Introduction to Globus for New Users SaaS for Research Data Management

Vas Vasiliadis
vas@uchicago.edu

NYSERNet– May 1, 2018

globus

Research data management today

Index?

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

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

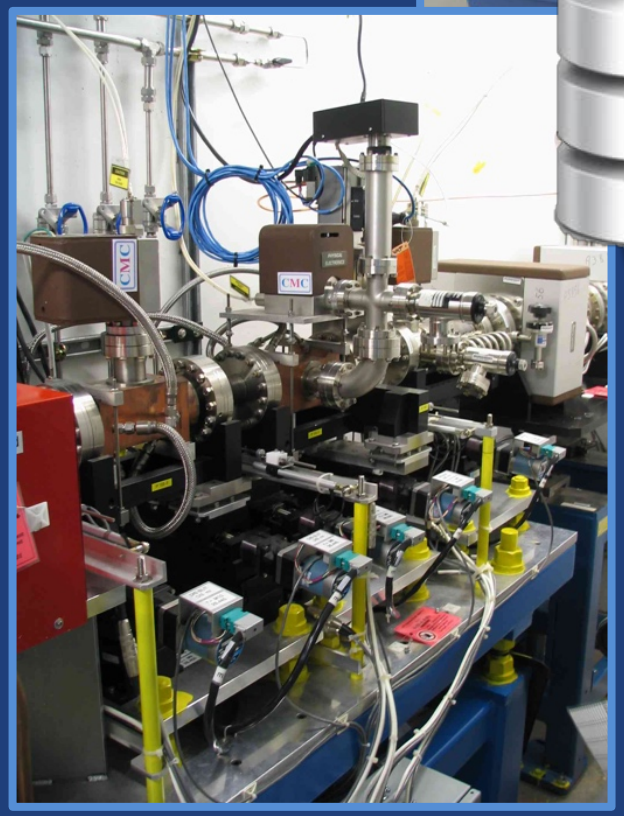
Move results to laptop, department, lab, …

Bridge to national cyberinfrastructure

Move datasets to supercomputer, national facility

Move results to campus (…)

