



Introduction to Globus for New Users SaaS for Research Data Management

Vas Vasiliadis
vas@uchicago.edu

NYSERNet– May 1, 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 (...)

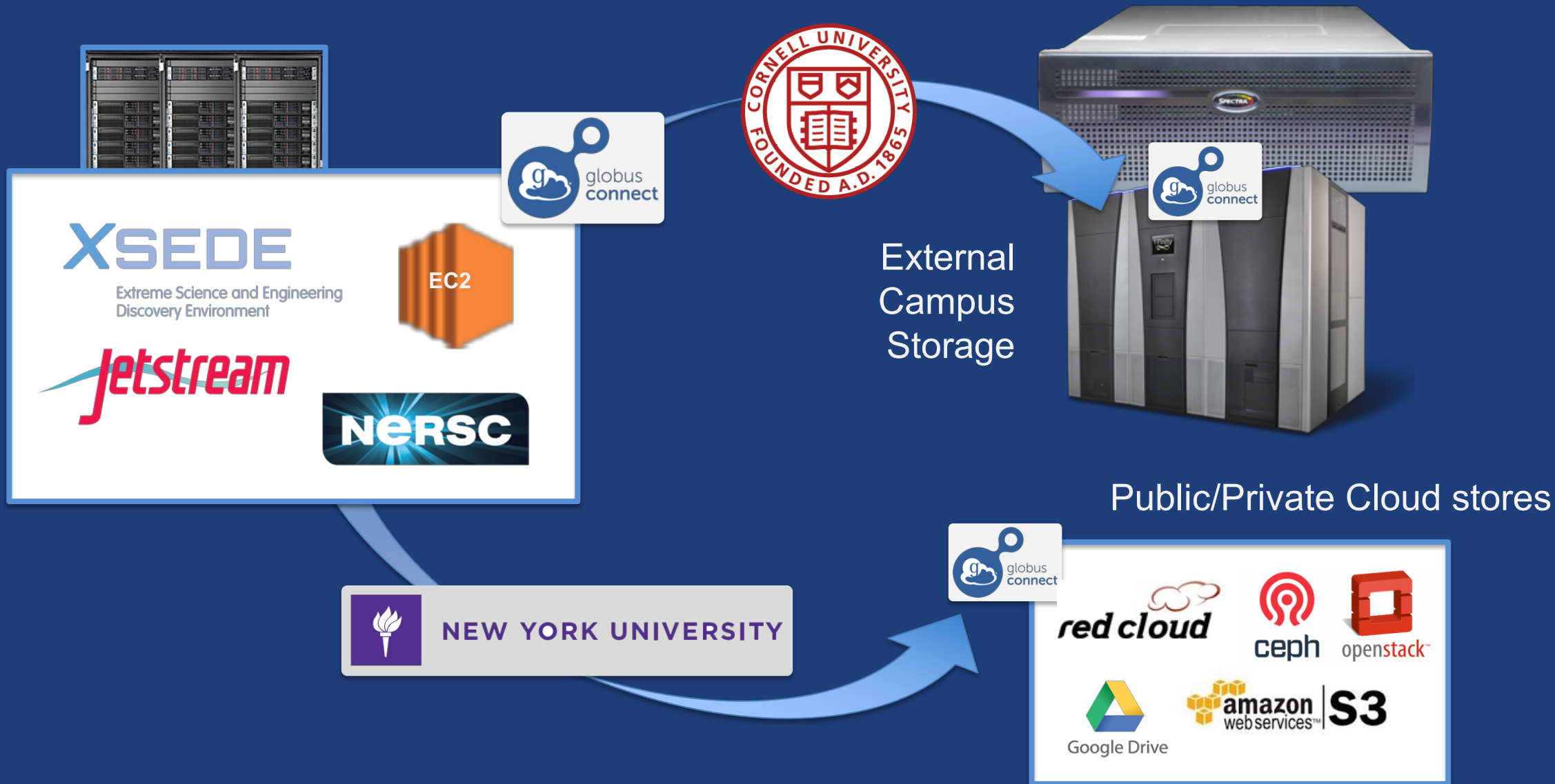


Bridge to instruments





Bridge to collaborators





Bridge to community/public



Project Repositories,
Replication Stores

XSEDE
Extreme Science and Engineering
Discovery Environment

Jetstream

EC2

NERSC

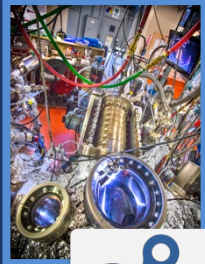
Public Repositories

globe connect

globe connect

Globus SaaS: Research data lifecycle

Instrument



Globus transfers files reliably, securely

2

Transfer

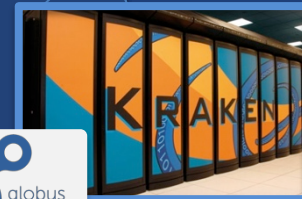
Compute Facility



