



# Registration: [bit.ly/ec2ip](https://bit.ly/ec2ip)

1. Select an empty row in the spreadsheet
2. Enter your name and email address
3. Make a note of the IP address displayed

Slides and useful links:  
[globusworld.org/tutorials](https://globusworld.org/tutorials)



# Introduction to Globus: SaaS for Research Data Management

Vas Vasiliadis  
vas@uchicago.edu

Rachana Ananthakrishnan  
rachana@globus.org

Stanford University – February 8, 2018





# Research data management today



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

Index?





Globus delivers...

Big 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**

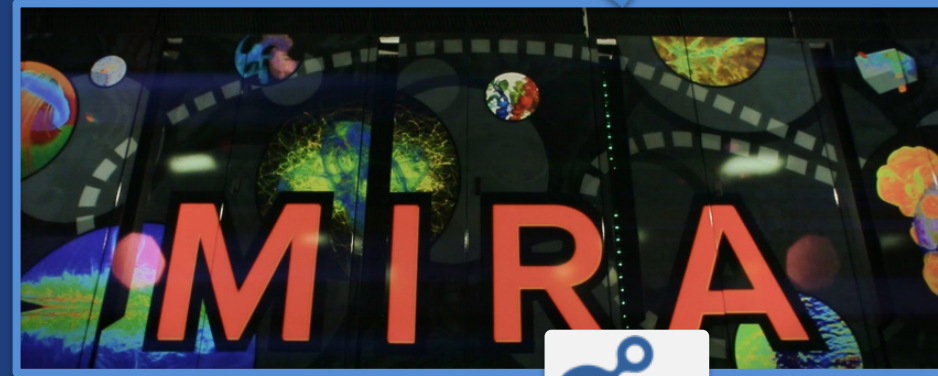


**Move results to laptop, department, lab, ...**



Bridge to national cyberinfrastructure

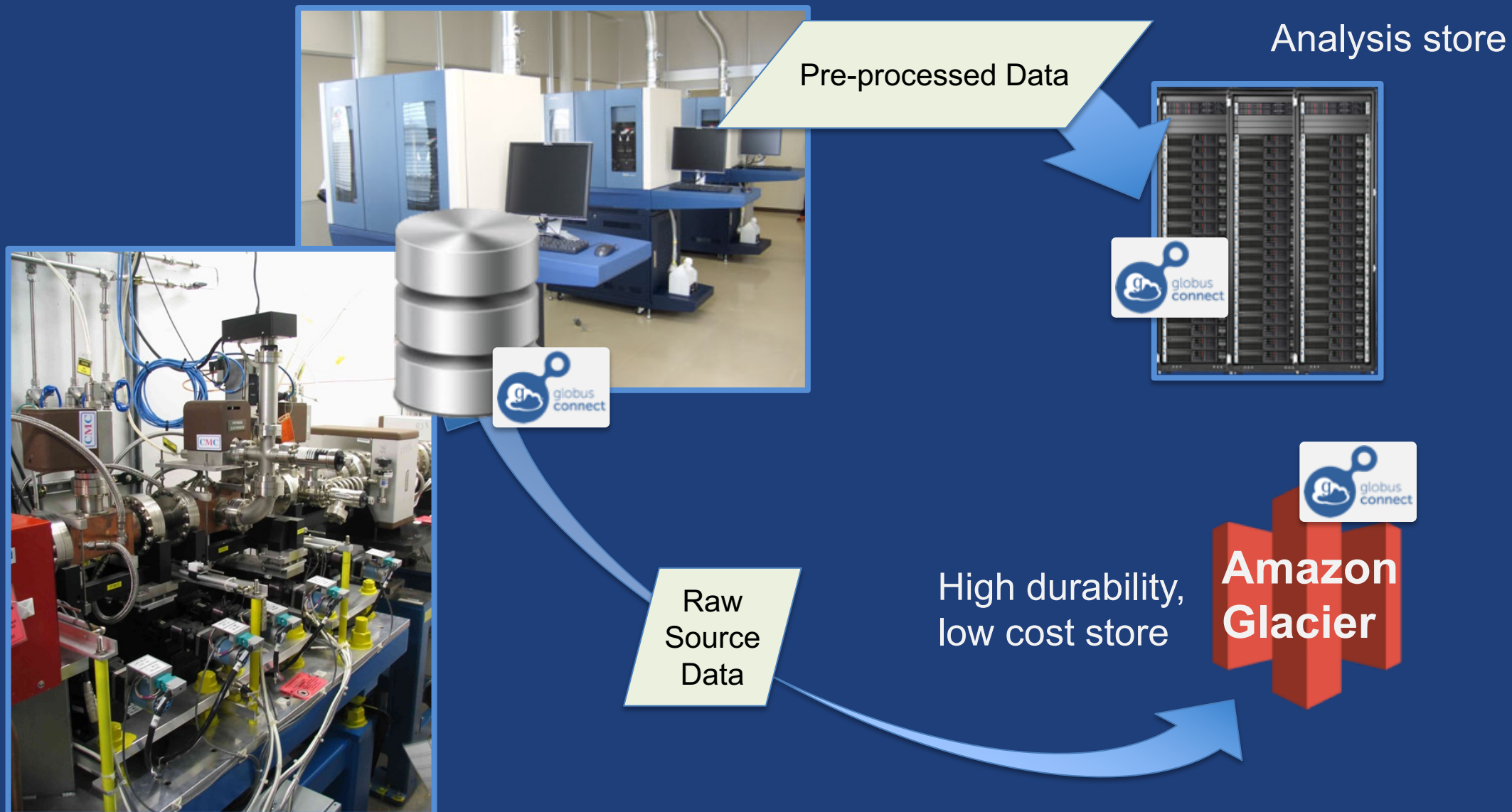
Move datasets to supercomputer,  
national facility



Move results to campus (...)



