

# What is Globus?

Rick Wagner  
[rick@globus.org](mailto:rick@globus.org)

LUG18 – April 26, 2018





# Research data management today



How do we...  
...move?  
...share?  
...discover?  
...reproduce?

Index?





Globus delivers...

Secure, reliable, 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



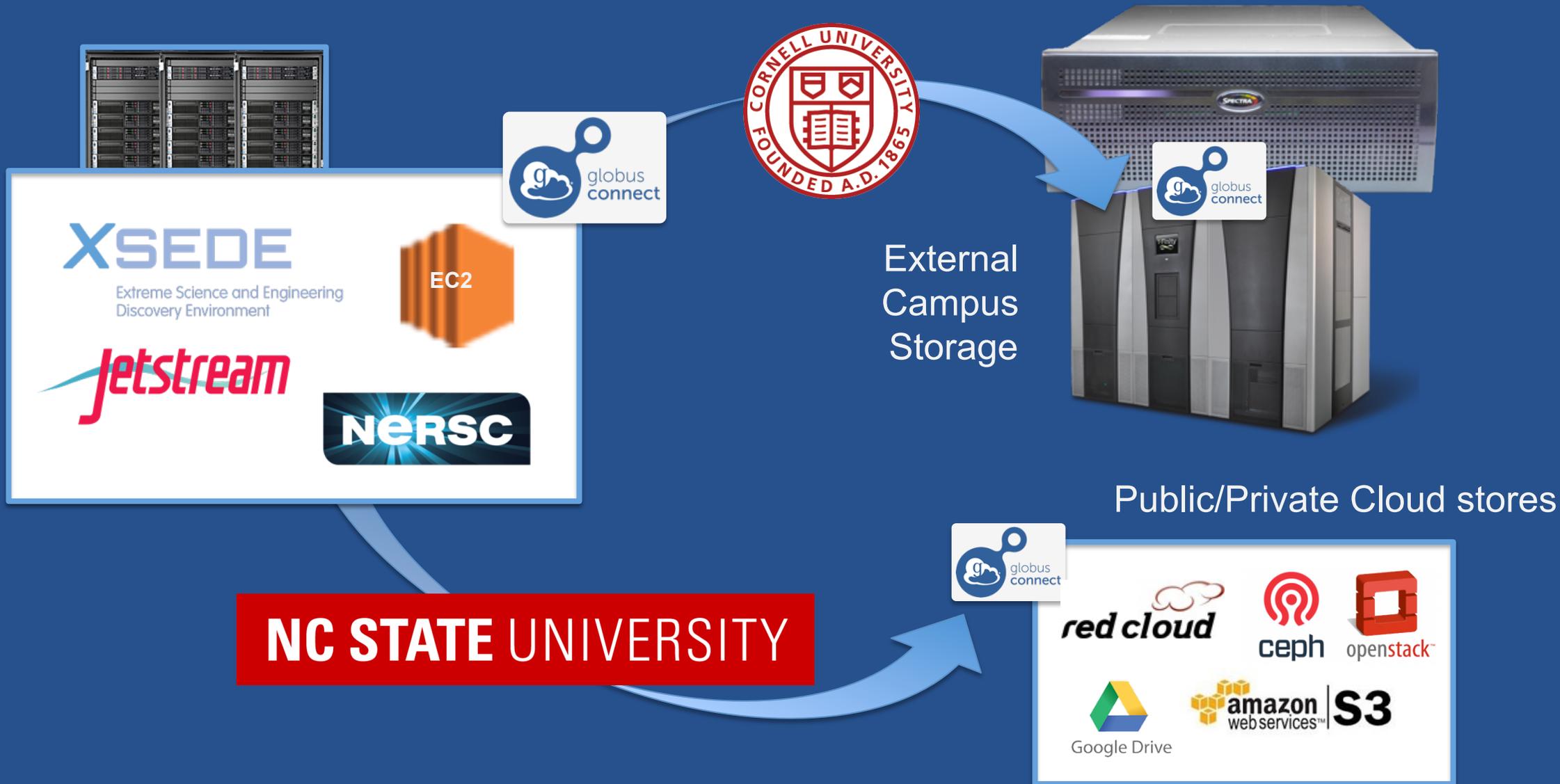
Move results to campus (...)