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

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

6



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

5

Publish

6

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

8



Discover

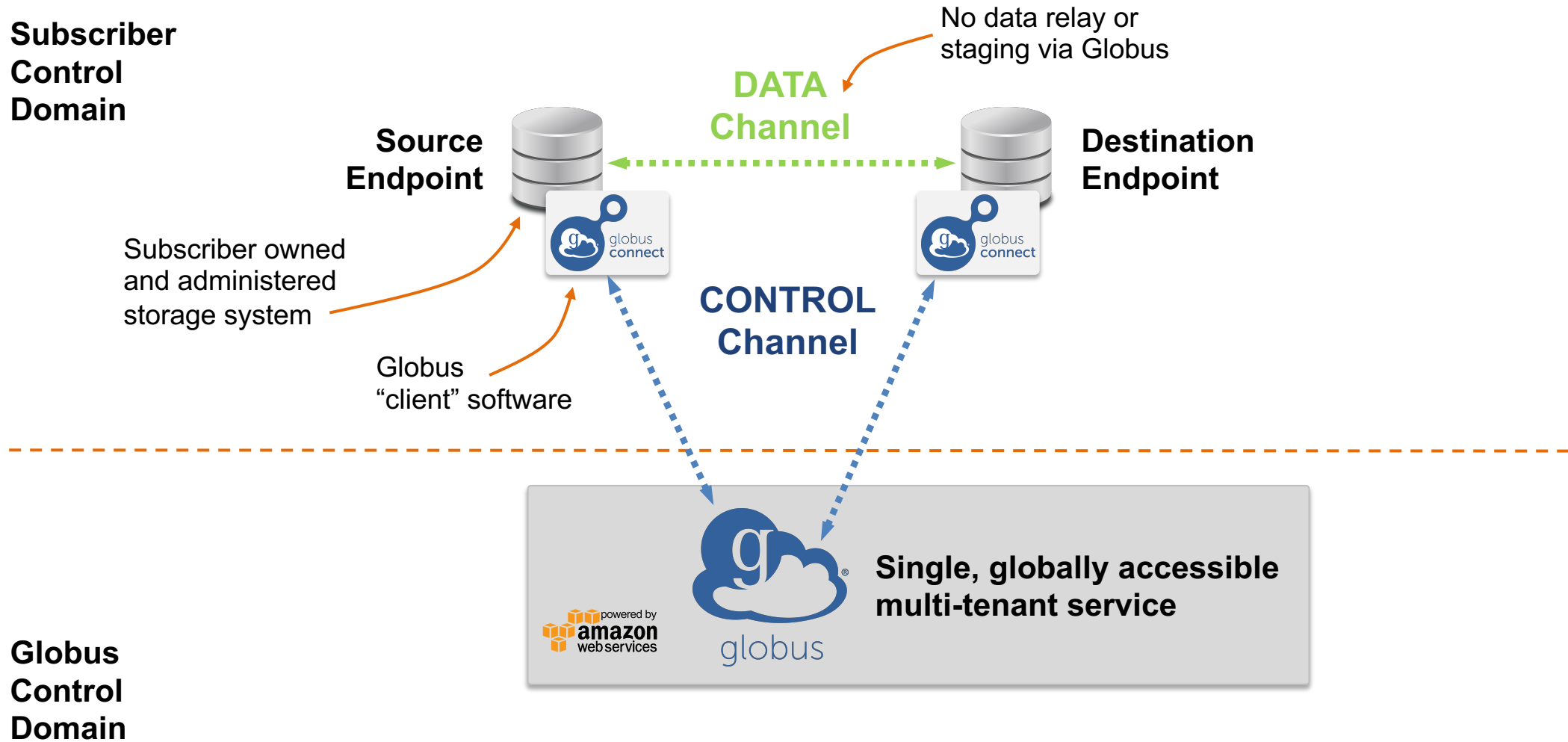
Personal Computer



- Access any storage...
- ...with an existing identity
- ...using a Web browser



Conceptual architecture: Hybrid SaaS



Endpoints: A Storage Abstraction

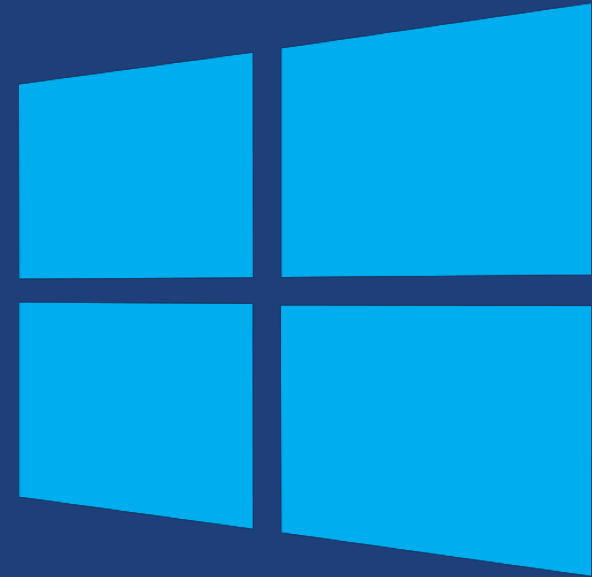
- **Endpoint: Storage abstraction**
- **All transfers happen between two endpoints**
- **Testing endpoints**
 - Globus Tutorial Endpoints
(initial endpoint validation)
 - ESnet Test Endpoints
(diverse dataset samples)