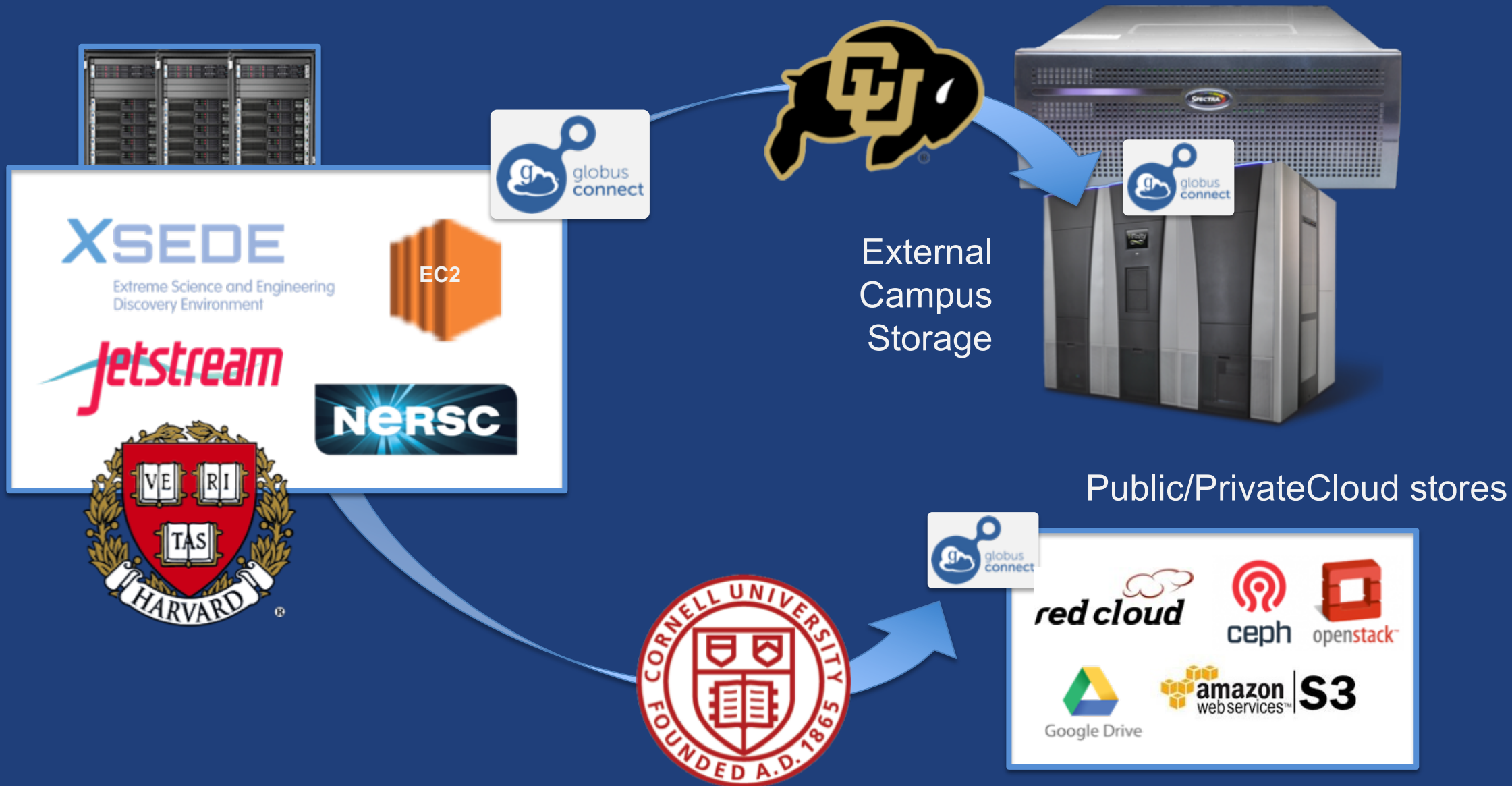
# Bridge to instruments







# Bridge to collaborators





# Bridge to community/public



Project Repositories,  
Replication Stores



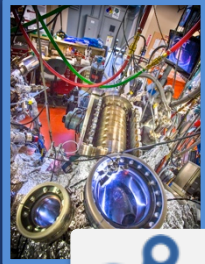
Public Repositories





# 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

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

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

8



Discover



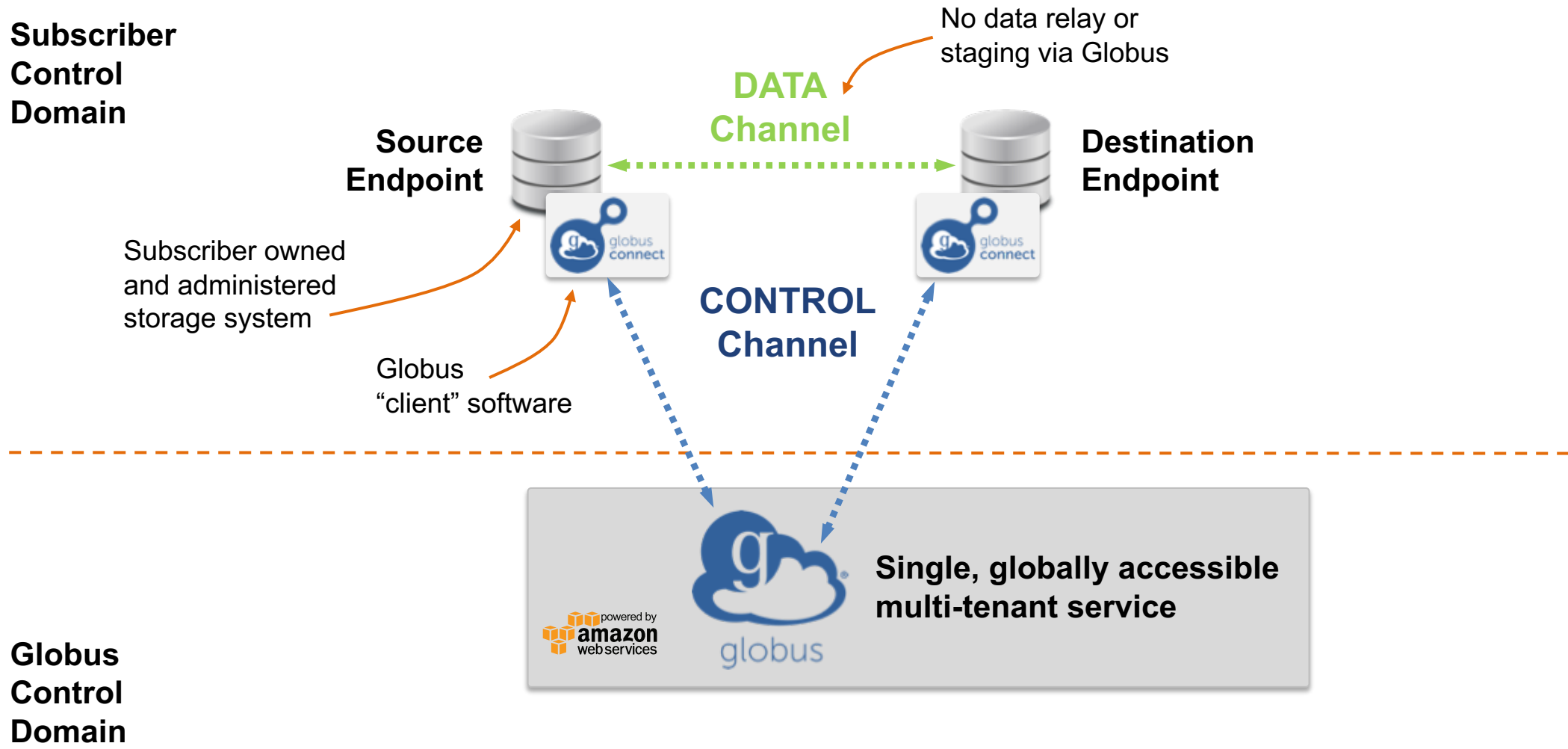
Personal Computer



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

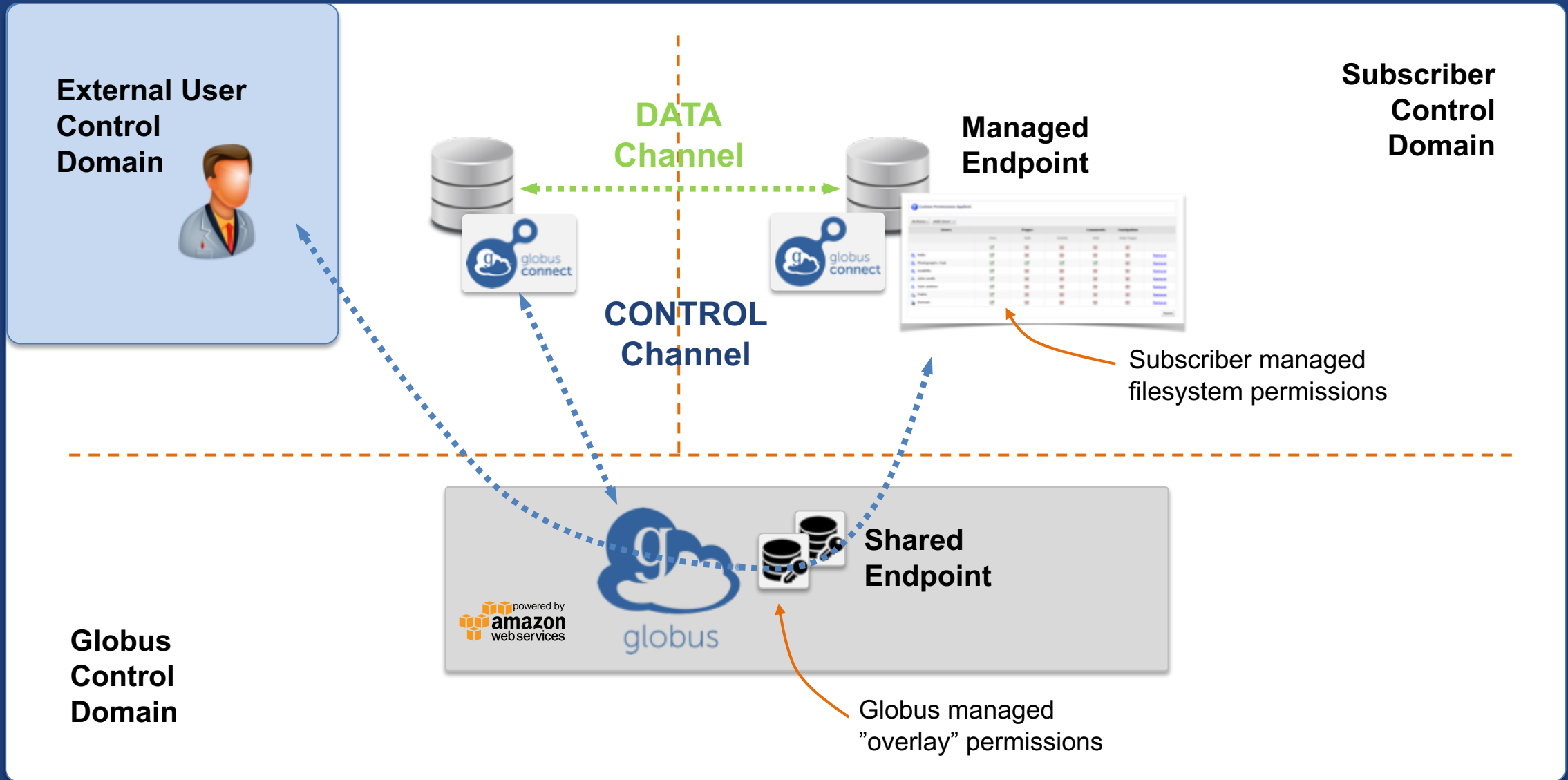


# Conceptual architecture: Hybrid SaaS





# 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**



Demonstration  
**File Transfer**  
**File Sharing**  
**Group Management**



# Data Publication and Discovery

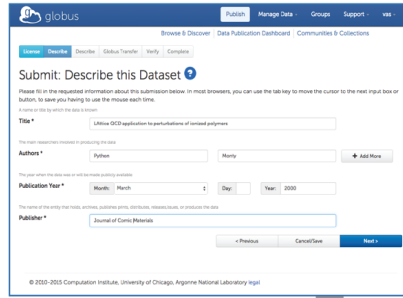
The screenshot shows the Materials Data Facility (MDF) community home page. At the top, the Globus logo and name are visible, along with "Log In" and "Sign Up" links. A blue banner below the header reads: "To submit a dataset or view datasets that have restricted access, please log in." Below this is a search bar with the text "Search" and a magnifying glass icon. The main content area features the MDF logo, which consists of a cluster of colorful circles (blue, green, yellow, orange, red) surrounding the text "MATERIALS DATA FACILITY". Below the logo, the text reads: "Materials Data Facility Community home page". A paragraph describes the MDF as a scalable repository for materials scientists. Another paragraph 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>

