Introduction to Research Data Management Using Globus

Rick Wagner
rick@globus.org

La Jolla, CA — August 4, 2017

globus

Research data management today

Index?

How do we...

...move?

...share?

...discover?

...reproduce?

Globus delivers…

Data transfer, sharing, publication, and discovery…

…directly from your own storage systems…

…via software-as-a-service

Globus enables...

Campus Bridging

…within and beyond campus boundaries

Bridge to campus HPC

Move datasets to campus research computing center

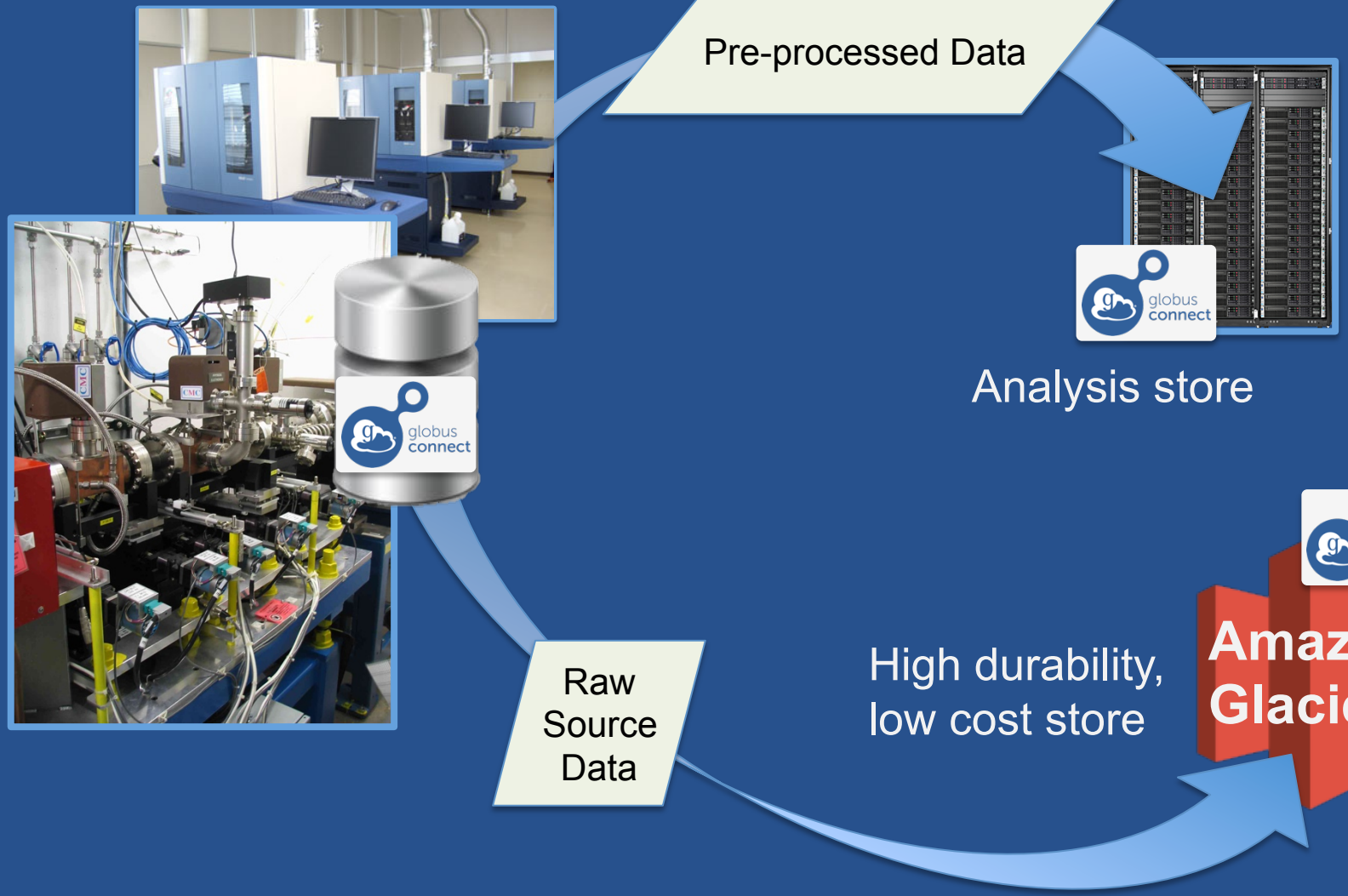
Move results to laptop, department, lab, …