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



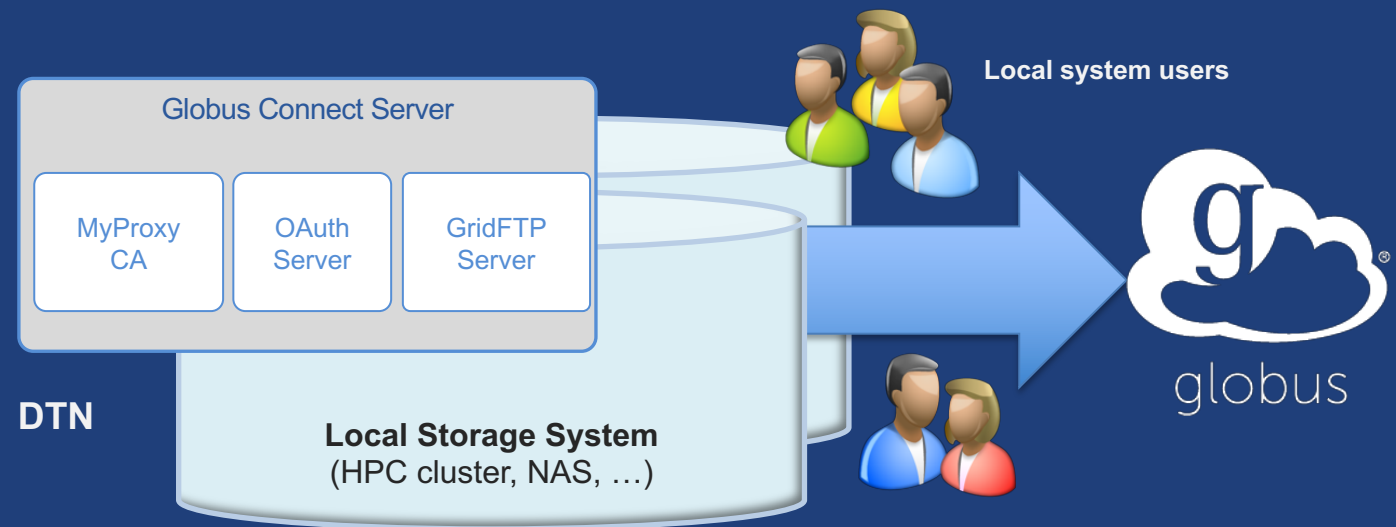
Globus Connect Personal



- **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/



Demonstration

File Transfer

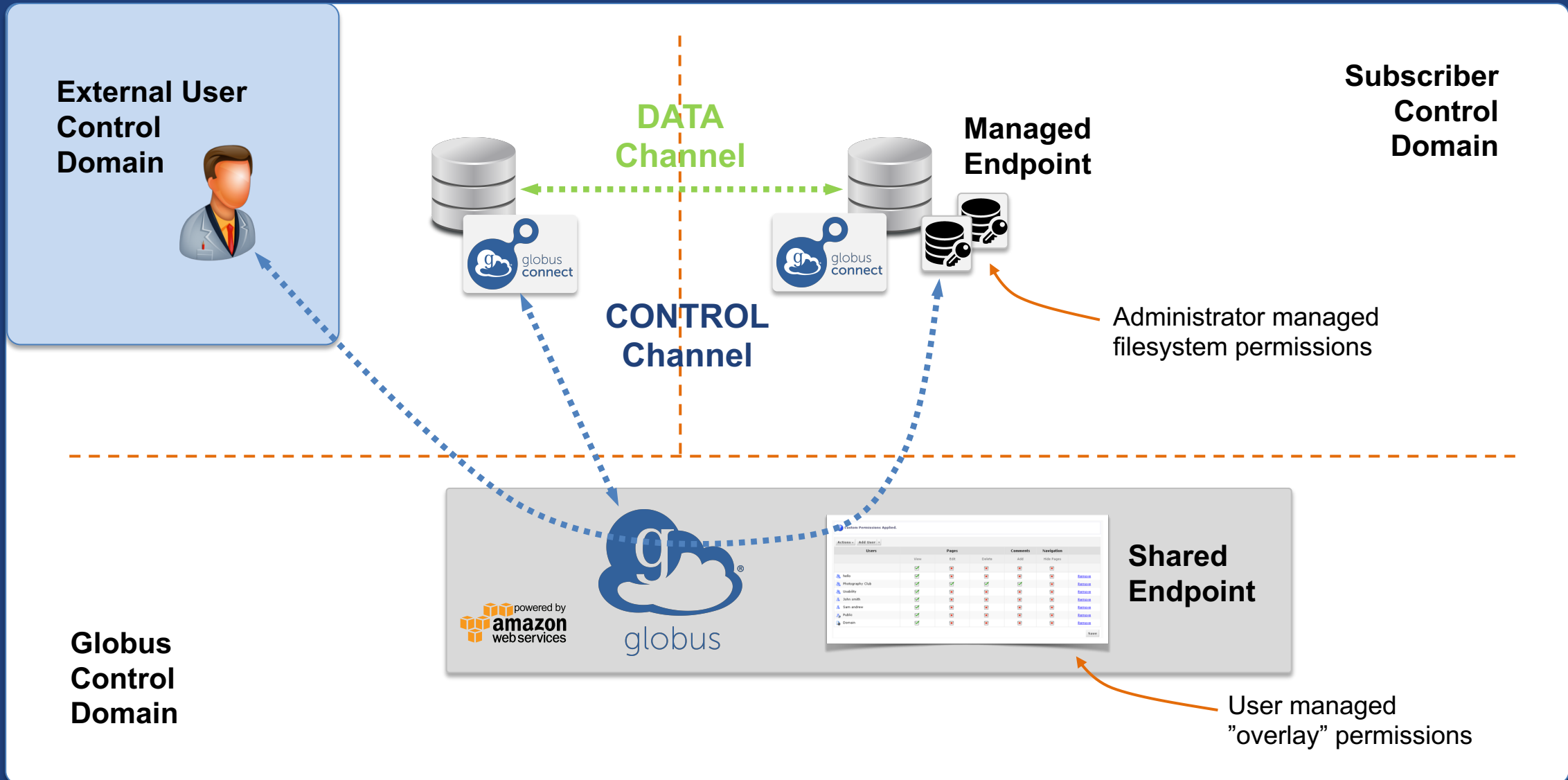


Data Sharing

- **Select the directory and create a “shared endpoint”**
 - ...just another type of endpoint
 - ...but does not require activation
- **Share with user/group/all Globus users**
- **Roles delegate control/monitor rights to endpoints**