# Bridge to instruments





# Bridge to collaborators



# Bridge to community/public



Project Repositories,  
Replication Stores



Public Repositories



# Globus SaaS: Research data lifecycle

Instrument



Globus transfers files reliably, securely

2

Transfer

Compute Facility



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



Publication Repository

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

1



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

3

Share

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

5

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

6

Publish

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

8

Discover

- Use a Web browser
- Access any storage
- Use an existing identity

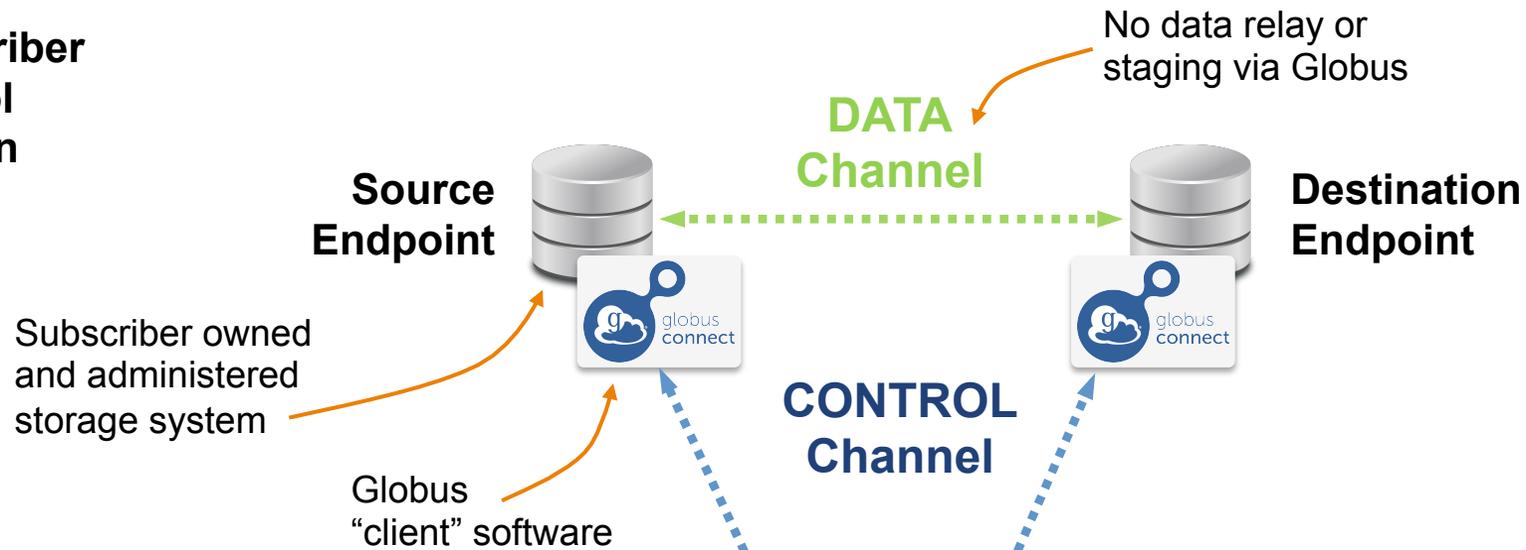