Bridge to national cyberinfrastructure

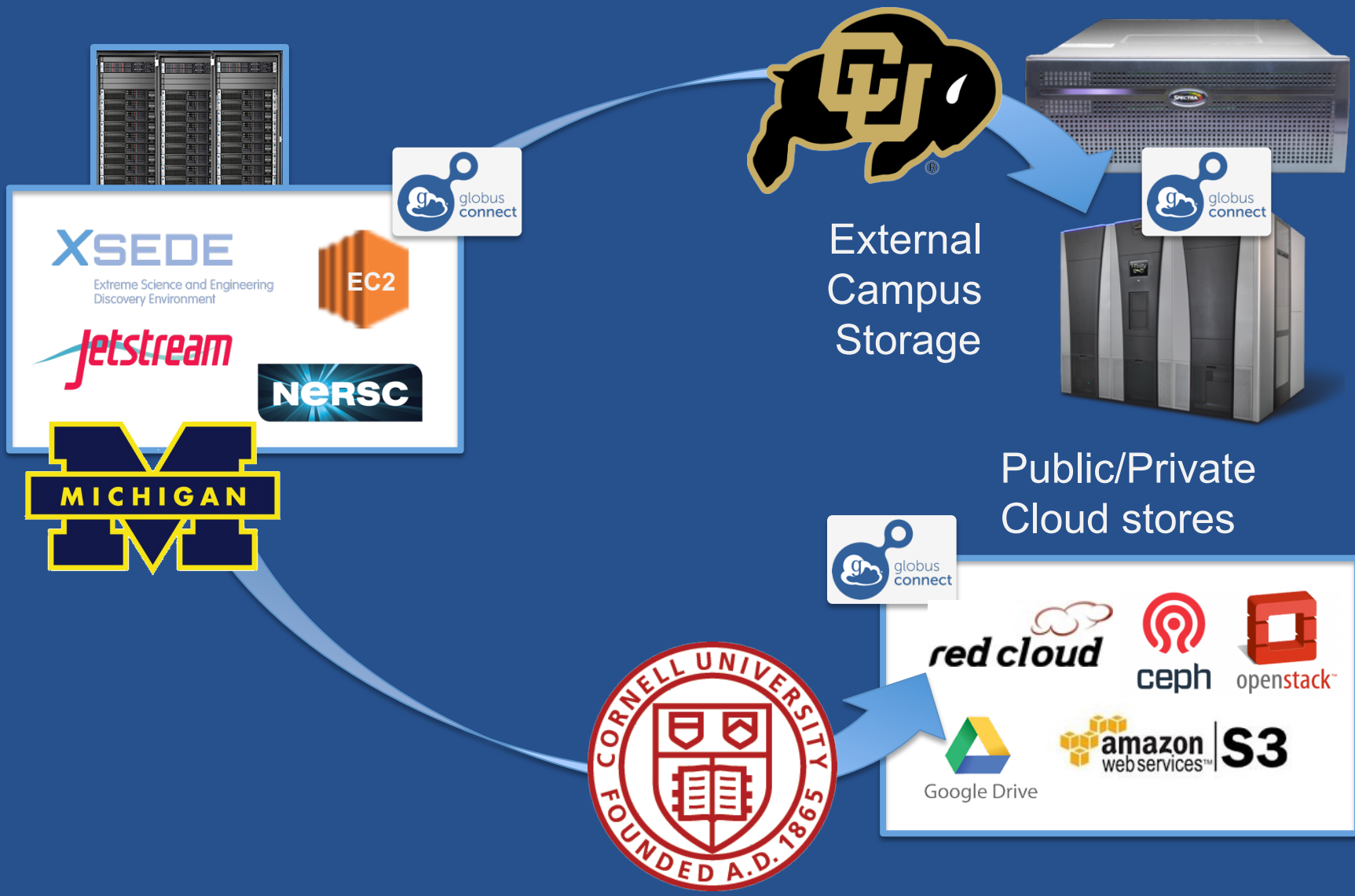
Move datasets to supercomputer, national facility

Move results to campus (…)

Bridge to instruments



Bridge to collaborators



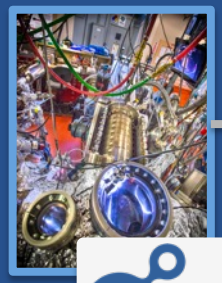
Bridge to community/public

Project Repositories,
Replication Stores

Public Repositories

Globus and the research data lifecycle

Instrument



Compute Facility

Globus transfers files reliably, securely

2

Transfer

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

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

5

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

Personal Computer

6

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

Publish

6

Publication Repository

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

8

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

Discover

- Access via Web browser
- Use any storage system
- Use existing identity

Why use Globus?

- **Simplicity**
  - Consistent UI across systems
  - Easy access to collaborators
- **Reliability and performance**
  - “Fire-and-forget” file transfer
  - Maximized WAN throughput
- **Operational efficiency**
  - Low overhead SaaS model
  - Highly automatable: CLI, RESTful API
- **Access to a large and growing community**

Demonstration

File Transfer

How can I use Globus on my system?

Globus Connect…

Makes your storage system a Globus endpoint

Globus Connect Personal



- **Installers do not require admin access**
- **Zero configuration; auto updating**
- **Handles NATs**

Move files to/from your laptop

1. Go to: www.globus.org/login
2. Select your institution from the list and click “Continue”
3. Authenticate with your institution’s identity system
4. Install Globus Connect Personal
5. Move file(s) between your laptop and another endpoint (e.g. ESnet)

Demonstration

File Sharing

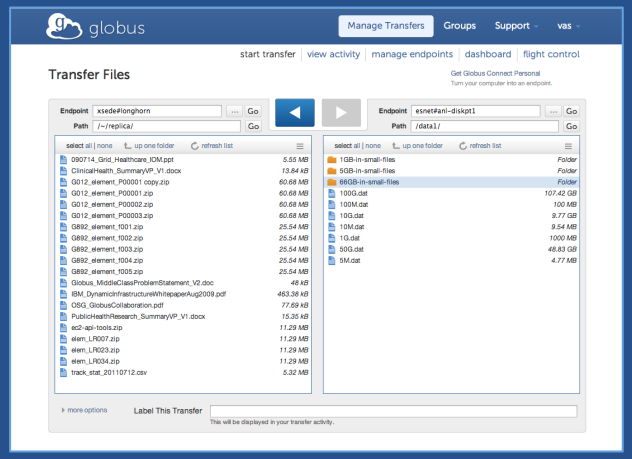
Federated Identity

How can I integrate Globus into my research workflows?

Globus serves as…

A platform for building science gateways, portals and other web applications in support of research and education