Conceptual architecture: Sharing





Demonstration

File Sharing



Groups

- **Sharing: Access permissions for multiple people**
- **Roles: Endpoint management and monitoring**
- **Key concepts**
 - Visibility
 - Membership model
 - Subgroups
 - Settings (policies, additional data, terms)
 - Roles (delegating authority to others)



Demonstration **Group Management**



The Globus Web App - Hidden in Plain Sight

- **The Hamburger Menu**
 - Varies by endpoint/storage type
 - A great place to get the link to a share
- **Transfer Settings**
 - label –see recognizable names in activity monitor
 - sync - only transfer new or changed files
 - delete files on destination that do not exist on source
 - preserve source file modification times
 - verify file integrity after transfer **
 - encrypt transfer
- **Unified search: Endpoints, users, groups**



Bookmarks

- **Just like browser bookmarks – frequently used, or maybe not used frequently enough!**
- **Creating a bookmark**
- **Using a bookmark**
- **Sorting and filtering**
- **Editing and deleting**



Data Publication and Discovery

The screenshot shows the Materials Data Facility (MDF) community home page. At the top left is the Globus logo and the word "globus". To the right are "Log In" and "Sign Up" links. A blue notification bar states: "To submit a dataset or view datasets that have restricted access, please log in." Below this is a search bar with the placeholder text "Search" and a magnifying glass icon. The main content area features the heading "Materials Data Facility Community home page" and a large logo for "MATERIALS DATA FACILITY" composed of colorful circles. Below the logo, a paragraph describes the MDF as a scalable repository for materials scientists. It mentions that the MDF is a pilot project funded by NIST and serves as the first pilot community of the National Data Service. A link is provided to contact Ben Blaiszik (blaiszik@uchicago.edu) to begin publishing data. At the bottom, there is a "Browse" section with four buttons: "Issue Date", "Author", "Title", and "Subject".

<https://publish.globus.org>

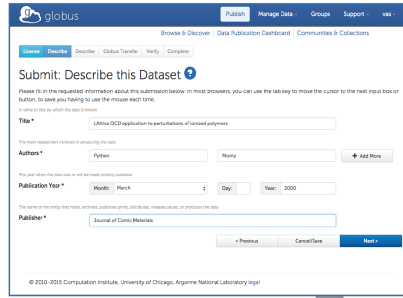


Globus data publication framework

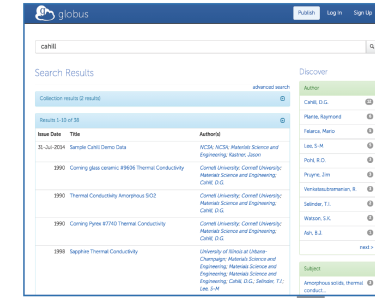
Identifier	URL	Handle	DOI	
Description	None	Standard	Domain-specific	Custom
Curation	None	Acceptance	Human-validated	Machine-validated
Access	Anonymous	Public	Embargoed	Collaborators
Preservation	Transient	Project Lifetime	Archive	“forever”