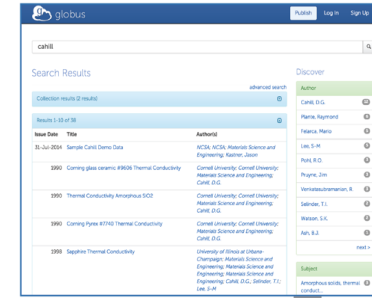




# Publish

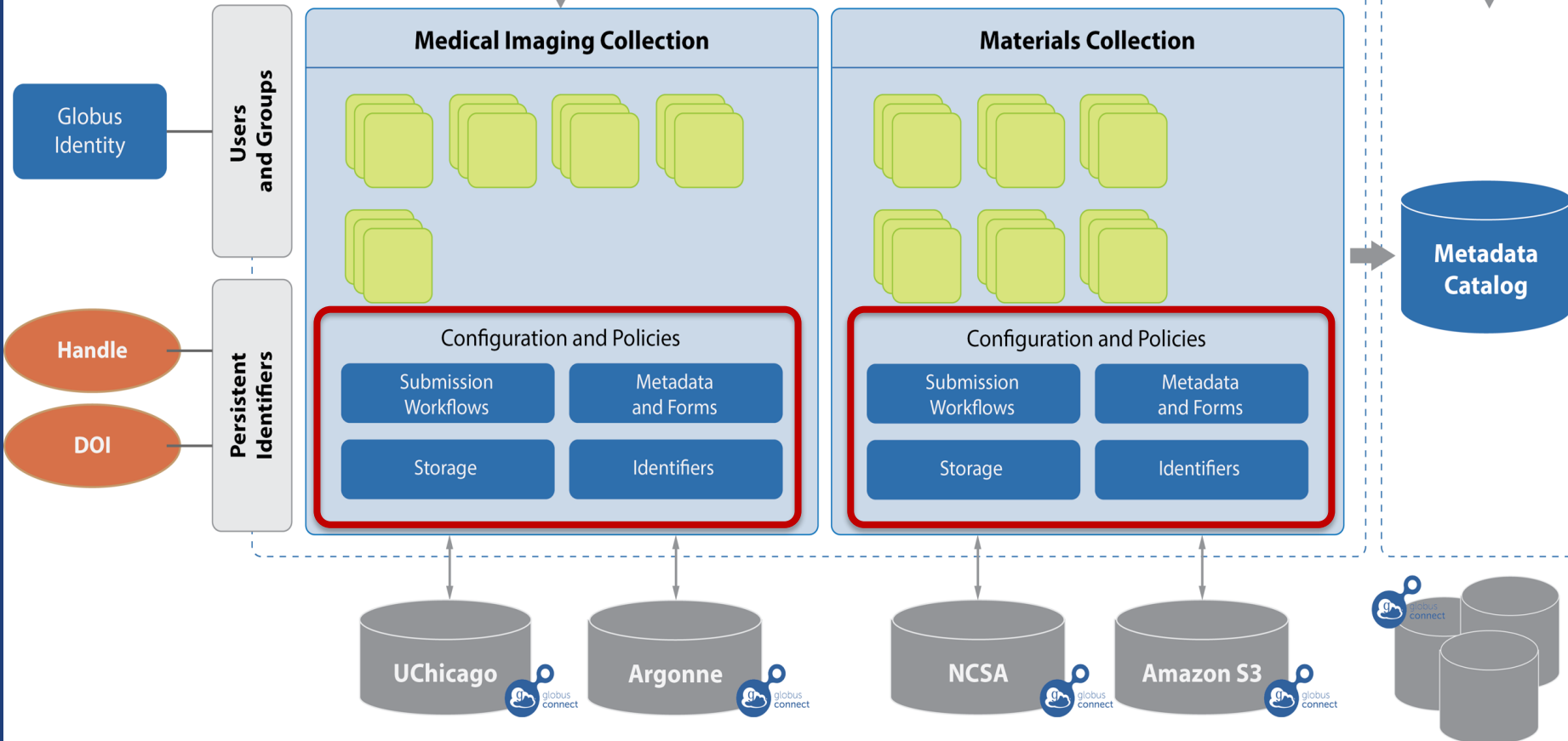


# Discover



## Globus Authentication

### Globus Data Publication





# Demonstration Data Publication



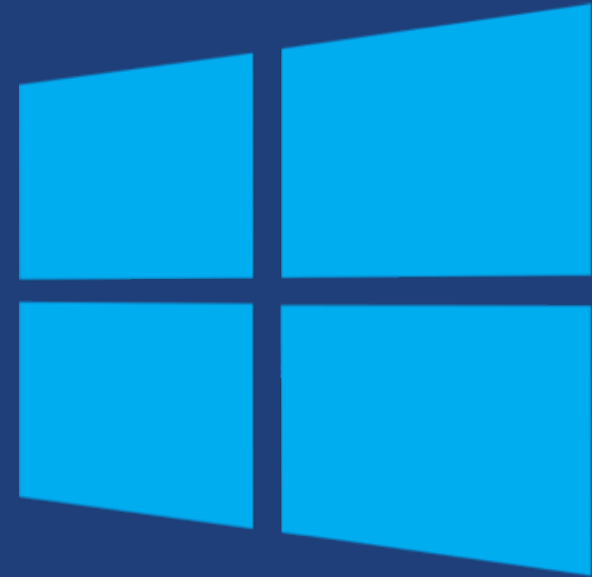
**How can I use Globus  
on my computer?**



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



# Globus Connect Personal



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



**Moving data between  
your laptop and  
another system**



# Exercise: Log in & transfer files

- **Go to: [www.globus.org/login](http://www.globus.org/login)**
- **Select your institution from the list and click “Continue” (or log in with a Globus ID)**
- **Authenticate with your institution’s identity system**
- **Install Globus Connect Personal**
- **Move some data between an ESnet test endpoint and your laptop**



# Sharing Data





# Exercise: Share files

- **Join the “Tutorial Users” groups**
  - Go to “Groups”, search for “tutorial”
  - Select group from list, click “Join Group”
- **Create a shared endpoint on your laptop**
- **Grant your neighbor permissions on your shared endpoint**
- **Access your neighbor’s shared endpoint**



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



**Globus serves as...**

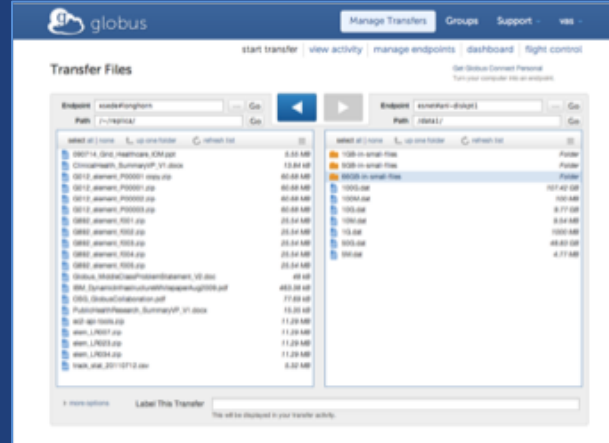
**...a platform for building  
science gateways, portals,  
and other web applications  
in support of research and  
education**