Use(r)-appropriate interfaces



Web

Globus service

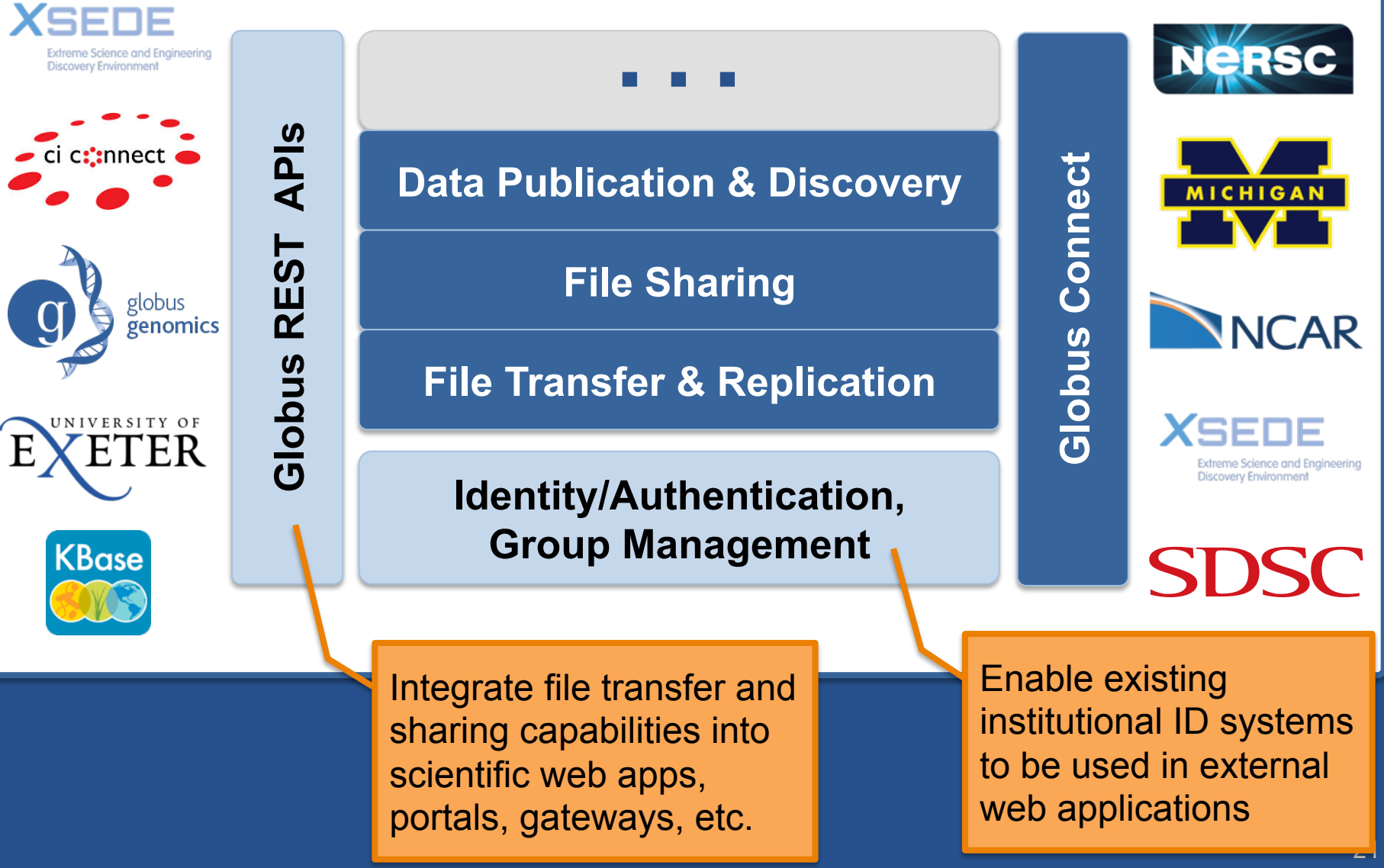
```
(globus-cli) jupiter:~ vas$ globus
Usage: globus [OPTIONS] COMMAND [ARGS]...

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

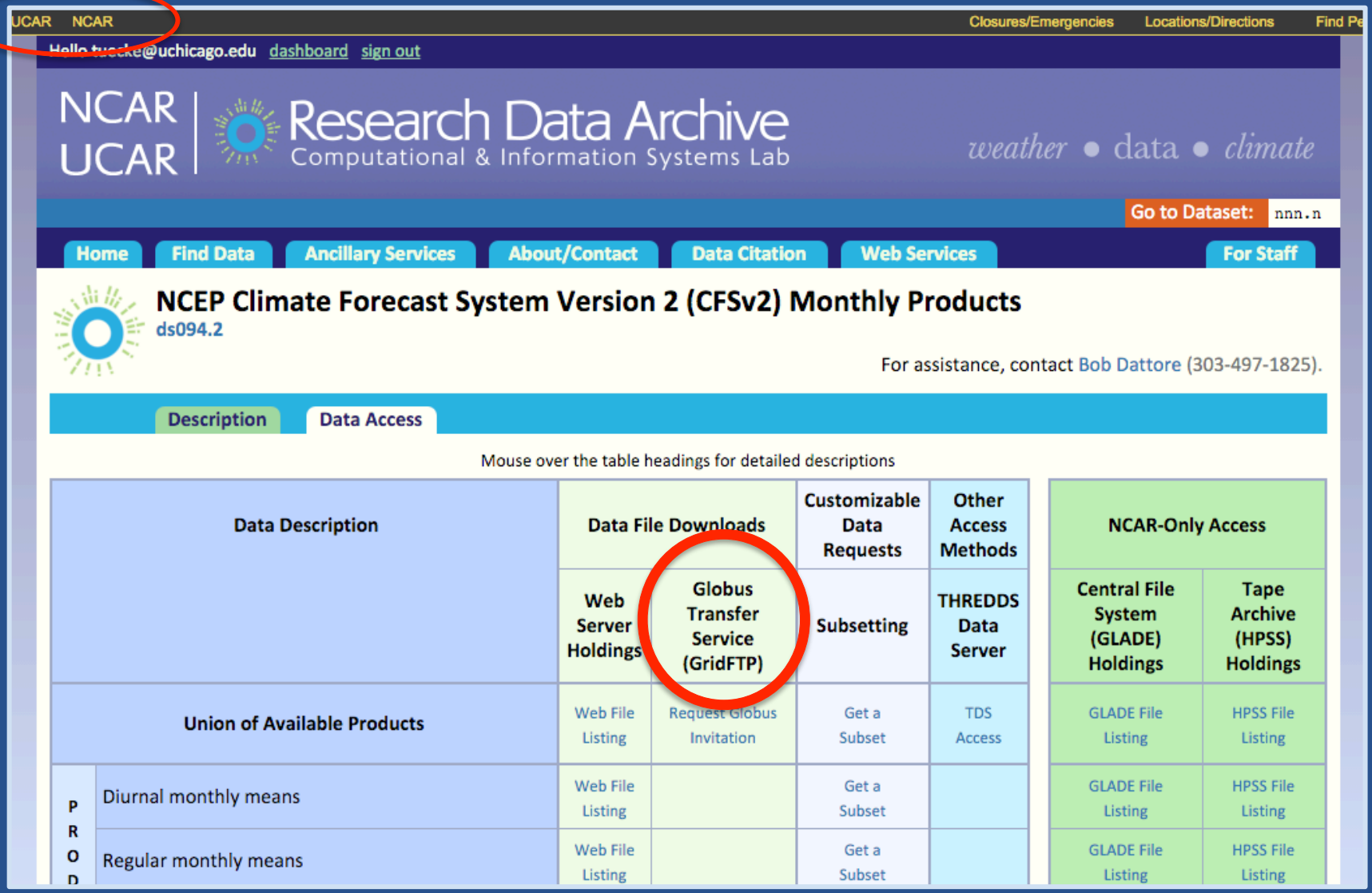
```
GET /endpoint/go%23ep1
PUT /endpoint/vas#my_endpt
200 OK
X-Transfer-API-Version: 0.10
Content-Type: application/json
…
```

Rest API

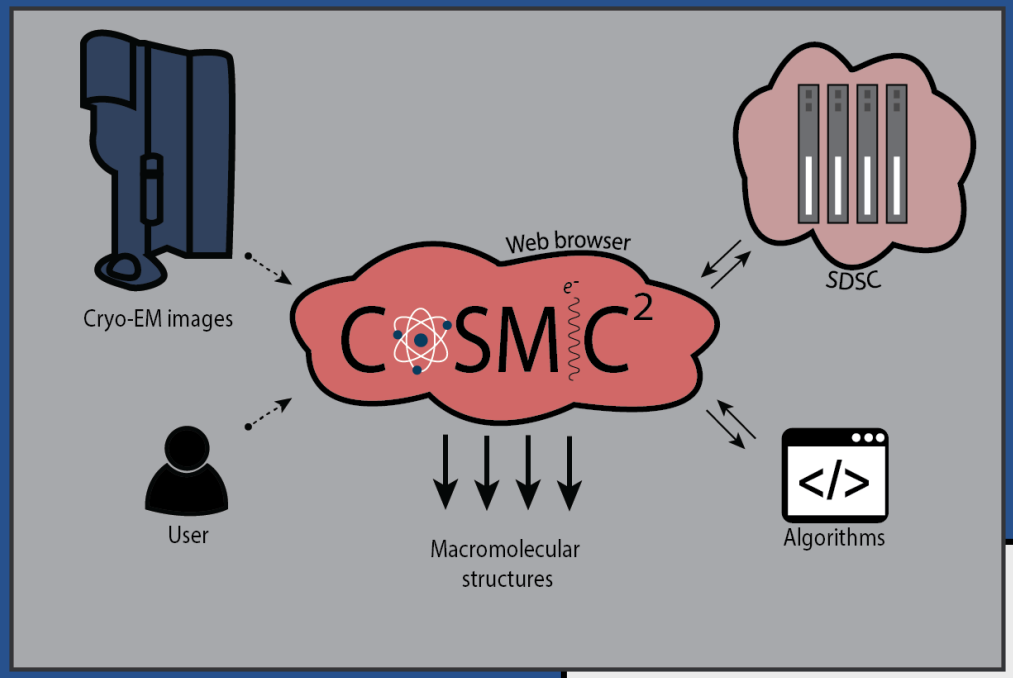
Globus as PaaS

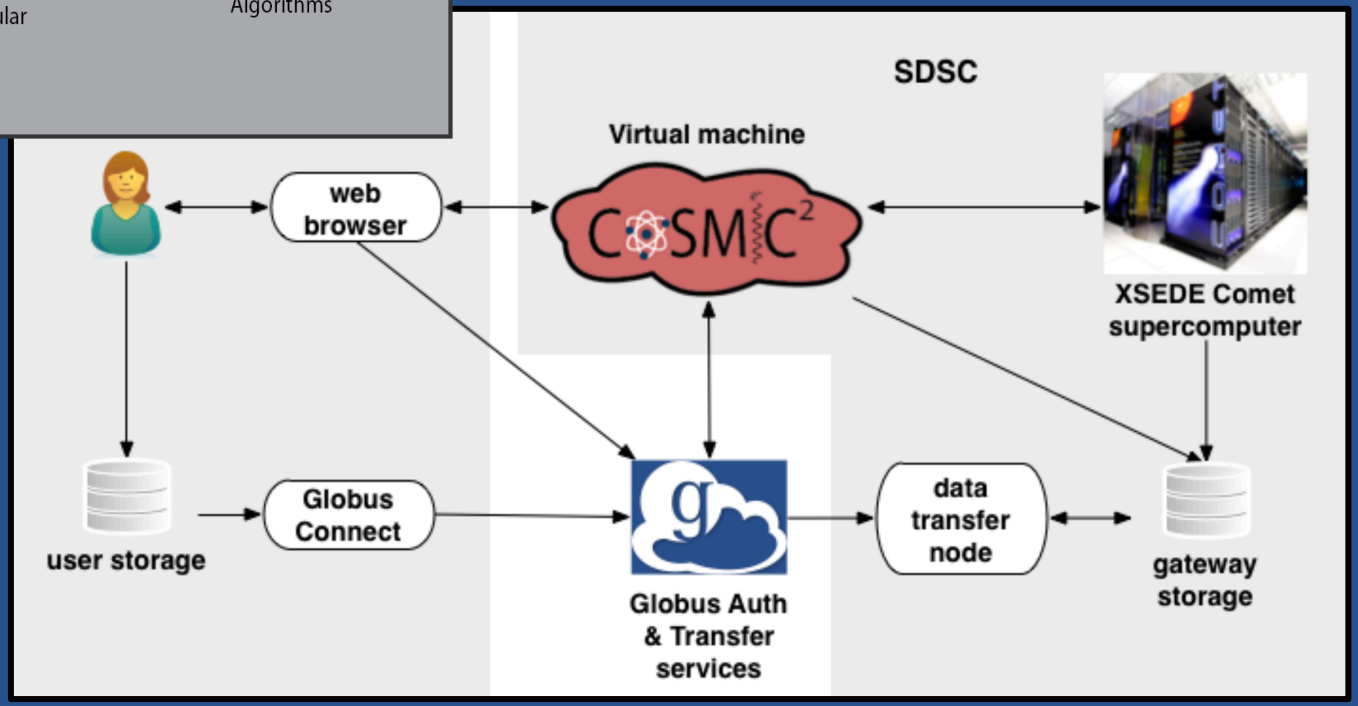


Data App: NCAR RDA

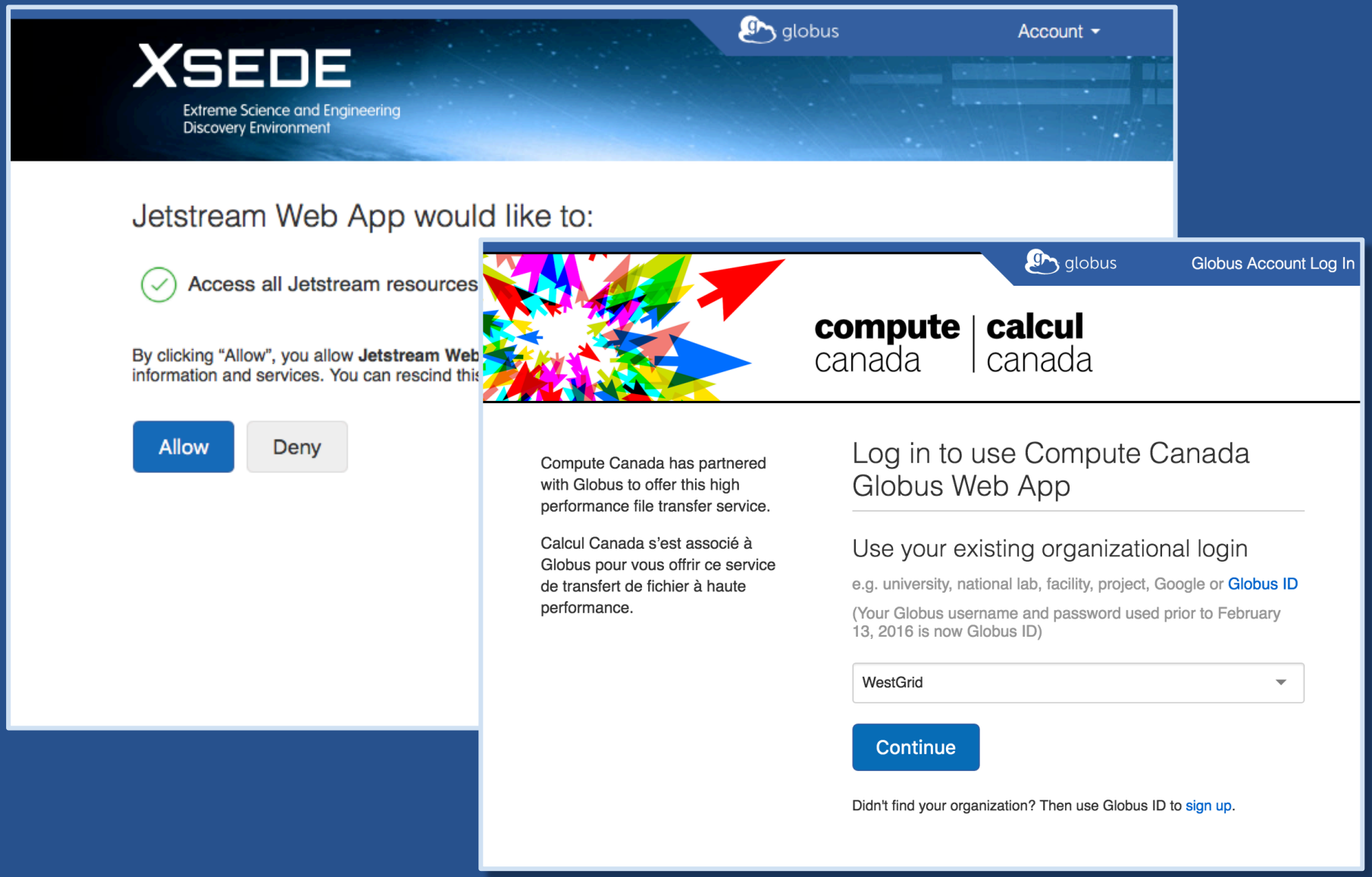


Gateway App: COSMIC² Cryo-EM

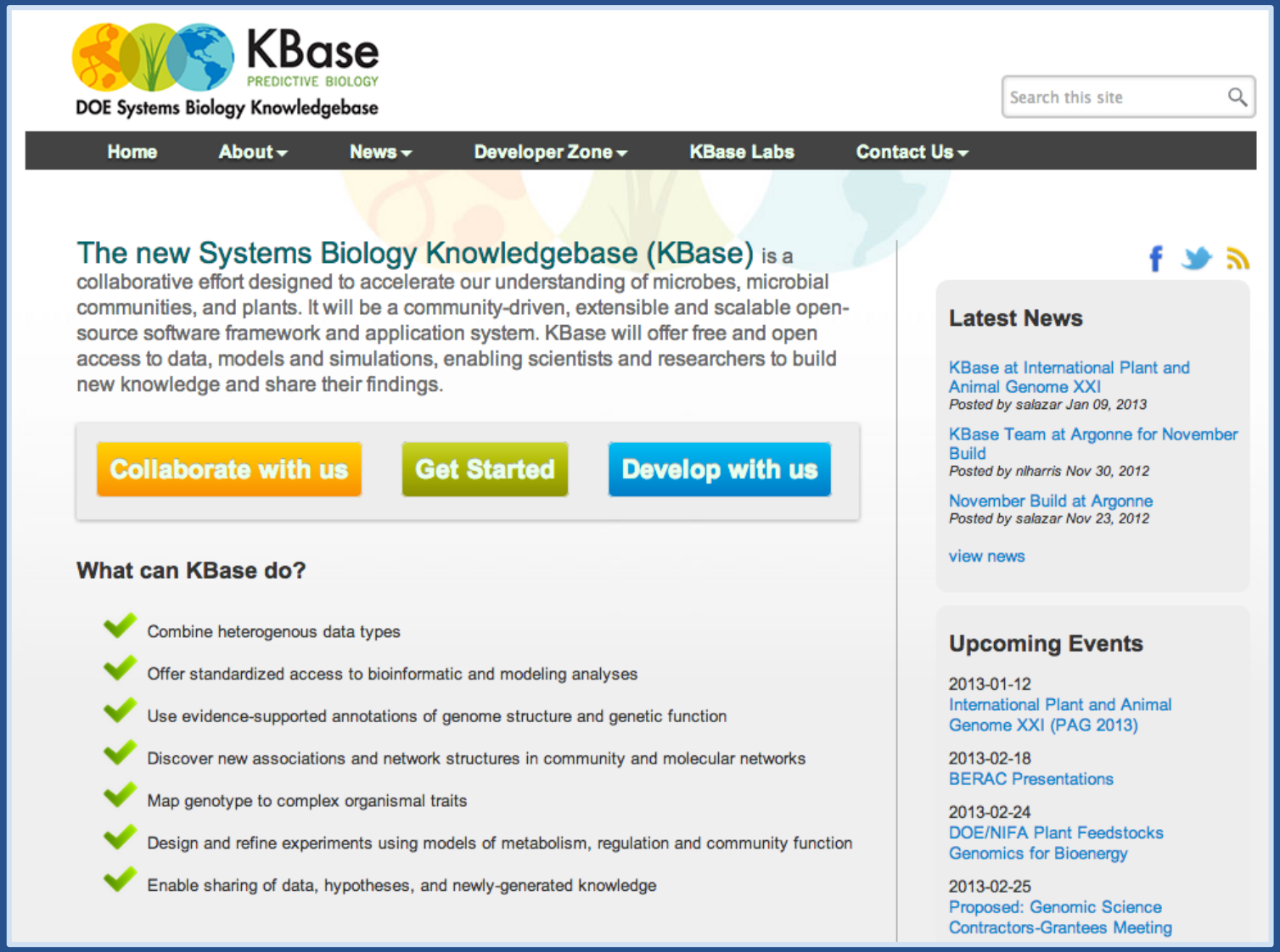




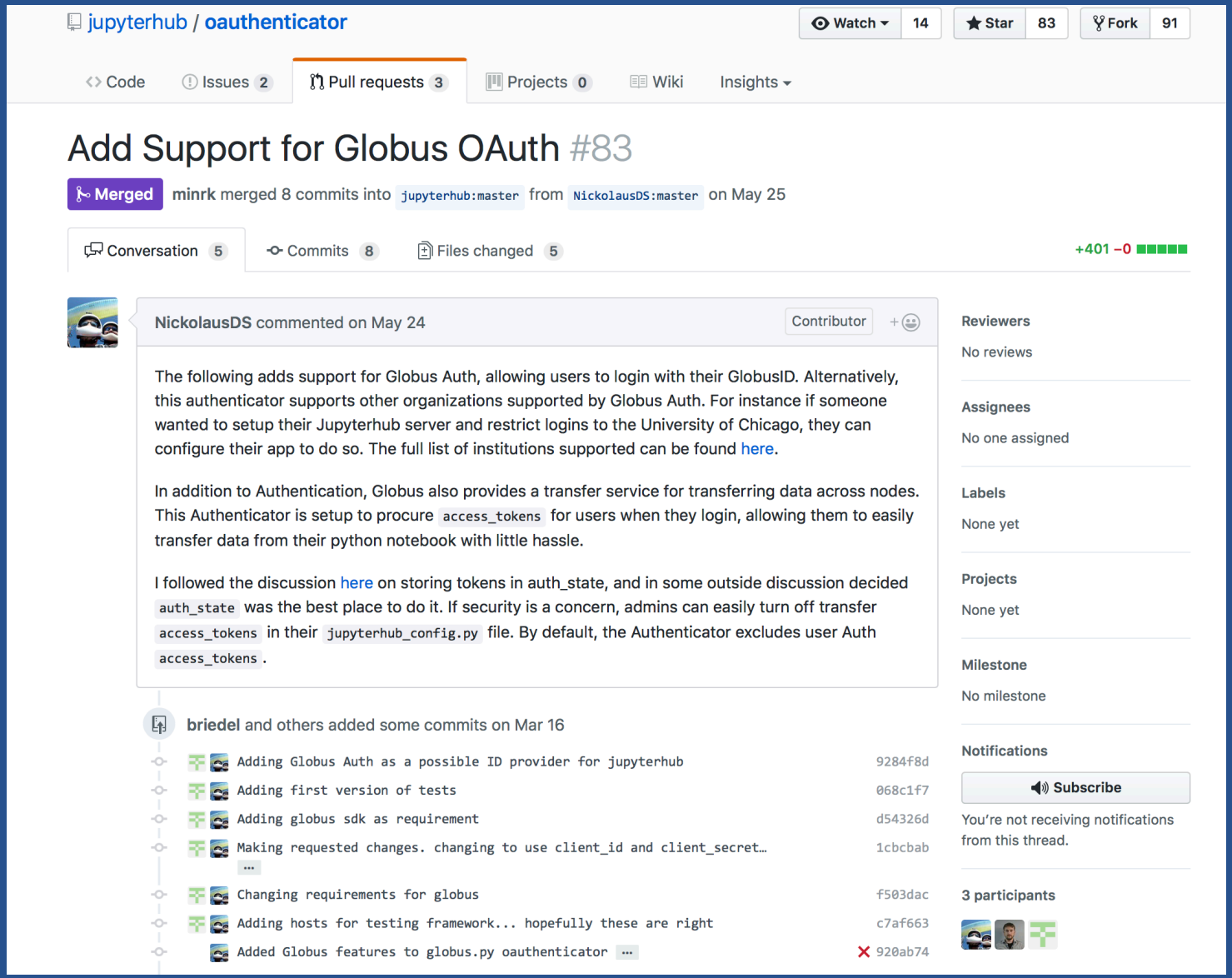