Publish

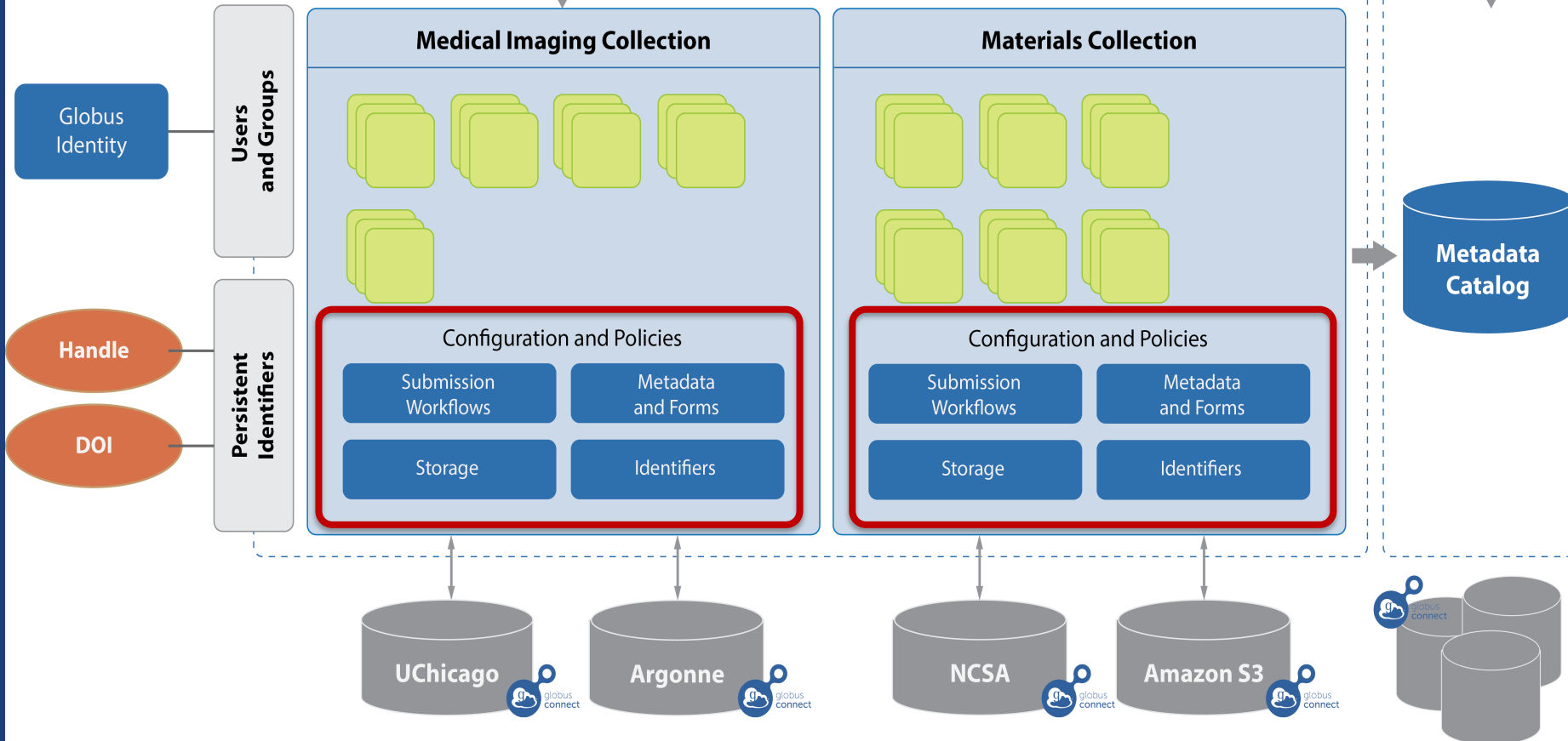


Discover



Globus Authentication

Globus Data Publication





Demonstration

Data Publication



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 integrate
Globus into my
research workflows?**



Globus serves as...

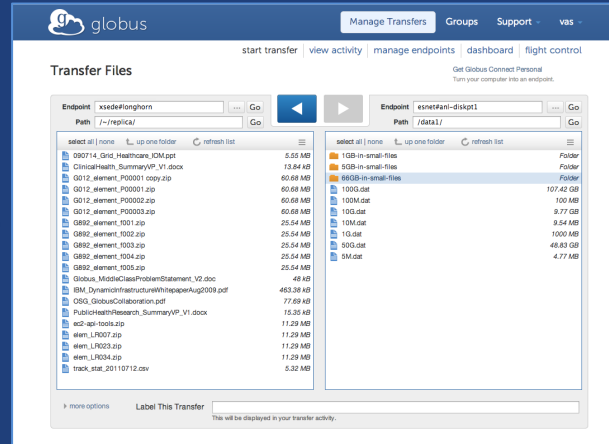
**...a platform for building
science gateways, portals,
and other web applications
in support of research and
education**