Bridge to instruments



Pre-processed Data

Analysis store

globus
connect

Raw
Source
Data

High durability,
low cost store

globus
connect

Amazon
Glacier

Bridge to collaborators

globus connect

XSEDE
Extreme Science and Engineering Discovery Environment

EC2

Jetstream

NERSC

CORNELL UNIVERSITY
FOUNDED A.D. 1865

External Campus Storage

NEW YORK UNIVERSITY

globus connect

Public/Private Cloud stores

red cloud

ceph

openstack

Google Drive

amazon web services S3

Bridge to community/public

globus connect

Project Repositories,
Replication Stores

globus connect

XSEDE
Extreme Science and Engineering Discovery Environment

EC2

Jetstream

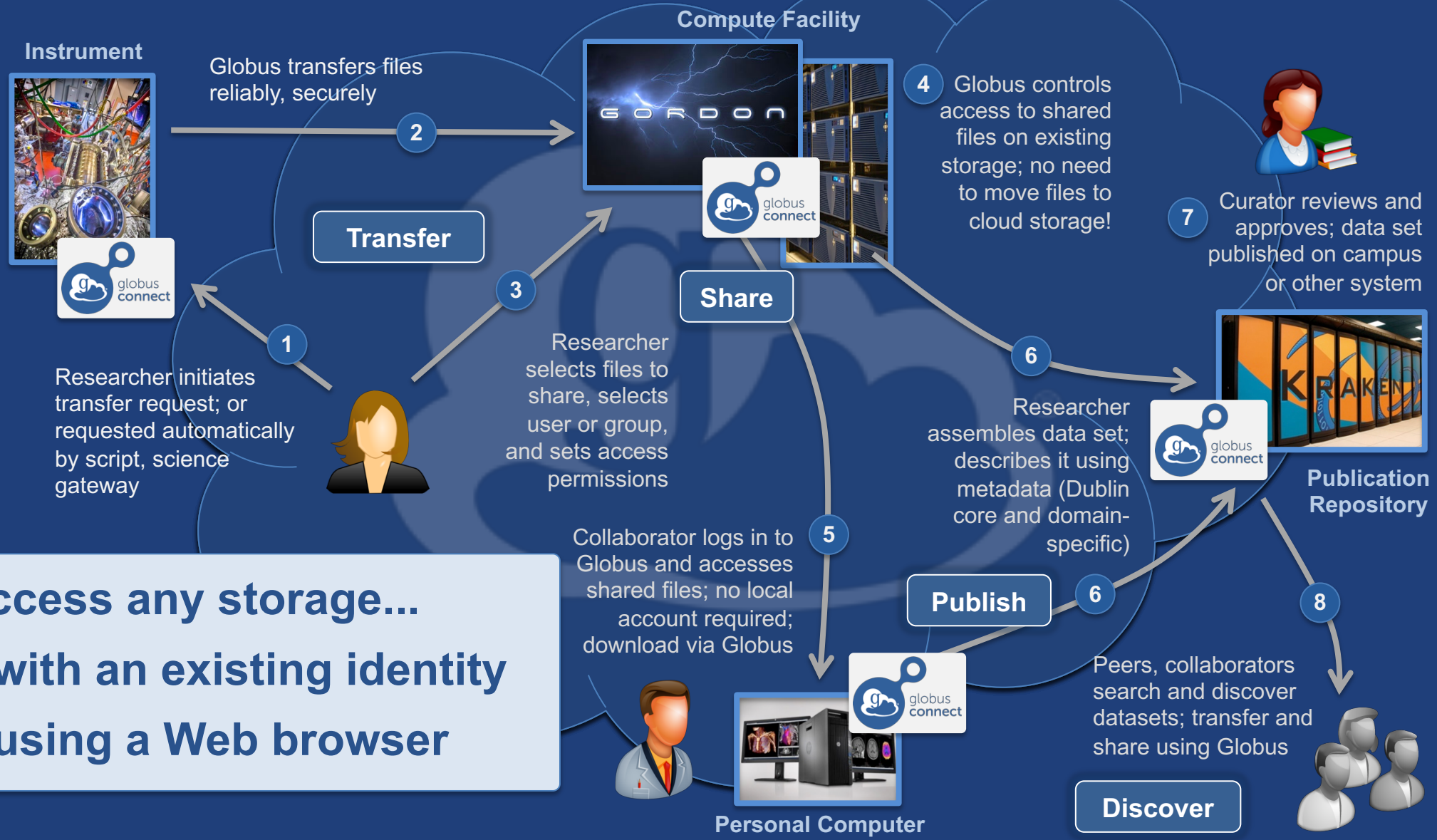
NERSC

Public Repositories

globus connect

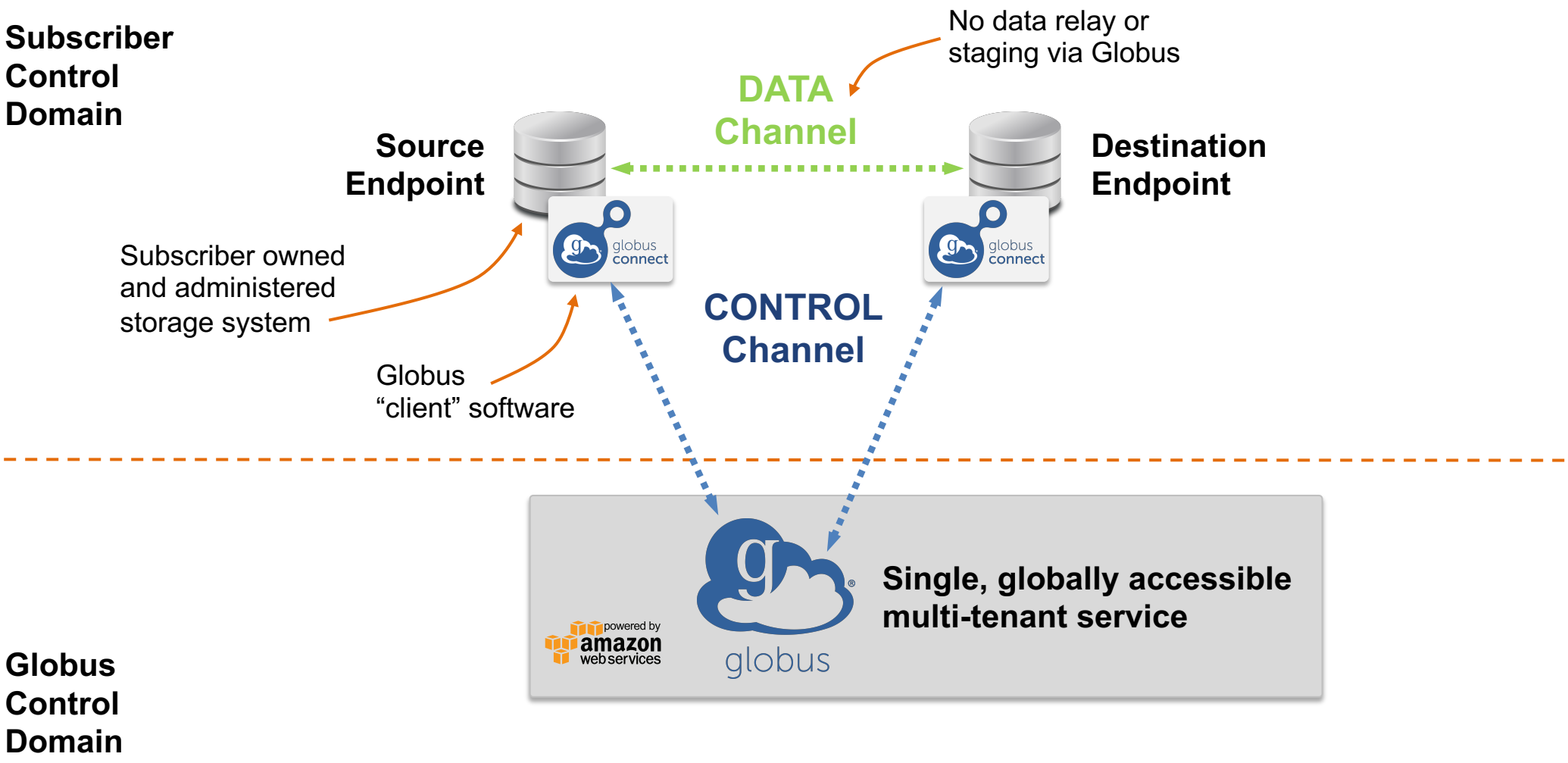
globus connect

Globus SaaS: Research data lifecycle