Globus PaaS: National Resource Access



Globus PaaS: Identity Management



Globus PaaS: JupyterHub



jupyterhub / oauthenticator

Watch 14 | Star 83 | Fork 91

Code | Issues 2 | Pull requests 3 | Projects 0 | Wiki | Insights

Add Support for Globus OAuth #83

Merged **minrk** merged 8 commits into `jupyterhub:master` from `NickolausDS:master` on May 25

Conversation 5 | Commits 8 | Files changed 5 | +401 −0

NickolausDS commented on May 24 — Contributor

The following adds support for Globus Auth, allowing users to login with their GlobusID. Alternatively, this authenticator supports other organizations supported by Globus Auth. For instance if someone wanted to setup their Jupyterhub server and restrict logins to the University of Chicago, they can configure their app to do so. The full list of institutions supported can be found here.

In addition to Authentication, Globus also provides a transfer service for transferring data across nodes. This Authenticator is setup to procure `access_tokens` for users when they login, allowing them to easily transfer data from their python notebook with little hassle.

I followed the discussion here on storing tokens in auth_state, and in some outside discussion decided `auth_state` was the best place to do it. If security is a concern, admins can easily turn off transfer `access_tokens` in their `jupyterhub_config.py` file. By default, the Authenticator excludes user Auth `access_tokens`.