# 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.
```

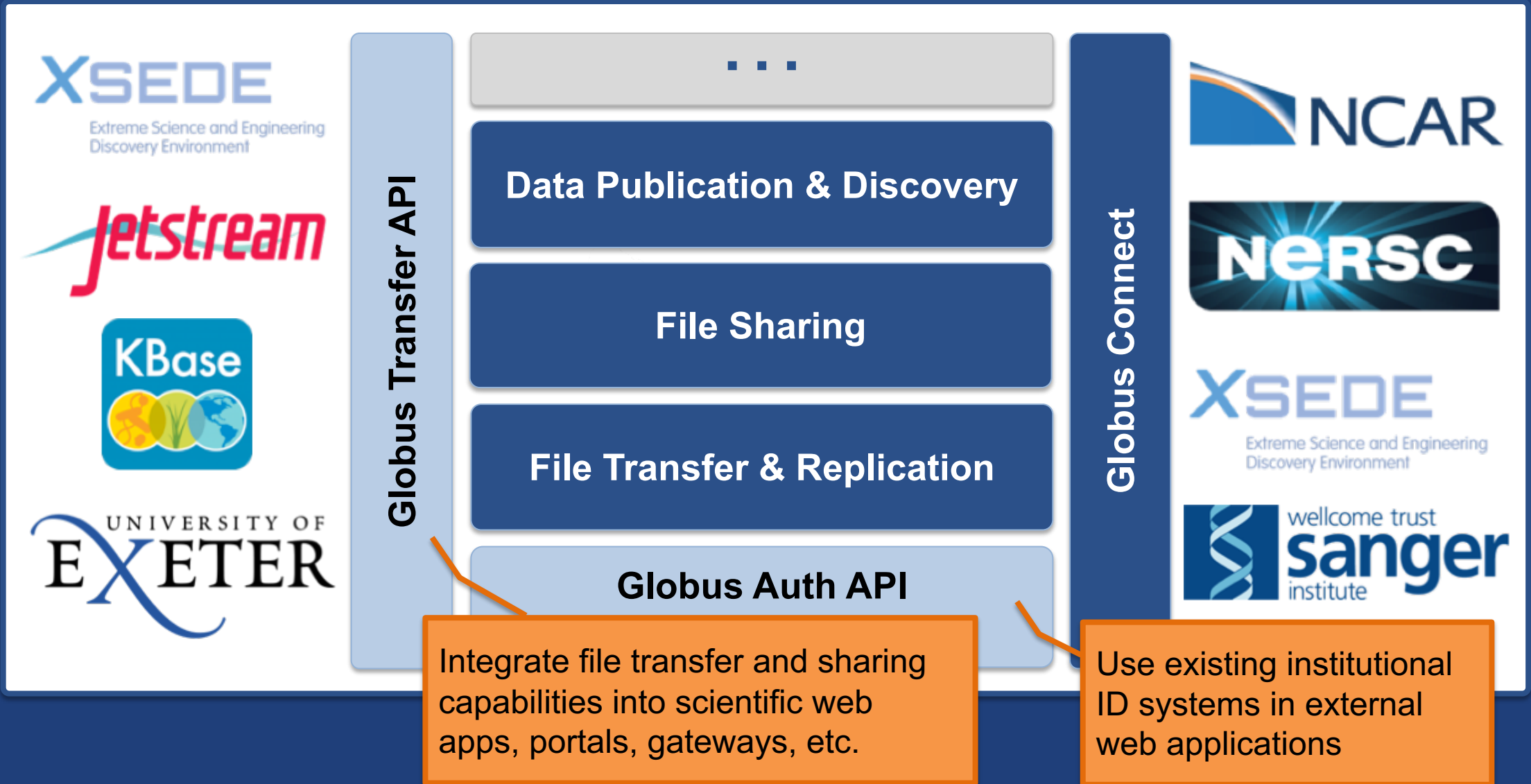
CLI

```
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






# Data Portal: NCAR RDA


UCAR NCAR Closures/Emergencies Locations/Directions Find Pe

Hello [tucke@uchicago.edu](#) [dashboard](#) [sign out](#)

NCAR UCAR |  **Research Data Archive**  
Computational & Information Systems Lab *weather • data • climate*

Go to Dataset:

[Home](#) [Find Data](#) [Ancillary Services](#) [About/Contact](#) [Data Citation](#) [Web Services](#) [For Staff](#)

 **NCEP Climate Forecast System Version 2 (CFSv2) Monthly Products**  
ds094.2

For assistance, contact [Bob Dattore](#) (303-497-1825).

[Description](#) [Data Access](#)

Mouse over the table headings for detailed descriptions

Data Description		Data File Downloads		Customizable Data Requests	Other Access Methods	NCAR-Only Access	
		Web Server Holdings	<b>Globus Transfer Service (GridFTP)</b>	Subsetting	THREDDS Data Server	Central File System (GLADE) Holdings	Tape Archive (HPSS) Holdings
Union of Available Products		Web File Listing	Request Globus Invitation	Get a Subset	TDS Access	GLADE File Listing	HPSS File Listing
P R O D	Diurnal monthly means	Web File Listing		Get a Subset		GLADE File Listing	HPSS File Listing
	Regular monthly means	Web File Listing		Get a Subset		GLADE File Listing	HPSS File Listing



# Data Distribution: ARM Climate Research Facility

### Data Selection Summary

Signed in as ANANTHAKRISHNANR1.

mergesonde1mace c1 @ fkb M1 [Generate Citation](#) 274 file(s) // 6014 MB

Order Complete Datastream  Extract Specific Measurements

Note: All variables will be delivered for this datastream.

**Measurement** : Atmospheric temperature  
**Variable** : Temperature // temp

2007-04-01

2007-12-31

Combine files by datastream

File format

Remove data flagged by Data Quality Reports (DQR) of type  Incorrect  Suspect

**Data Delivery Options**

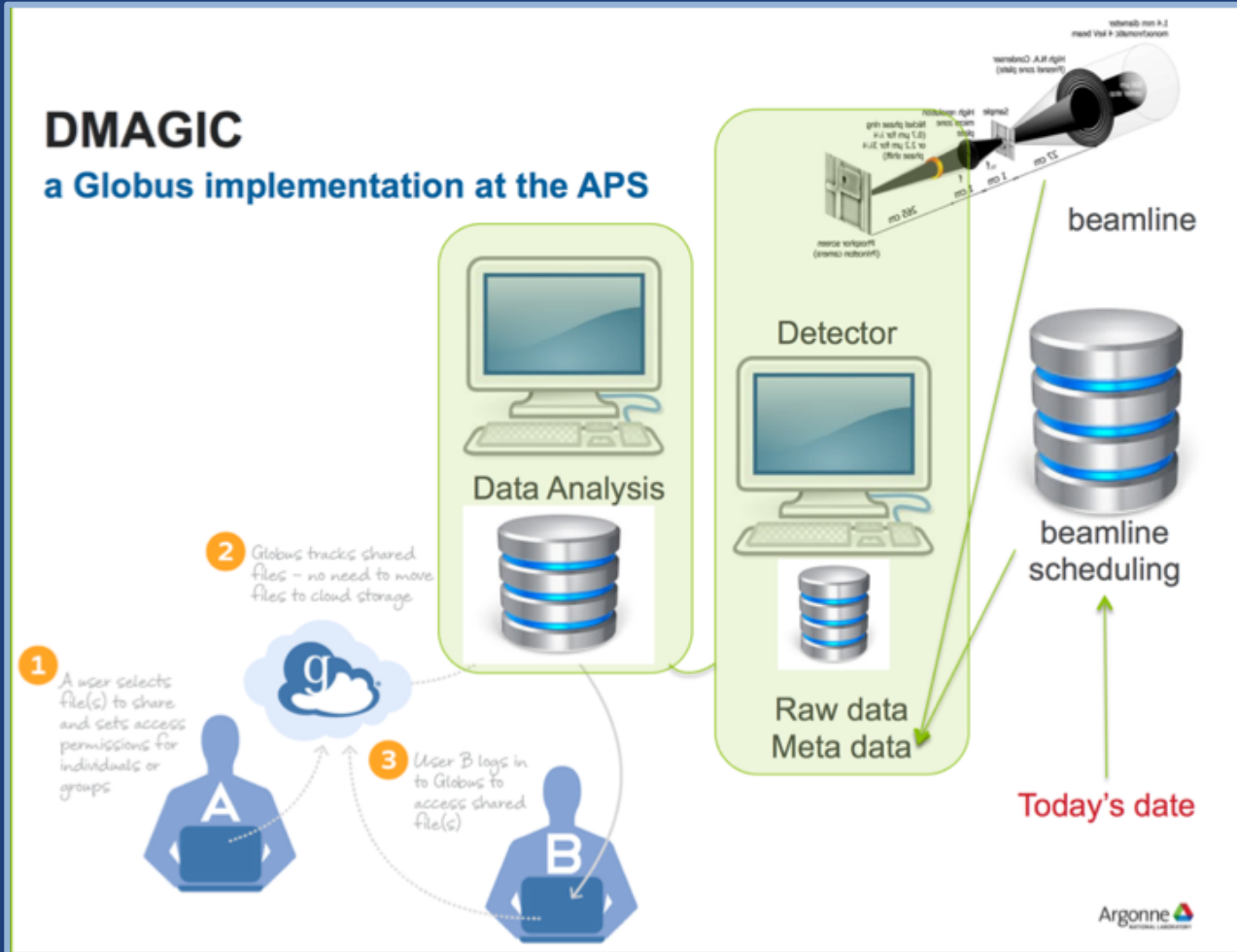
- FTP
- Globus
- THREDDS
- Dropbox

Extraction options only apply when "Extract Specific Measurements" is selected.