Personal Computer

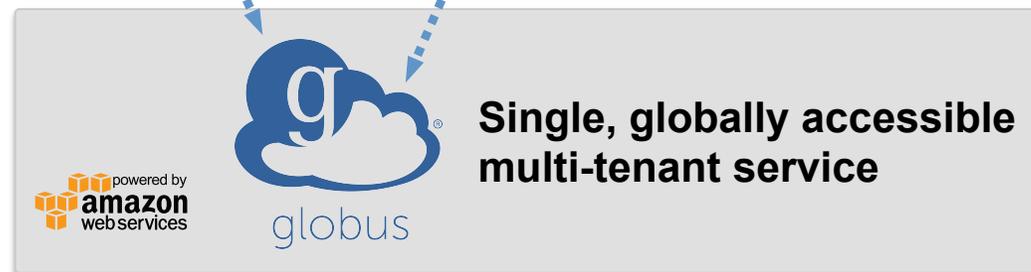


# Conceptual architecture: Hybrid SaaS

**Subscriber  
Control  
Domain**

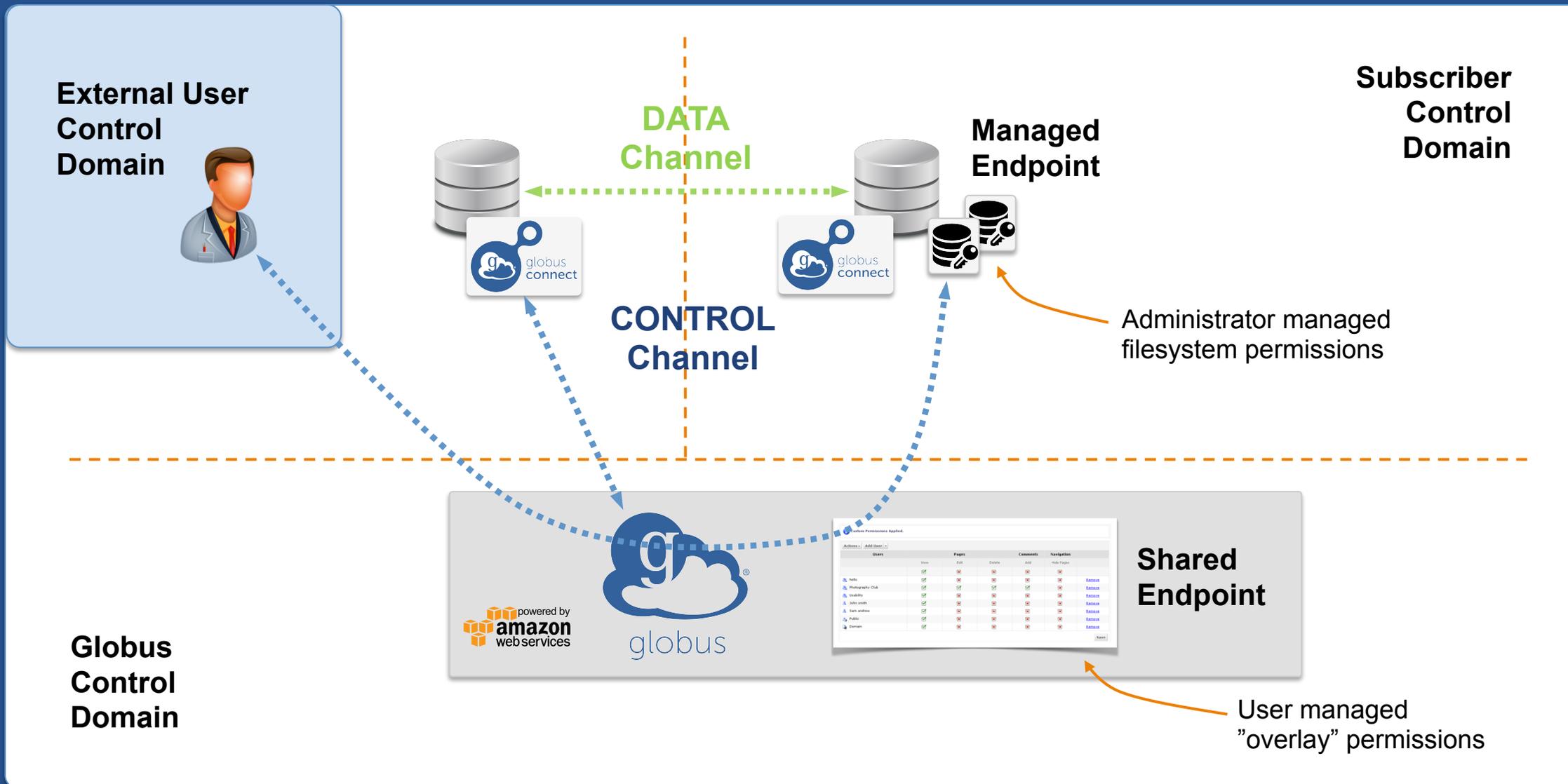


**Globus  
Control  
Domain**





# Conceptual architecture: Sharing





# 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**



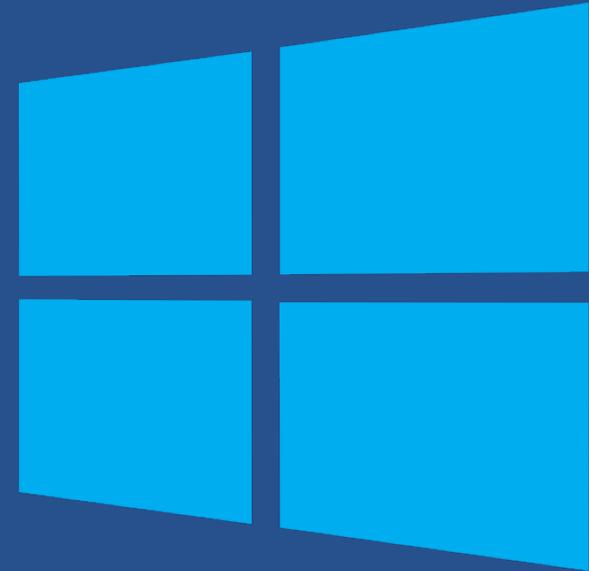
**How can I use Globus  
on my computer?**



