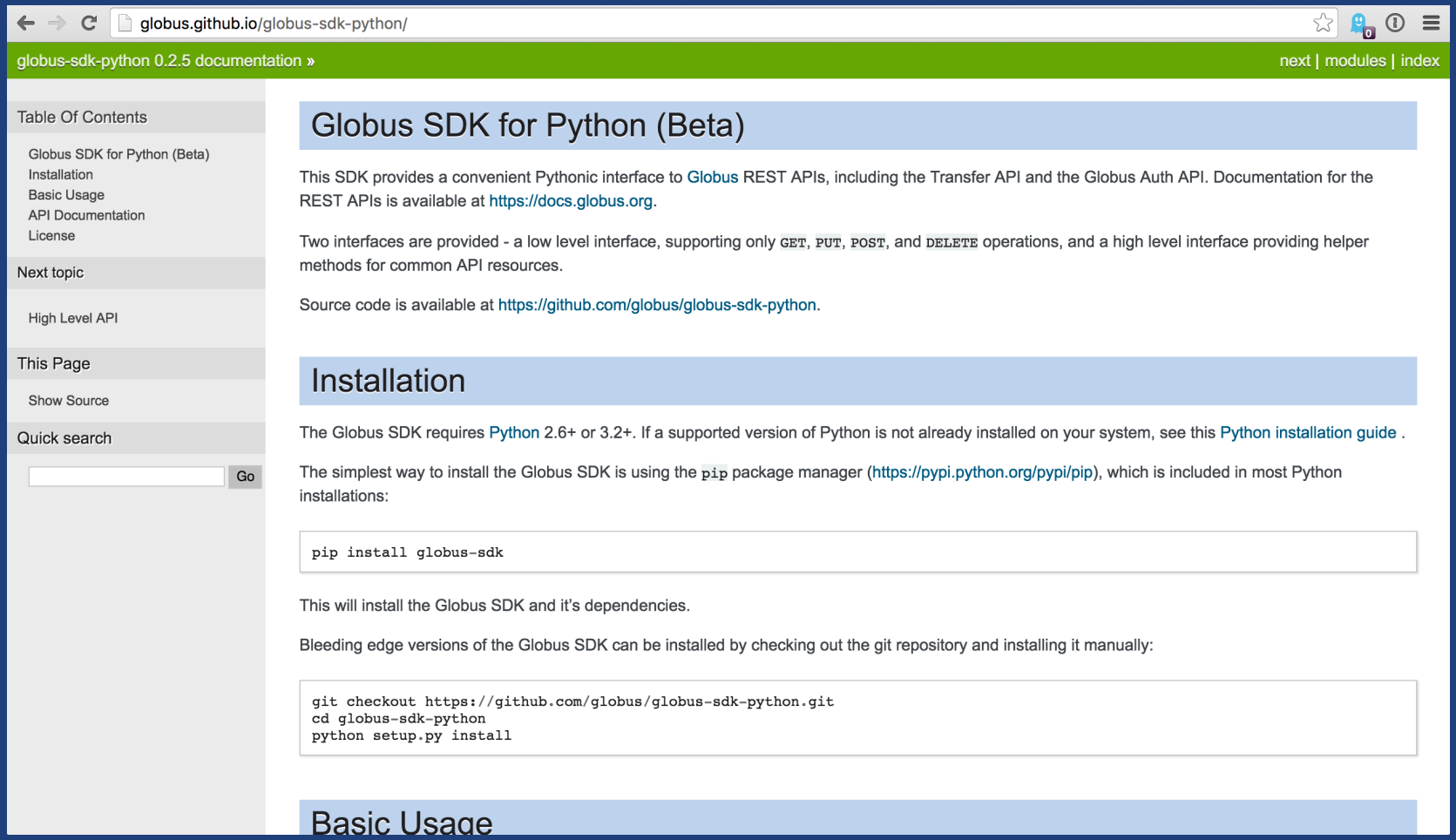
Globus Transfer API

Nearly all Globus Web App functionality implemented via public Transfer API



https://docs.globus.org/api/transfer/

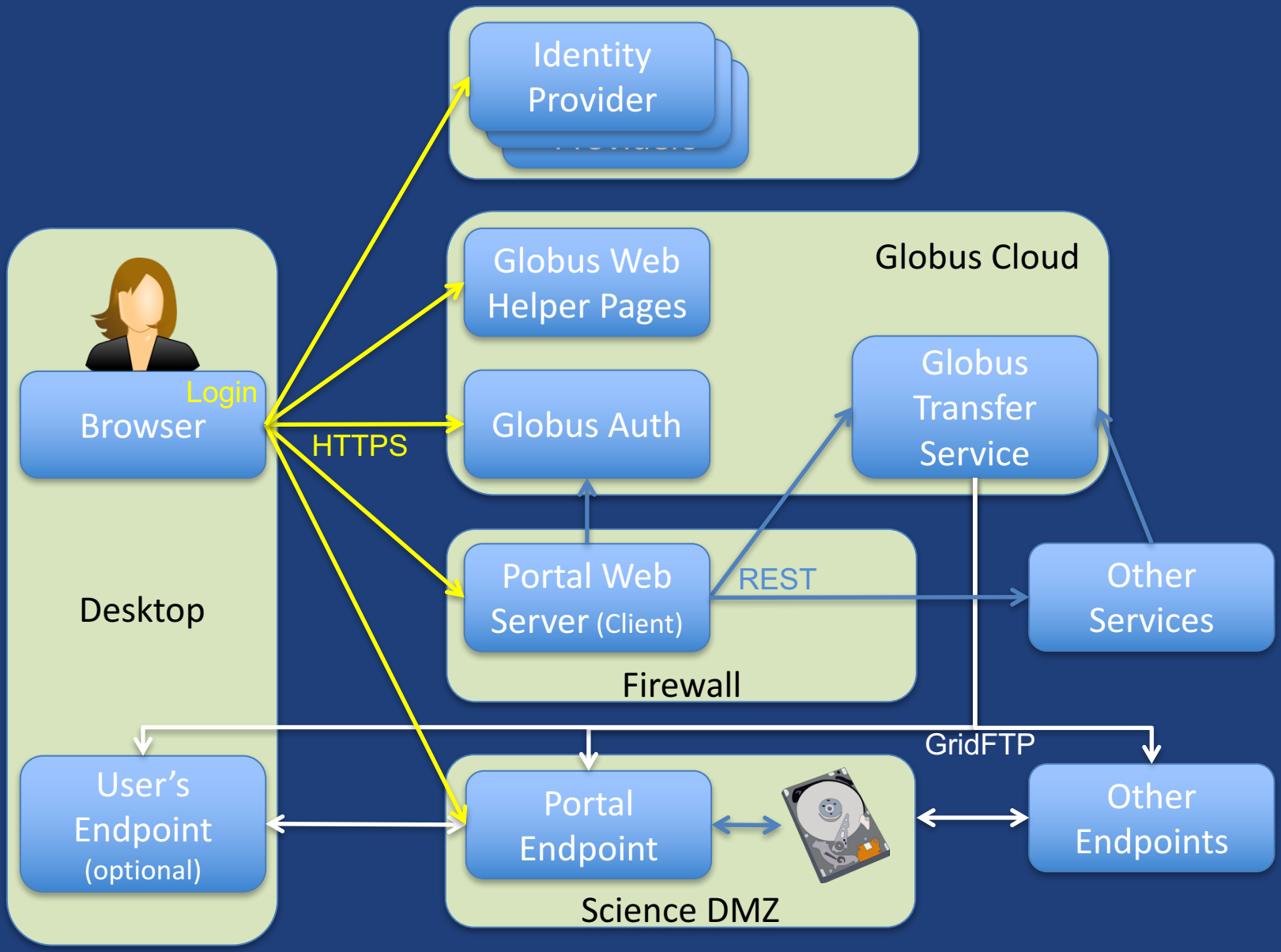
Globus SDK Python



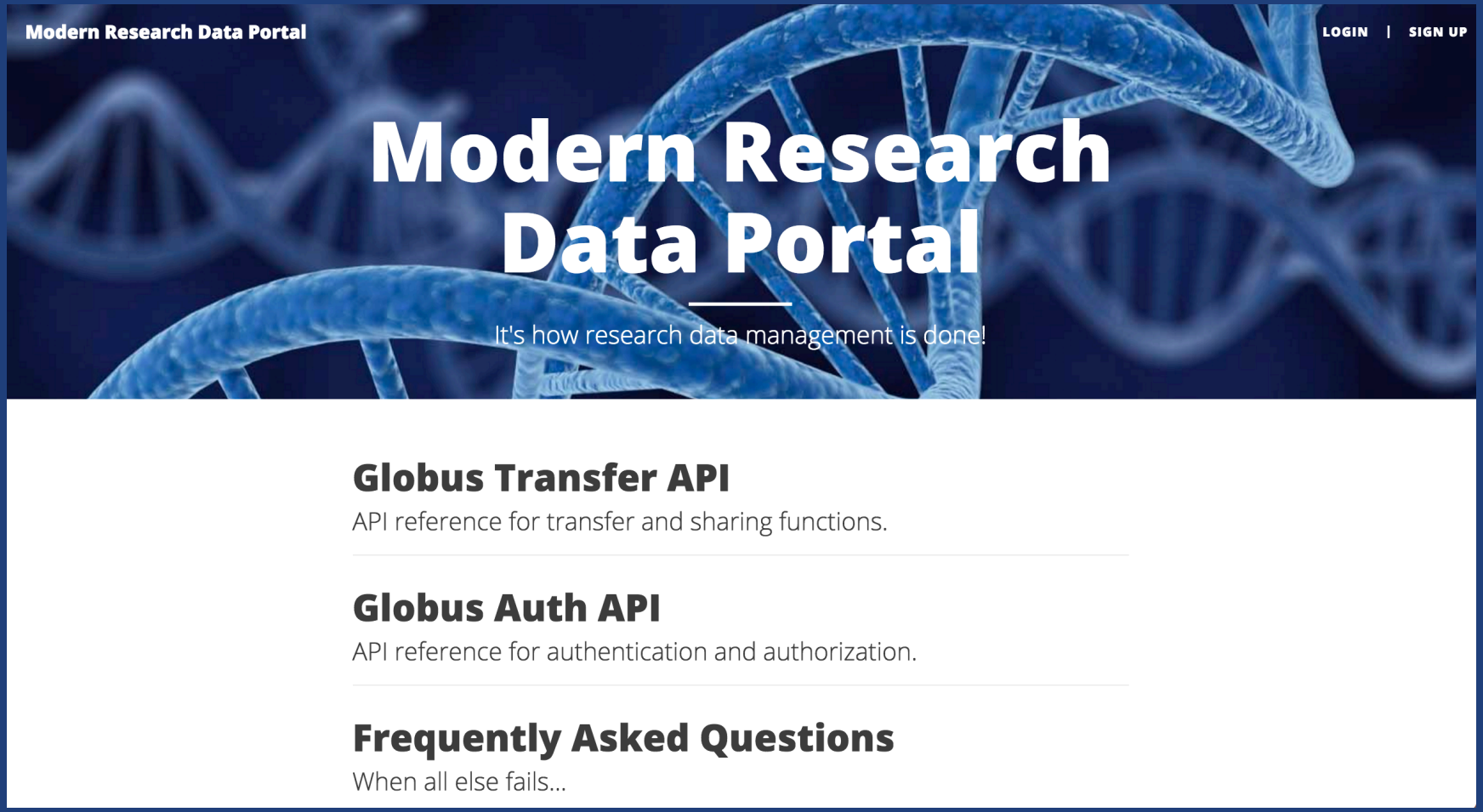
pip install globus-sdk-python

https://github.com/globus/globus-sdk-python