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

Conceptual architecture: Hybrid SaaS

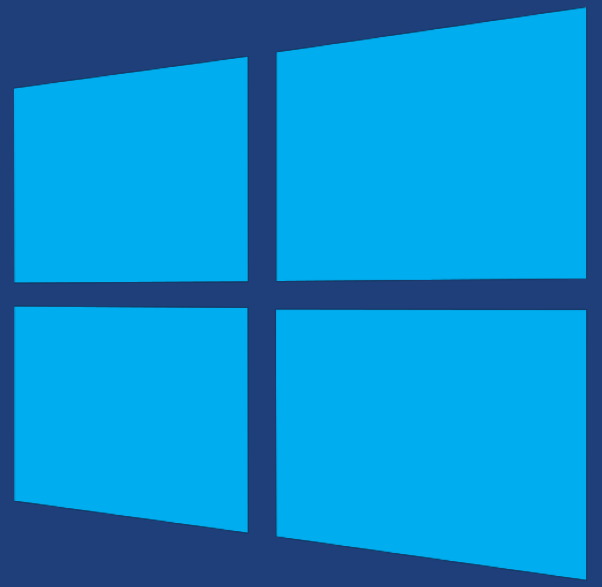


Endpoints: A Storage Abstraction

- **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

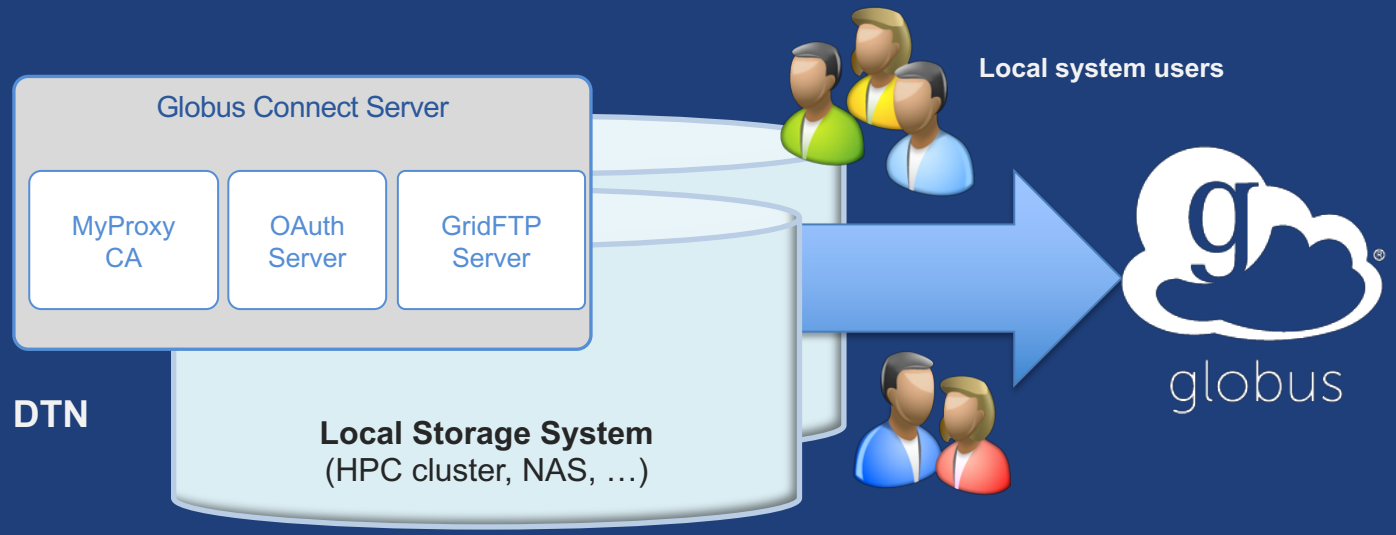
Globus Connect Personal



- **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**



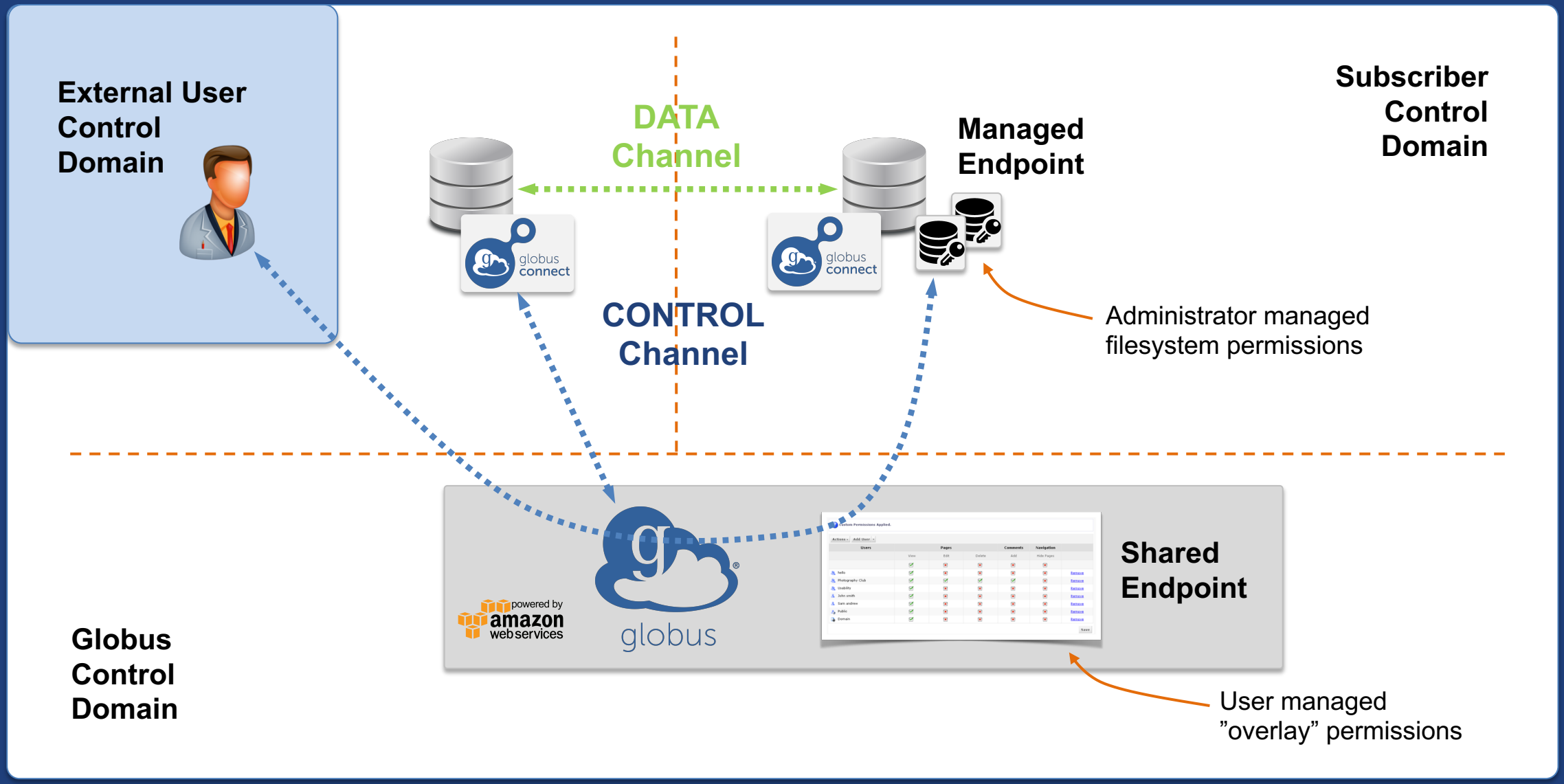
docs.globus.org/globus-connect-server-installation-guide/