...makes your  
storage system a  
**Globus endpoint**



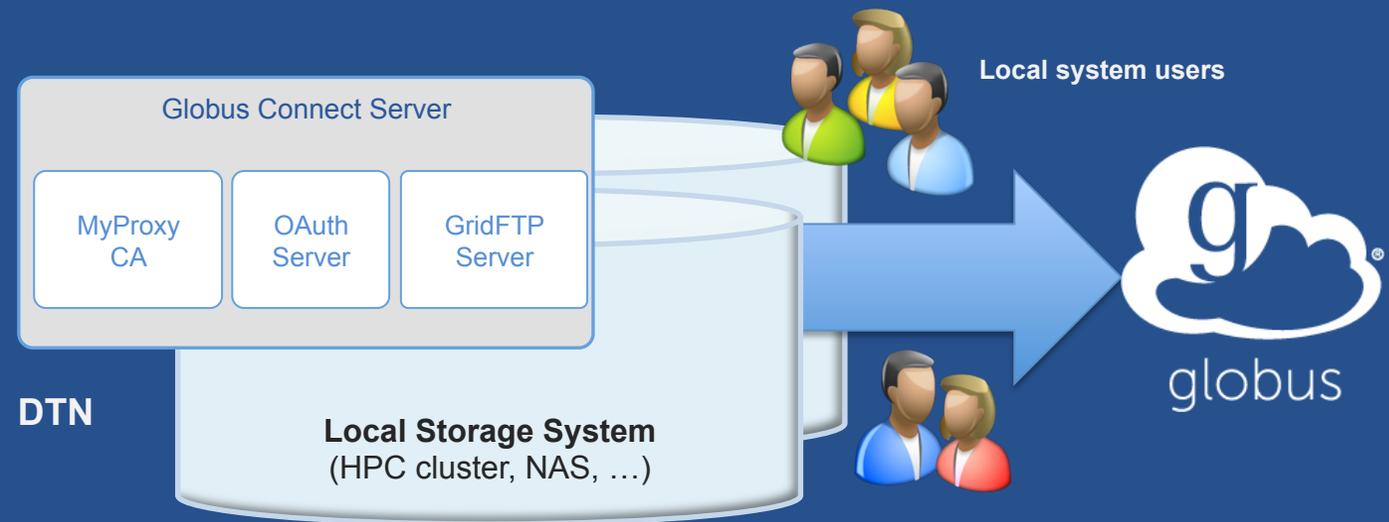
# Globus Connect Personal



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

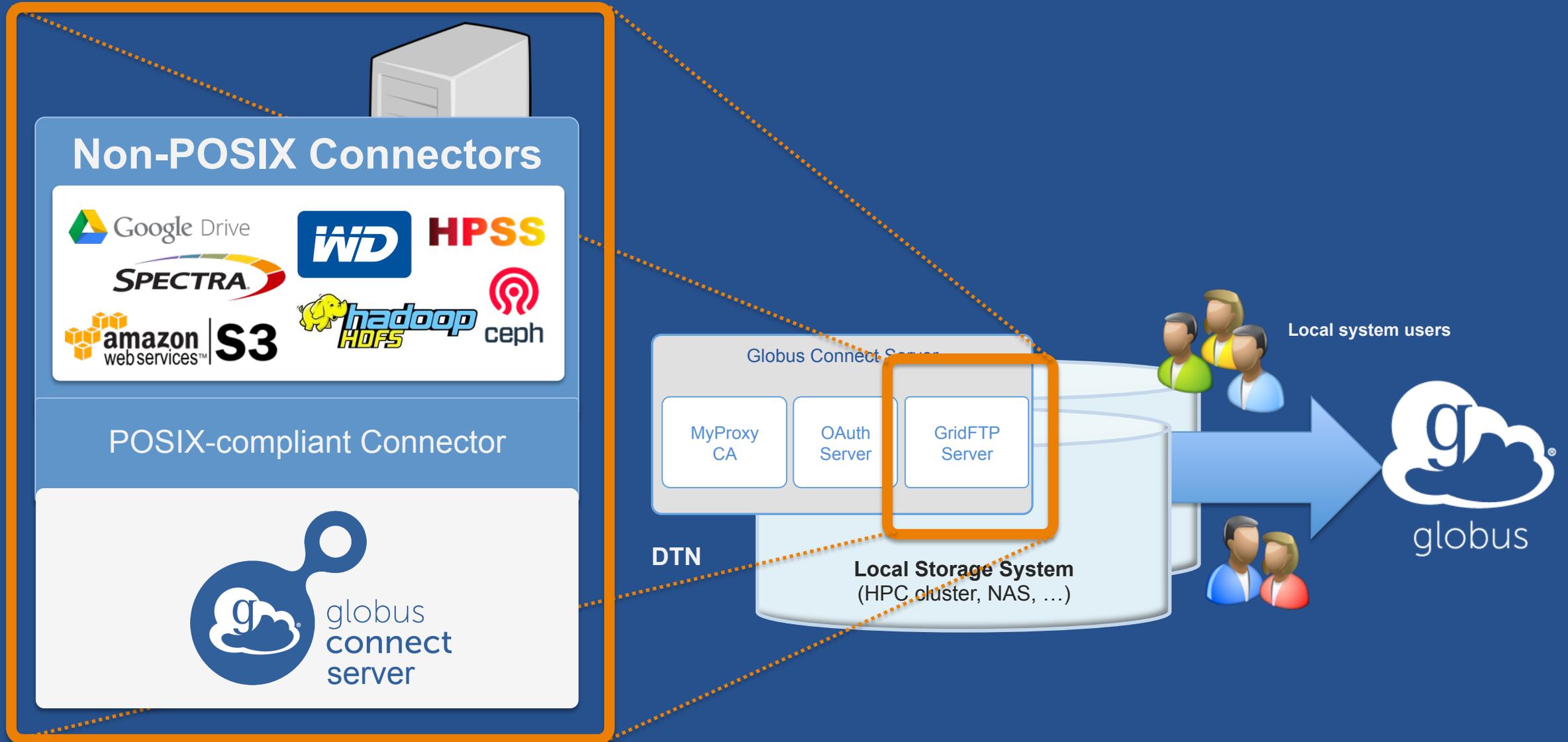
# Globus Connect Server

- Makes your storage accessible via Globus
- Multi-user server, installed and managed by sysadmin
- Default access for all local accounts
- Native packaging  
Linux: DEB, RPM



[docs.globus.org/globus-connect-server-installation-guide/](https://docs.globus.org/globus-connect-server-installation-guide/)

# Globus Connect Server





**How can I integrate  
Globus into my  
research workflows?**



**Globus serves as...**

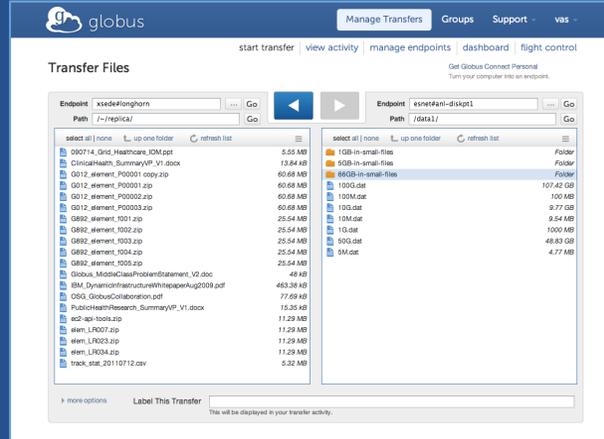
**...a platform for building  
science gateways, portals,  
and other web applications  
in support of research and  
education**