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

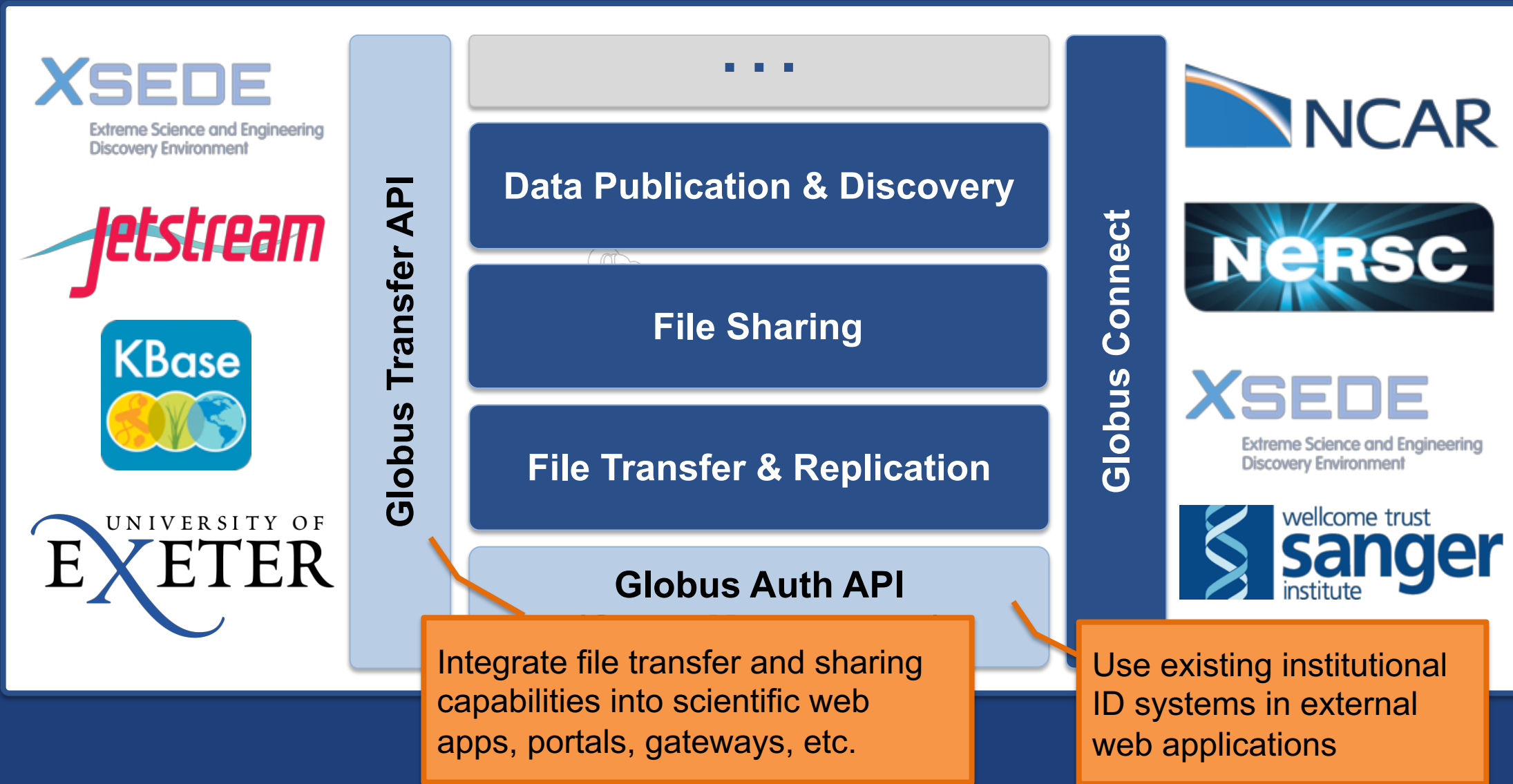
```
GET /endpoint/go%23ep1
PUT /endpoint/vas#my_endpt
200 OK
X-Transfer-API-Version: 0.10
Content-Type: application/json
...
```

Rest API





Globus as PaaS





Globus PaaS developer resources

globus-sdk-python 0.2.5 documentation »

Globus SDK for Python (Beta)

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>.

Python SDK

Installation

The Globus SDK requires Python 2.6+ or 3.2+. If a suitable version is not installed, you can install it 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
cd globus-sdk-python
python setup.py install
```

Modern Research Data Portal

It's how research data management is done!

LOGIN | SIGN UP

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 Transfer"
tutorial_endpoint_2 = "ddb59af0-6d04-11e5-ba46-22000b92c6ec" # endpoint "Globus Auth"
tutorial_users_group = "50b6a29c-63ac-11e4-8062-22000ab68755" # group "Tutorial Users"
```