Original files will be delivered as part of all orders.



# Data Distribution: APS - DMagic



[dmagic.readthedocs.io](https://dmagic.readthedocs.io)

DMagic

latest

Search docs

About DMagic

Install directions

Development

API reference

Examples

Frequently asked questions

Docs > DMagic

Edit on GitHub

DMagic is an open-sourced Python toolbox to perform data management and data sharing for users of the Imaging Group of the Advanced Photon Source.





# Analysis App: Wellcome Sanger

Sanger Imputation Service **Beta**

Home About Instructions Resources Status

## Sanger Imputation Service

This is a free genotype **imputation** and **phasing** service provided by the [Wellcome Trust Sanger Institute](#). You can upload GWAS data in VCF or 23andMe format and receive imputed and phased genomes back. Click [here](#) to learn more and [follow us on Twitter](#).

### Before you start

Be sure to [read through the instructions](#).

You will need to set up a free account with [Globus](#) and have [Globus Connect](#) running at your institute or on your computer to transfer files to and from the service.

### Ready to start?

If you are ready to upload your data, please fill in the details below to **register an imputation and/or phasing job**. If you need more information, see the [about](#) page.

Full name

Organisation

Email address

What is this [?](#)

Globus user identity

[Next](#)

### News [@sangerimpute](#)

**11/05/2016**  
Thanks to [EAGLE](#), we can now return **phased data**. The HRC panel has been updated to r1.1 to fix a [known issue](#). See [ChangeLog](#) for more details.

**15/02/2016**  
Globus API changed, please see [updated instructions](#).

**17/12/2015**  
New status page and reworked internals. See [ChangeLog](#).

**09/11/2015**  
Pipeline updated to add some features requested by users. See [ChangeLog](#).

[See older news...](#)



# National Resource Access

The image shows two overlapping web pages. The background page is the XSEDE website, featuring a dark blue header with the XSEDE logo and the text "Extreme Science and Engineering Discovery Environment". Below the header, it says "Jetstream Web App would like to:" followed by a green checkmark icon and the text "Access all Jetstream resources". Below this, there is a paragraph of text and two buttons: "Allow" and "Deny".

The foreground page is the Compute Canada login page. It has a colorful geometric pattern on the left side. The header includes the "globus" logo and "Globus Account Log In". The main heading is "compute | calcul canada" with "canada" below each word. The text reads: "Log in to use Compute Canada Globus Web App" and "Use your existing organizational login". Below this, it lists examples of organizational logins: "e.g. university, national lab, facility, project, Google or [Globus ID](#)". A note states: "(Your Globus username and password used prior to February 13, 2016 is now Globus ID)". There is a dropdown menu showing "WestGrid" and a blue "Continue" button. At the bottom, it says: "Didn't find your organization? Then use Globus ID to [sign up](#)."



# Identity Management



DOE Systems Biology Knowledgebase

[Home](#) [About](#) [News](#) [Developer Zone](#) [KBase Labs](#)

The new **Systems Biology Knowledgebase (KBase)** is a collaborative effort designed to accelerate our understanding of microbes, microbial communities, and plants. It will be a community-driven, extensible and scalable open source software framework and application system. KBase will offer free and open access to data, models and simulations, enabling scientists and researchers to build new knowledge and share their findings.

Collaborate with us

Get Started

Develop with us

## What can KBase do?

- ✔ Combine heterogeneous data types
- ✔ Offer standardized access to bioinformatic and modeling analyses
- ✔ Use evidence-supported annotations of genome structure and genetic function
- ✔ Discover new associations and network structures in community and molecular networks
- ✔ Map genotype to complex organismal traits
- ✔ Design and refine experiments using models of metabolism, regulation and community function
- ✔ Enable sharing of data, hypotheses, and newly-generated knowledge



Accelerate research on campus by providing connective services for local, cloud and national cyberinfrastructure

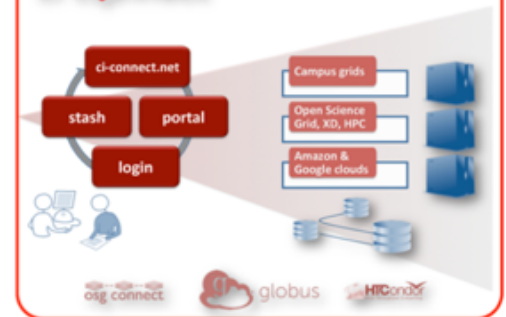
## Connecting Science to Cycles and Data

CI Connect offers campus IT organizations connective services for cyberinfrastructure. End-users quickly gain access to shared campus clusters while gaining access to distributed resources of national HPC or HTC facilities (such as the [Open Science Grid](#)) or cloud-based resources. CI Connect offers several capabilities based on [Globus](#), [HTCondor](#) and other HTC and data sharing technologies that augment your vision of providing scientific computational resources to research communities you support.

## Reliable, High Performance CI Services

CI Connect provides services to CI providers at US universities and laboratories following a simple set of principles:

ci connect



## Connected environments from hosted services

Resources of a campus cluster (or [campus grid](#)) can be

2013-02-25  
Proposed: Genomic Science  
Contractors-Grantees Meeting



# Globus PaaS developer resources

**Python SDK**

**Modern Research Data Portal**

**Requirements**

- You need to be in the tutorial users group for sharing: <https://www.globus.org/app/groups/50b6a29c-63e...>
- Installed Globus Python SDK