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**

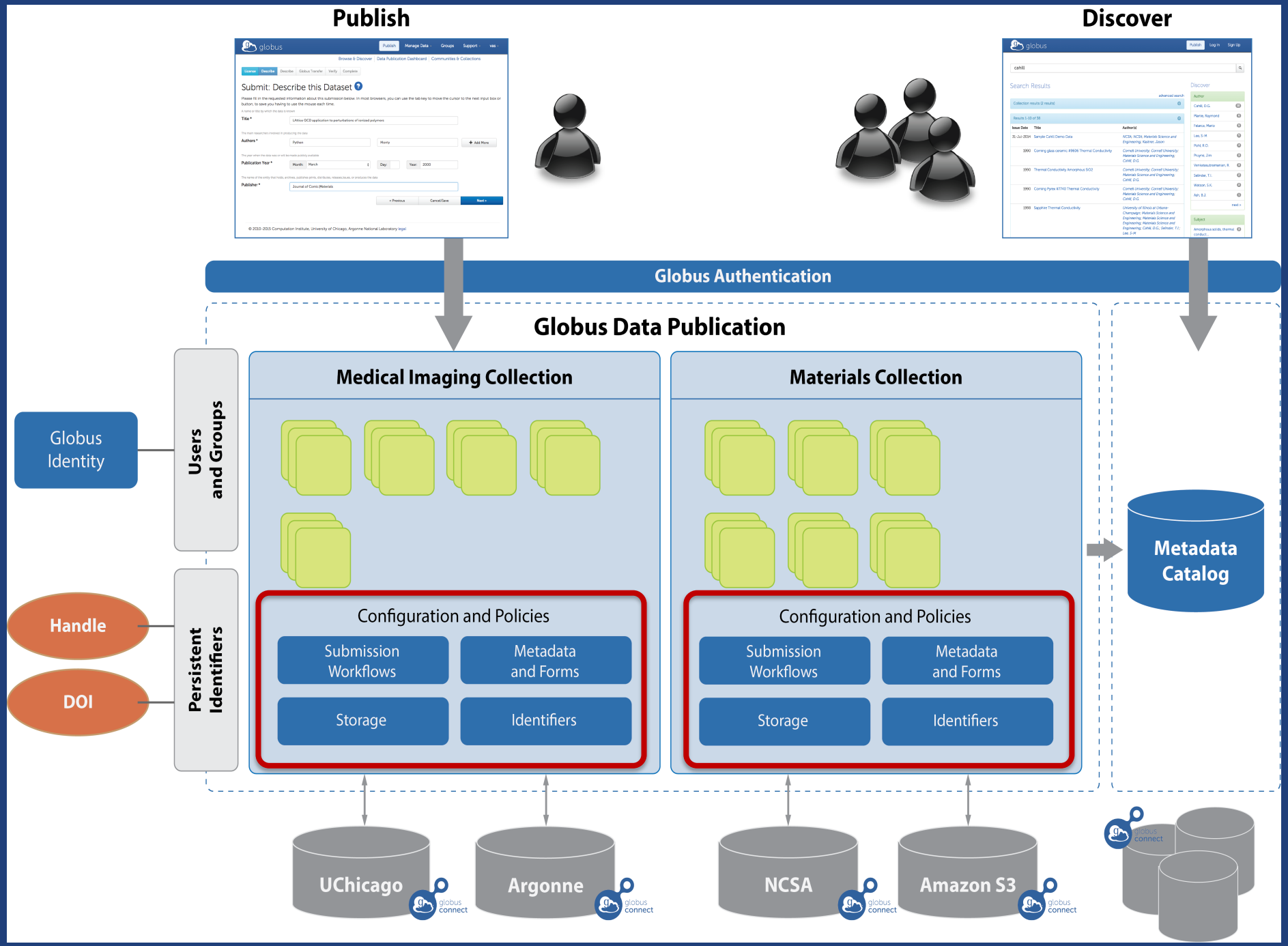
Bookmarks

- **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

https://publish.globus.org

Globus data publication framework


Publish
Discover
Globus Authentication
Globus Data Publication
Globus Identity
Users and Groups
Handle
DOI
Persistent Identifiers
Medical Imaging Collection
Materials Collection
Configuration and Policies
Submission Workflows
Metadata and Forms
Storage
Identifiers
Metadata Catalog
UChicago
Argonne
NCSA
Amazon S3
globus connect

Demonstration
Data Publication

Why use Globus?