briedel and others added some commits on Mar 16

Adding Globus Auth as a possible ID provider for jupyterhub	9284f8d
Adding first version of tests	068c1f7
Adding globus sdk as requirement	d54326d
Making requested changes. changing to use client_id and client_secret…	1cbcbab
Changing requirements for globus	f503dac
Adding hosts for testing framework... hopefully these are right	c7af663
Added Globus features to globus.py oauthenticator	920ab74

Reviewers
No reviews

Assignees
No one assigned

Labels
None yet

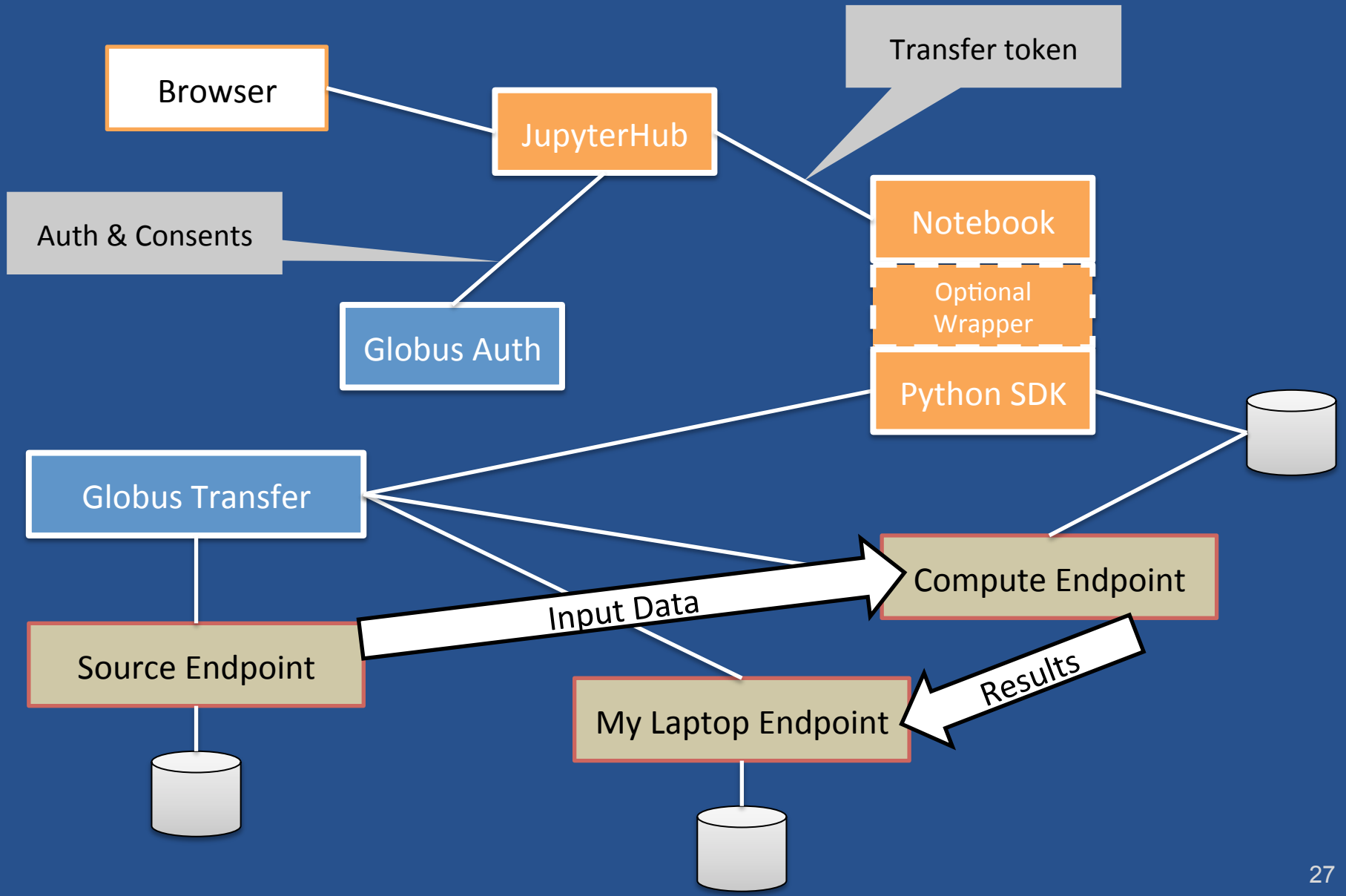
Projects
None yet

Milestone
No milestone

Notifications
Subscribe
You're not receiving notifications from this thread.