**Configuration**

```
In [15]: from __future__ import print_tutorial_endpoint_1 = "ddb59ae1-60v4-11e5-ba46-22000b92c6ec" # endpoint "Glo tutorial_endpoint_2 = "ddb59af0-6d04-11e5-ba46-22000b92c6ec" # endpoint "Glo tutorial_users_group = "50b6a29c-63ac-11e4-8062-22000ab68755" # group "Tutori
```

```
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

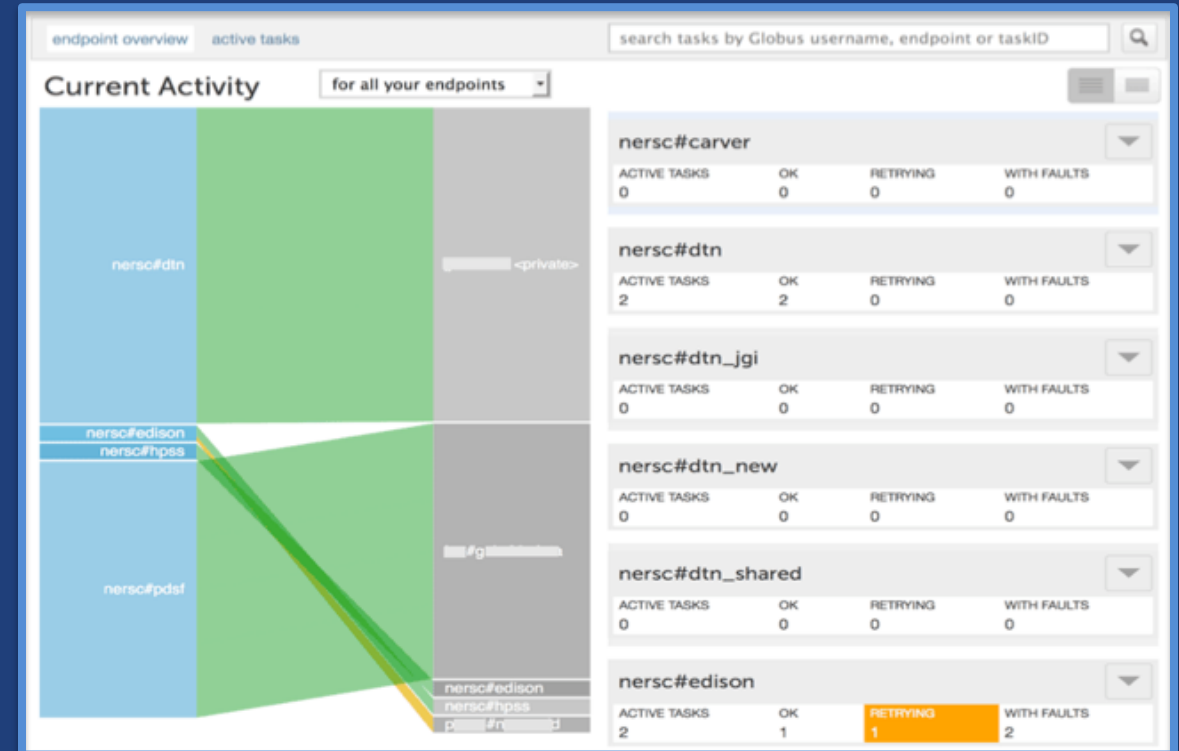
Argonne  
NATIONAL LABORATORY

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

**360 PB**  
transferred

**62 billion**  
tasks processed

**73,000**  
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

 ...and THANK YOU to our subscribers



JOHNS HOPKINS  
UNIVERSITY



Yale



HARVARD  
UNIVERSITY



CORNELL  
UNIVERSITY



VirginiaTech  
*Invent the Future*



THE UNIVERSITY OF  
CHICAGO



MICHIGAN STATE  
UNIVERSITY



Stanford  
University



Dartmouth

syngenta

NIST

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)



# Globus for System Administrators

Vas Vasiliadis  
[vas@uchicago.edu](mailto:vas@uchicago.edu)

Rachana Ananthakrishnan  
[rachana@globus.org](mailto:rachana@globus.org)

Stanford University – February 8, 2018





Get IP address: **[bit.ly/ec2ip](https://bit.ly/ec2ip)**

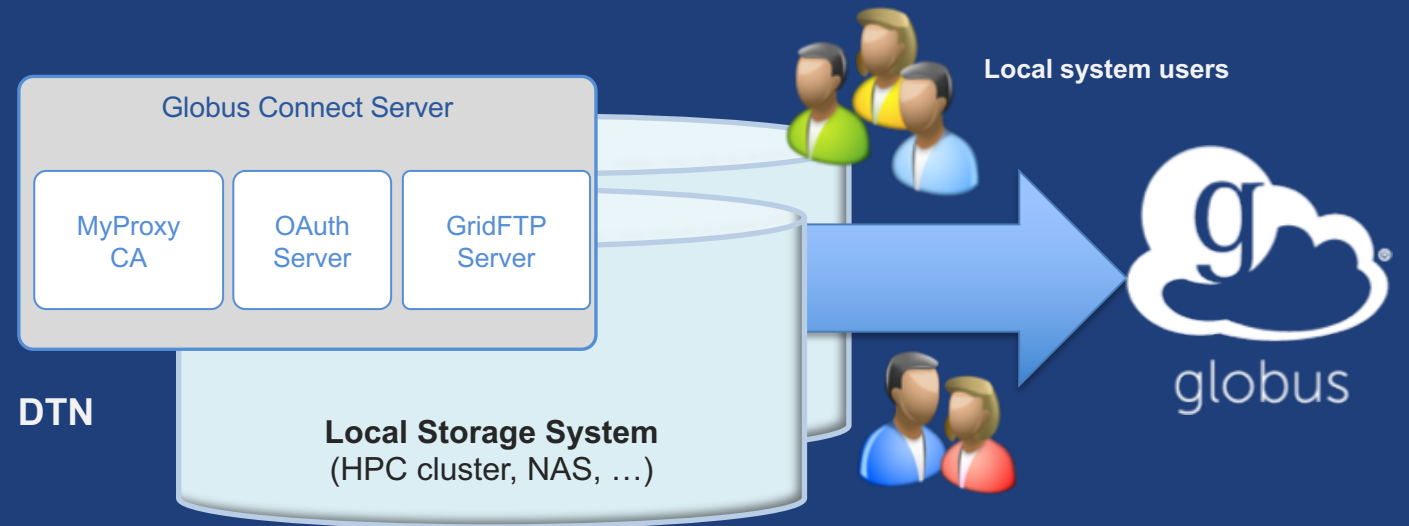
**'campusadmin'** - sudo privileges

**'researcher'** - regular user



# Globus Connect Server

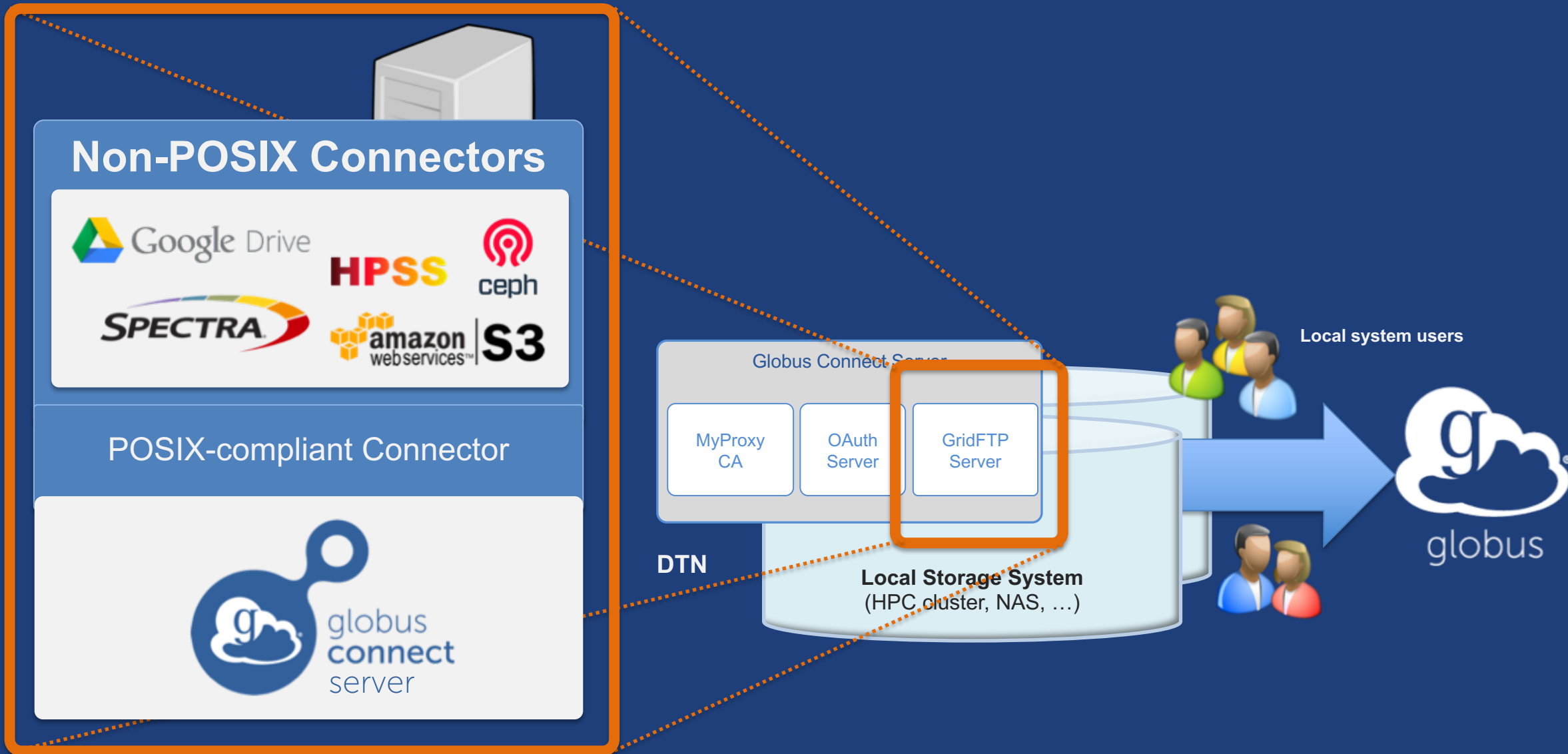
- **Makes your storage accessible via Globus**
- **Multi-user server, installed and managed by sysadmin**
- **Default access for all local accounts**
- **POSIX + connectors**
- **Native packaging  
Linux: DEB, RPM**



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