# Use(r)-appropriate interfaces



Globus service



Web

```
(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.
```

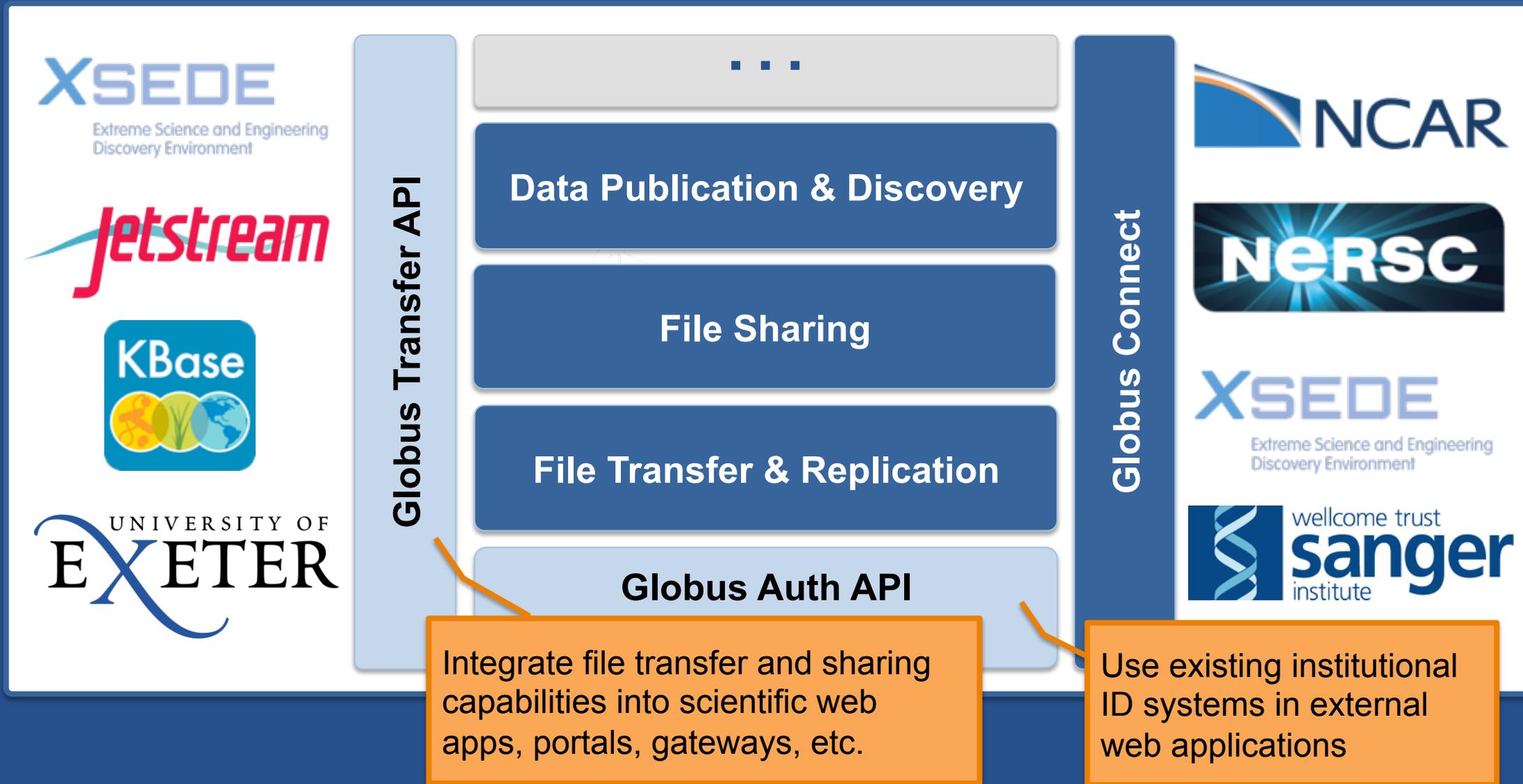
CLI

```
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





# 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

This SDK provides a convenient Pythonic interface to Globus REST APIs, including the Transfer API and the Globus Auth API. Documentation for the REST APIs is available at <https://docs.globus.org>.

Installation

The Globus SDK requires Python 2.6+ or 3.2+. If a suitable version of Python is not installed, you can install it from the Python website.

The simplest way to install the Globus SDK is using the following installation 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 Online"
tutorial_endpoint_2 = "ddb59af0-6d04-11e5-ba46-22000b92c6ec" # endpoint "Globus Online"
tutorial_users_group = "50b6a29c-63ac-11e4-8062-22000ab68755" # group "Tutorial Users Group"
```

## 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](https://docs.globus.org/api)

[github.com/globus](https://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