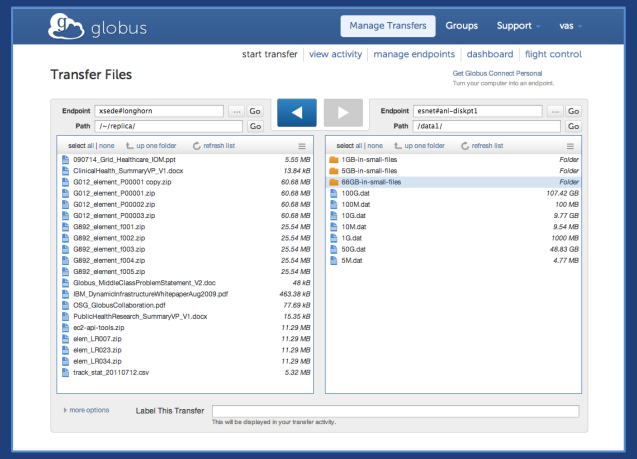
- **Simplicity**
  - Consistent UI across systems
  - Easy access to collaborators
- **Reliability and performance**
  - “Fire-and-forget” file transfer
  - Maximized WAN throughput
- **Operational efficiency**
  - Low overhead SaaS model
  - Highly automatable: CLI, RESTful API
- **Access to a large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