Prototypical research data portal



Sample python data portal



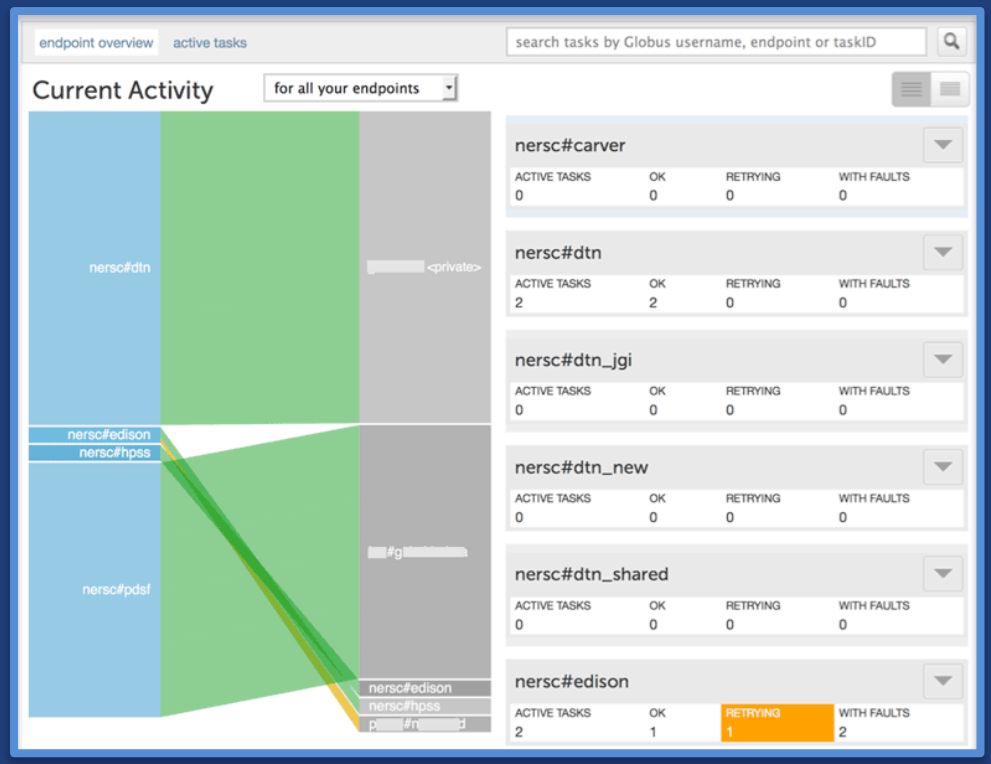
https://github.com/globus/globus-sample-data-portal

GlobusWorld Developer Workshops

https://www.globusworld.org/tour/

Globus Subscriptions

- **Globus Subscription**
  - Shared endpoints
  - Data publication
  - Management console
  - Usage reporting
  - Priority support
  - Application integration
  - HTTPS endpoints
- **Branded Web Site**
- **Alternate Identity Provider (InCommon is standard)**
- **Premium Storage Connectors**



globus.org/subscriptions

Summary

- Globus SaaS delivers advanced capabilities for researchers and research computing centers

- Globus PaaS opens new opportunities for developers creating applications for researchers