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



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



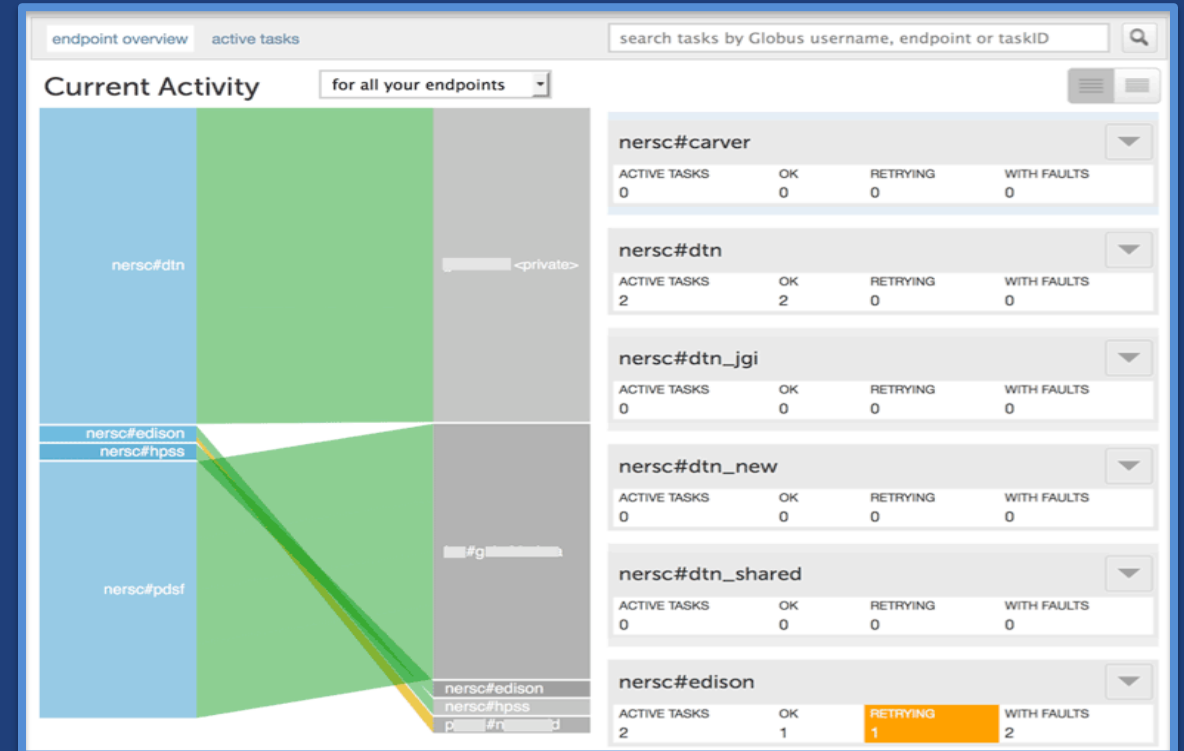
Thanks to our workshop sponsor!

Western Digital[®]



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)**





Globus by the numbers

1,042

most shared
endpoints
at a single
institution

400 PB

transferred

66 billion
files processed

100,000
users

24

Petabyte+
institutions

15,000

active transfer users

3 months

longest running transfer

20,000

active endpoints

500+

identity providers

1 PB

largest single
transfer to date

8,000

active shared
endpoints

99.9%+

availability



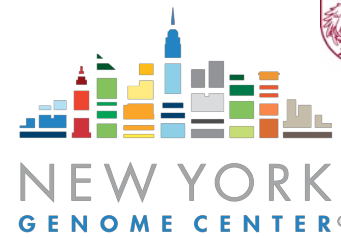
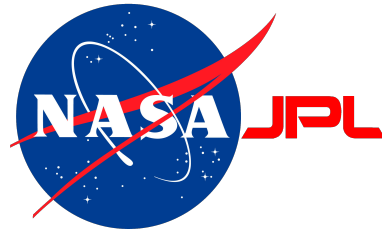
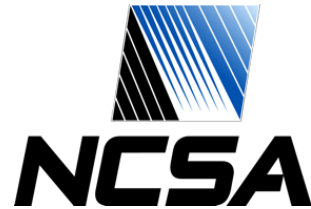
Our supporters



JOHNS HOPKINS
UNIVERSITY



Yale



THE UNIVERSITY OF
CHICAGO



Stanford
University



Dartmouth

SIMONS FOUNDATION





Join the Globus community

- Access the service: globus.org/login
- 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)