Web

Globus service

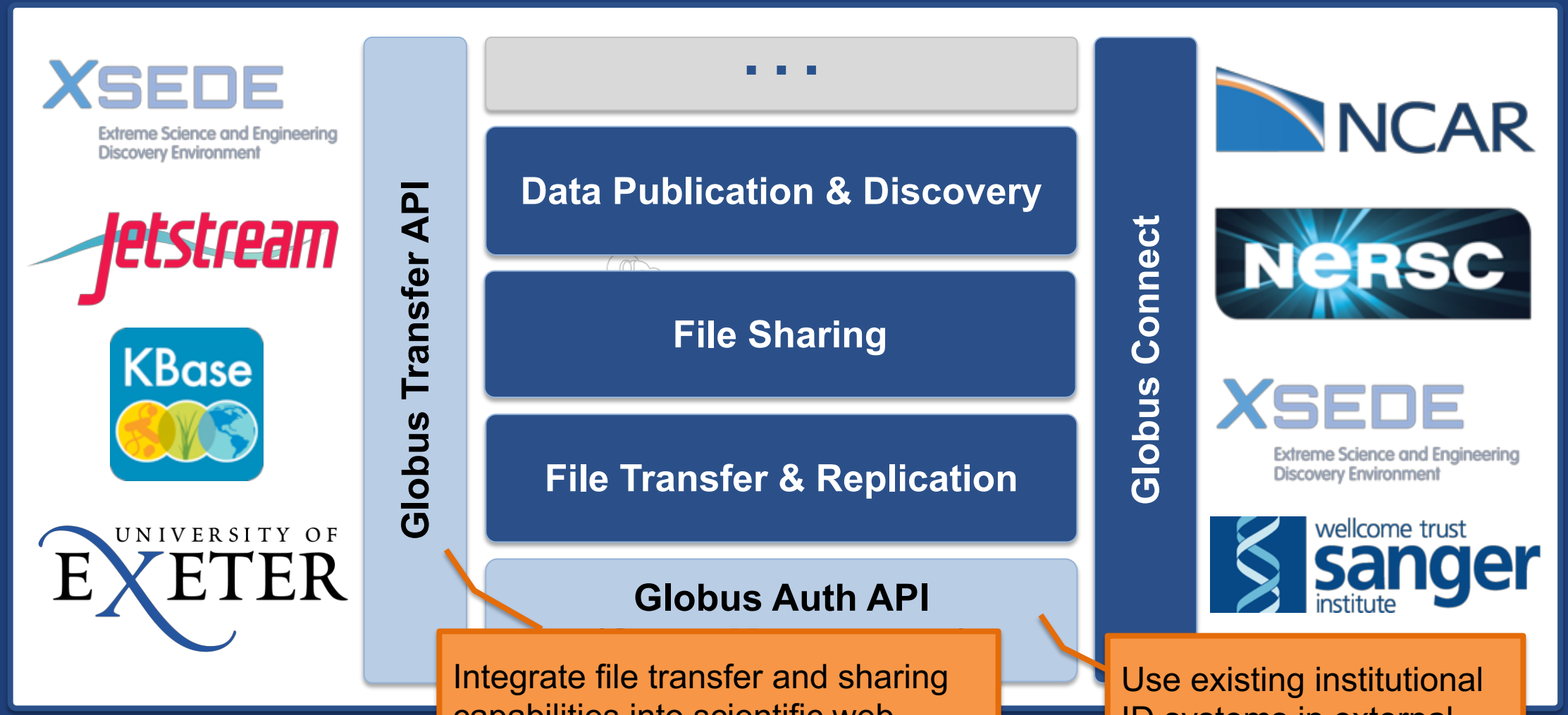
```
(globus-cli) jupiter:~ vas$ globus
Usage: globus [OPTIONS] COMMAND [ARGS]...

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

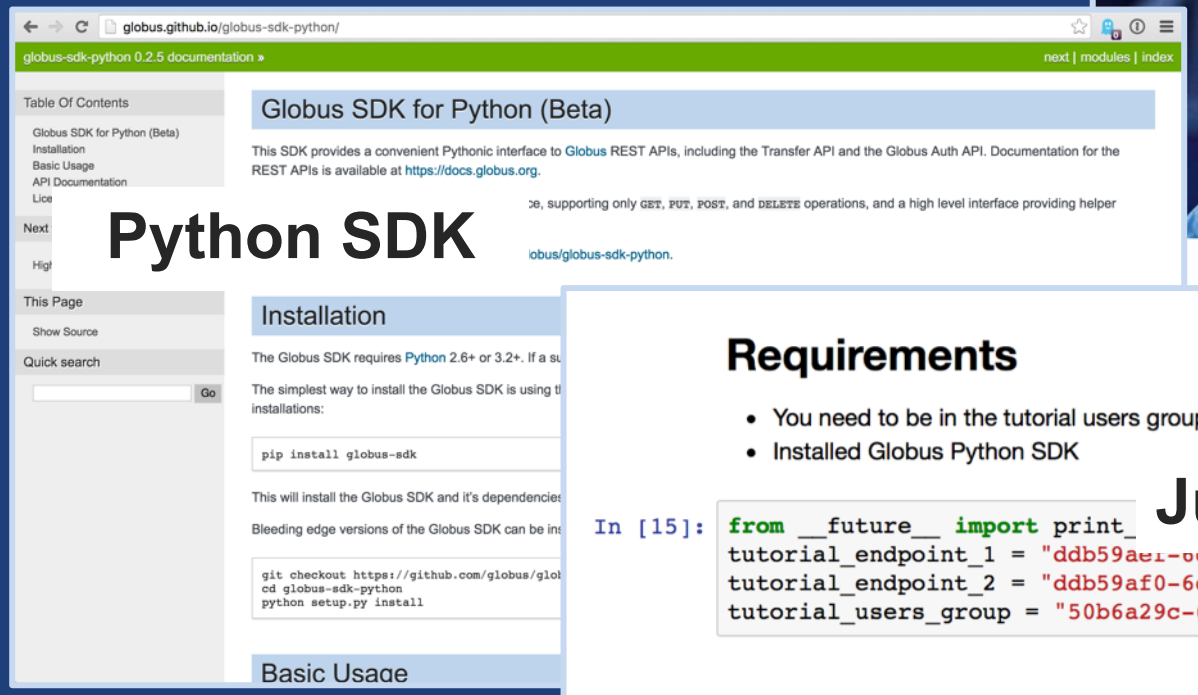
```
GET /endpoint/go%23ep1
PUT /endpoint/vas#my_endpt
200 OK
X-Transfer-API-Version: 0.10
Content-Type: application/json
...
```

Rest API