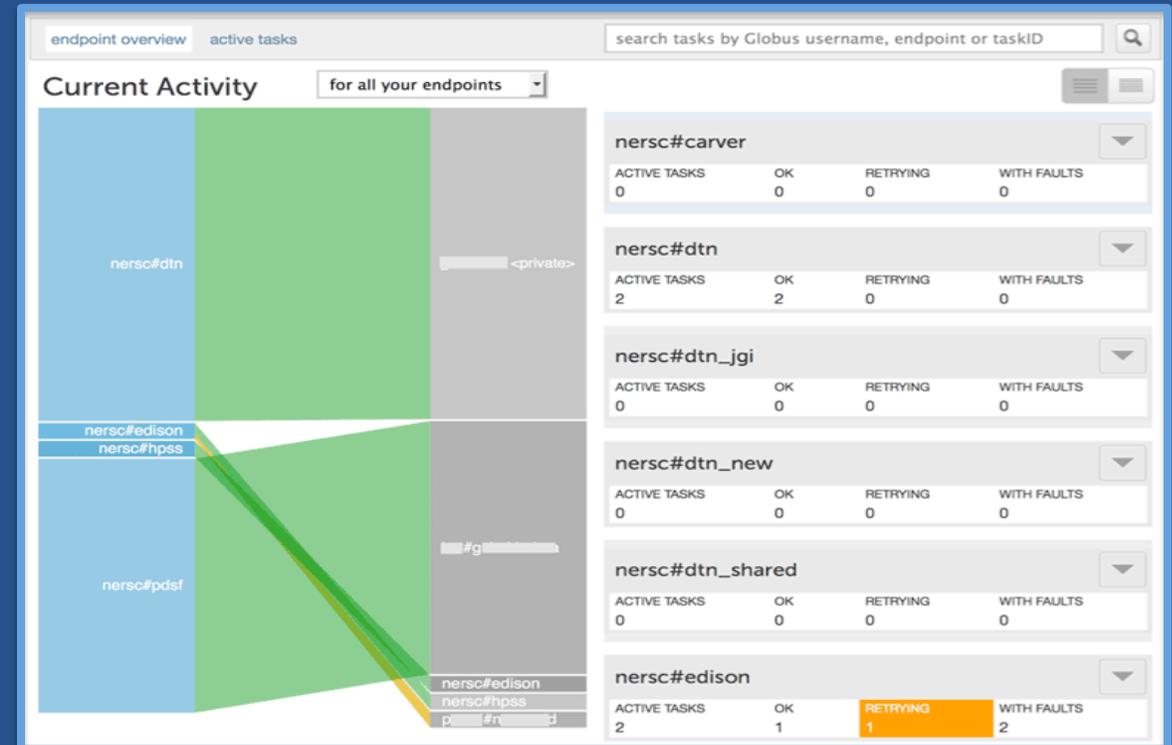
Argonne  
NATIONAL LABORATORY

powered by  
**amazon**  
web services



# Globus sustainability model

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





# Thank you to our users...

**48**

most server endpoints at a single organization

**384 PB**  
transferred

**64 billion**  
tasks processed

**76,000**  
registered users

**500**

100TB+ users

**14,000**  
active users

**3 months**  
longest running managed transfer

**10,000**  
active endpoints

**350+**  
federated identities

**1 PB**  
largest single transfer to date

**5,000**  
active shared endpoints

**99.5%**  
uptime

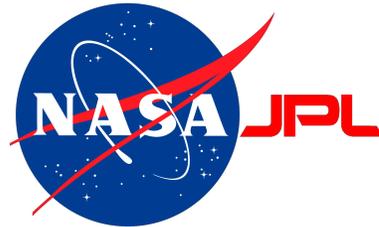
# Our supporters



JOHNS HOPKINS  
UNIVERSITY



Yale



THE UNIVERSITY OF  
CHICAGO

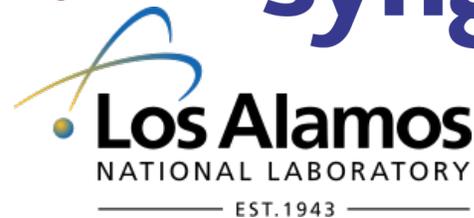


Stanford  
University



Dartmouth

SIMONS FOUNDATION





# Join the Globus community

- Access the service: [globus.org/login](https://globus.org/login)
- Create a personal endpoint: [globus.org/app/endpoints/create-gcp](https://globus.org/app/endpoints/create-gcp)
- Documentation: [docs.globus.org](https://docs.globus.org)
- Engage: [globus.org/mailing-lists](https://globus.org/mailing-lists)
- Subscribe: [globus.org/subscriptions](https://globus.org/subscriptions)
- Need help? [support@globus.org](mailto:support@globus.org)
- Follow us: [@globusonline](https://twitter.com/globusonline)