3 participants

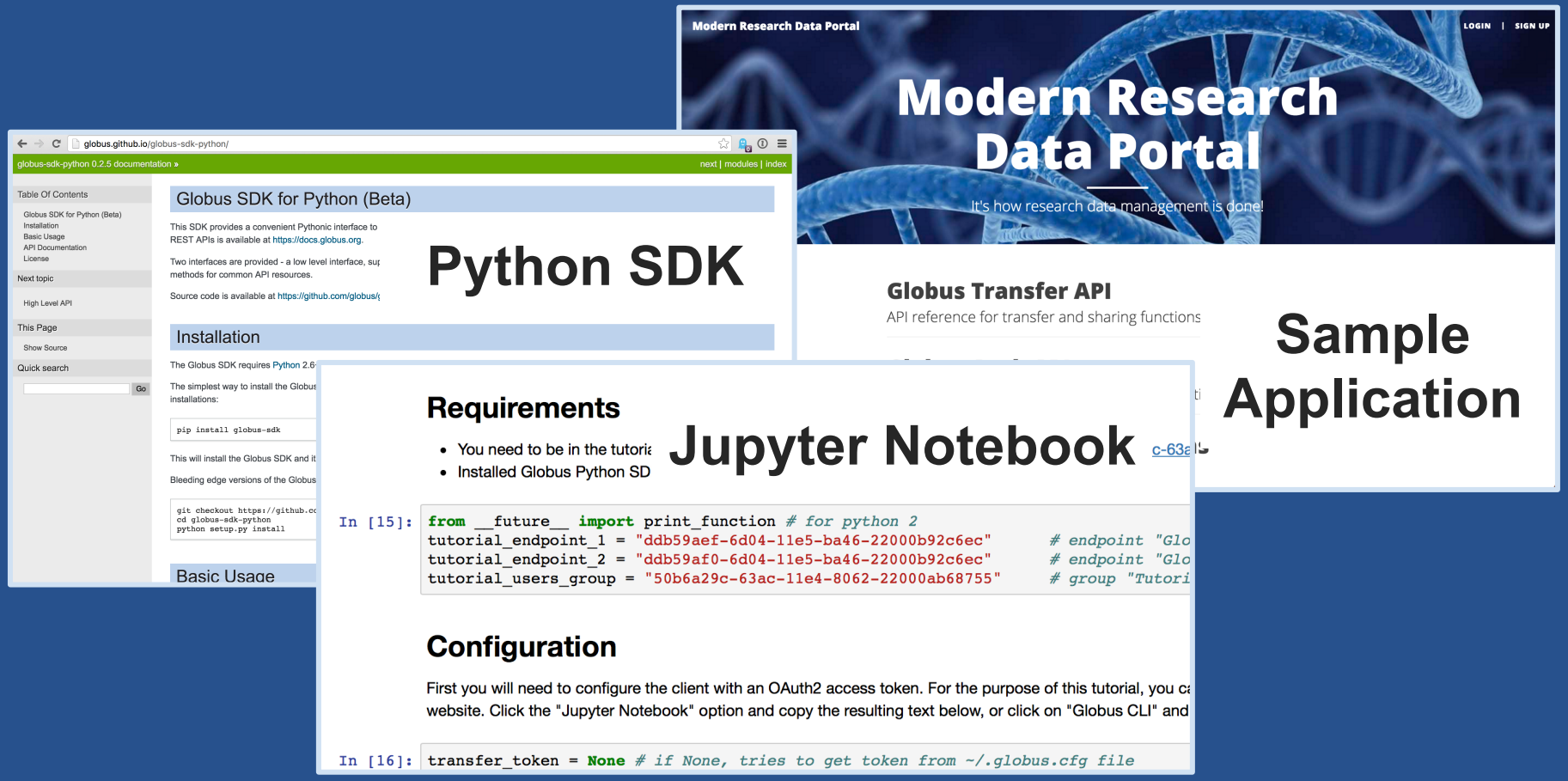
Globus PaaS: JupyterHub



Demonstration

JupyterHub

Globus PaaS developer resources



docs.globus.org/api **github.com/globus**

Thank you to our sponsors...

U.S. DEPARTMENT OF
ENERGY

THE UNIVERSITY OF
CHICAGO

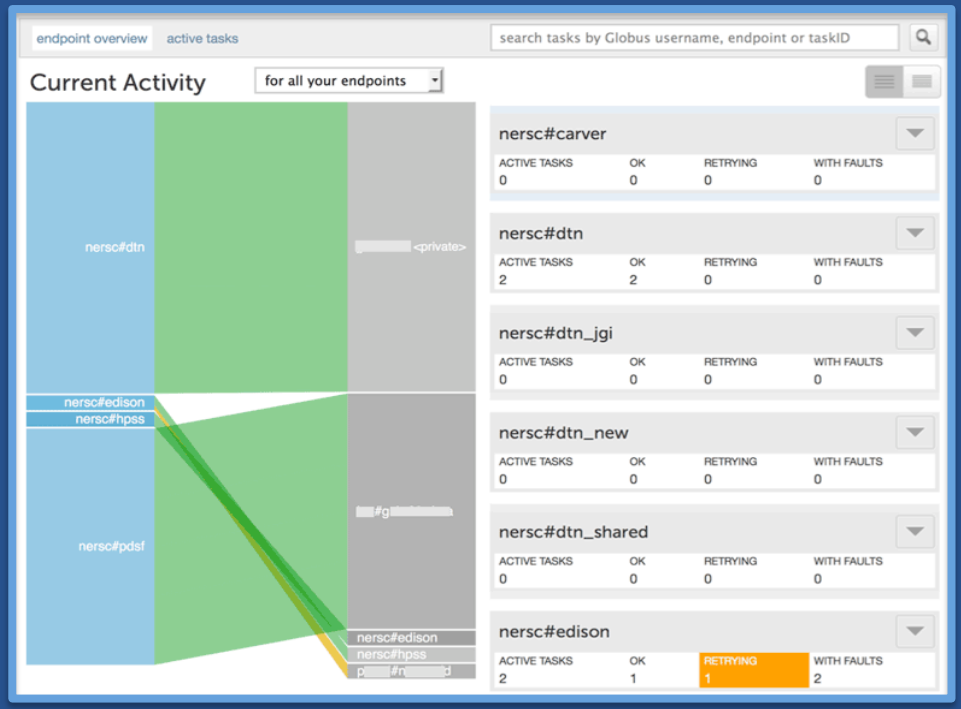
NIST
National Institute of
Standards and Technology
U.S. Department of Commerce

Argonne
NATIONAL LABORATORY

powered by
amazon
web services

Globus sustainability model

- **Standard Subscription**
  - Shared endpoints
  - Data publication
  - HTTPS support*
  - Management console
  - Usage reporting
  - Priority support
  - Application integration
- **Branded Web Site**
- **Premium Storage Connectors**
  - Amazon S3, Ceph, HPSS, Spectra, Google Drive, Box*, HDFS*
- **Alternate Identity Provider (InCommon is standard)**

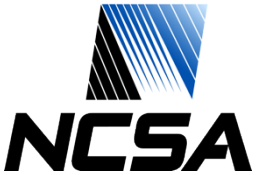


*Coming soon

Thank you to our users...

5 major services	**290PB** transferred	**47 Bn** files processed	**70,000** registered users
13 national labs use Globus	**10,000** active endpoints	**10,000** active users/year	**99.5%** uptime
65+ institutional subscribers	**1 PB** largest single transfer to date	**3 months** longest continuously managed transfer	**300+** federated campus identities

...and thank YOU, our subscribers!



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**