Globus as PaaS

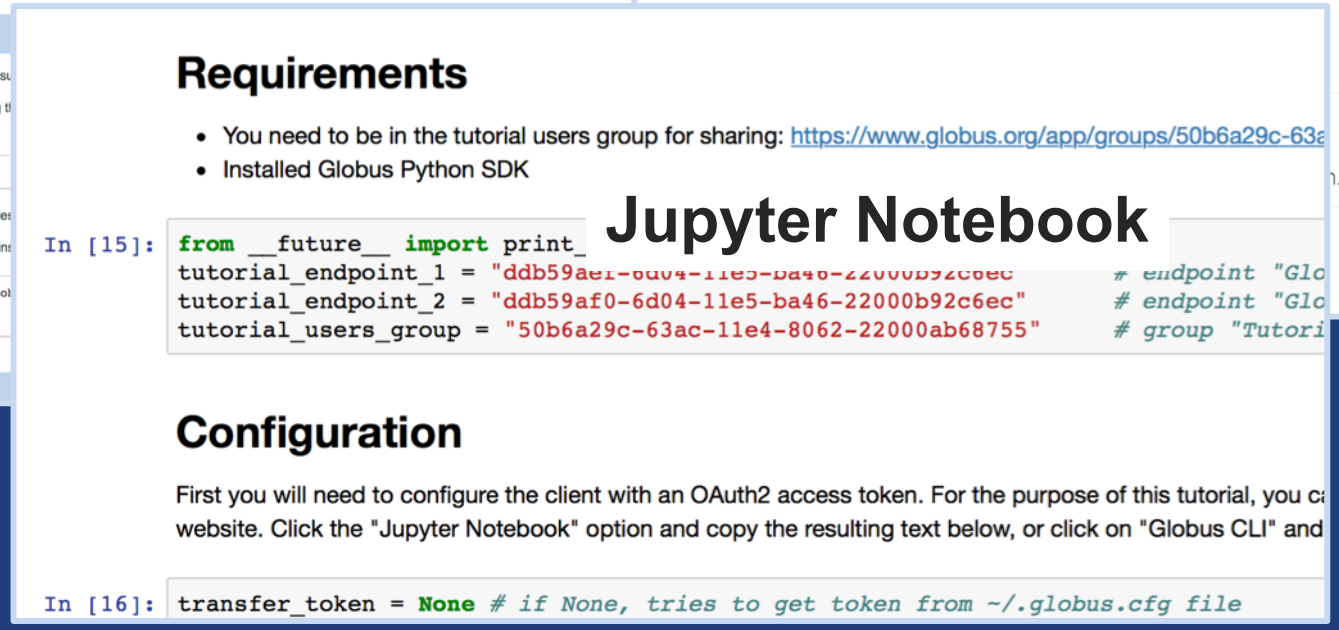


Globus PaaS developer resources



Python SDK

Sample Application



Jupyter Notebook

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

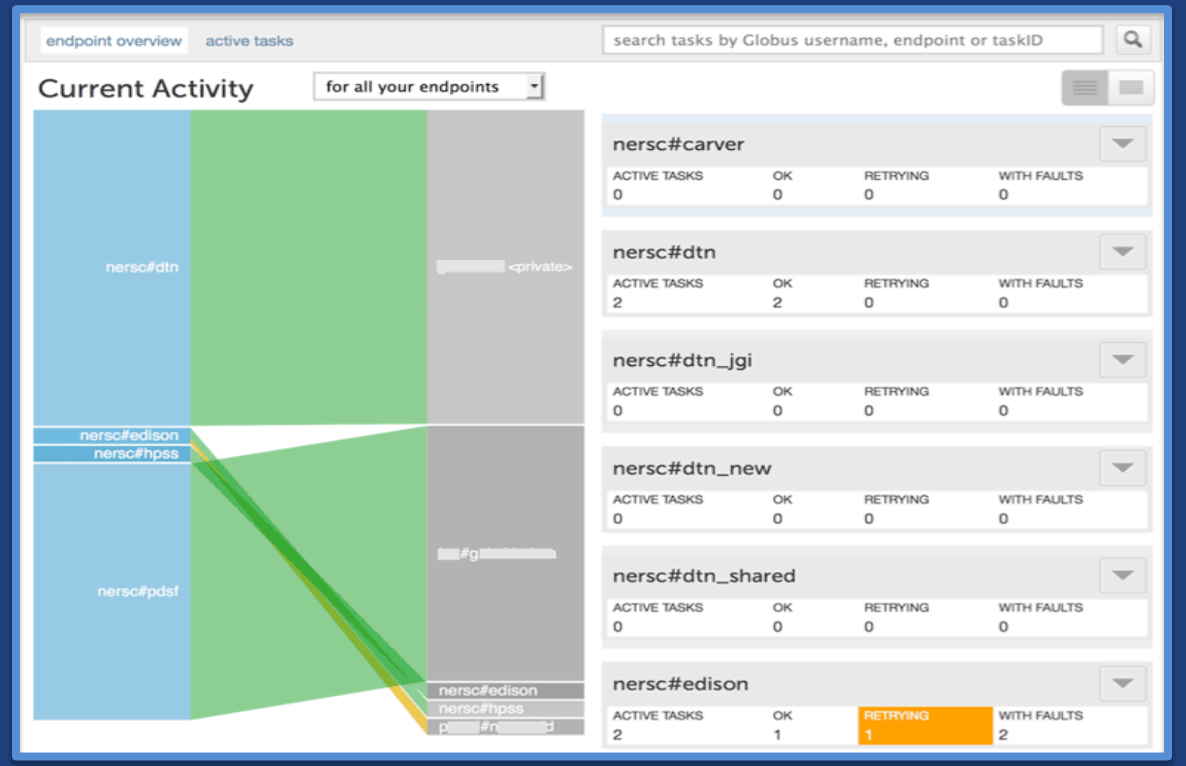
Argonne
NATIONAL LABORATORY

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

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**