# Globus Connect Server

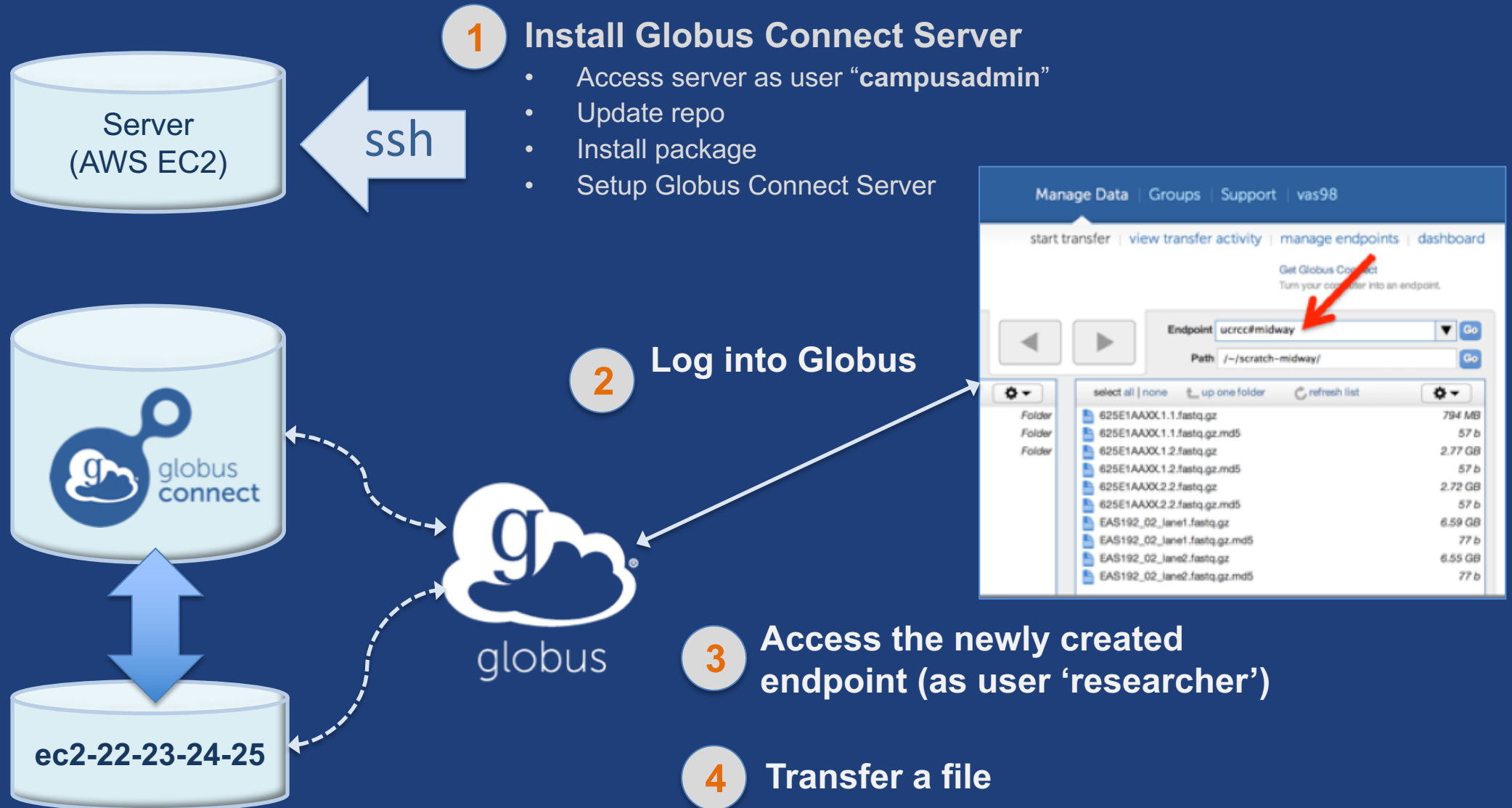


# Creating a Globus endpoint on your server

- **In this example, server = Amazon EC2 instance**
- **Installation and configuration of Globus Connect Server requires a Globus ID**
- **Go to `globusid.org`**
- **Click “create a Globus ID”**
  - Optional: associate it with your Globus account



# What we are going to do:





# Access your host

- **Get the IP address for your EC2 server**
- **Log in as user 'campusadmin':**  
`ssh campusadmin@<EC2_instance_IP_address>`
- **NB: Please sudo su before continuing**
  - User 'campusadmin' has sudo privileges





# Install Globus Connect Server

```
$ sudo su
$ curl -L0s http://toolkit.globus.org/ftppub/globus-
connect-server/globus-connect-server-
repo_latest_all.deb
$ dpkg -i globus-connect-server-repo_latest_all.deb
$ apt-get update
$ apt-get -y install globus-connect-server
$ globus-connect-server-setup
```

↑ Use your Globus ID username/password when prompted

**You have a working Globus endpoint!**



# Access the Globus endpoint

- **Go to Manage Data → Transfer Files**
- **Access the endpoint you just created**
  - Search for your EC2 host name in the Endpoint field
  - Log in as “researcher”; you will see the user’s home directory
- **Transfer files to/from a test endpoint (e.g. Globus Tutorial) and your EC2 endpoint**



# Configuring Globus Connect Server

# Endpoint configuration

- **Globus service “Manage Endpoints” page**
- **DTN (Globus Connect Server) config**
  - `/etc/globus-connect-server.conf`
  - Standard .ini format: `[Section] Option = value`
  - To enable changes you must run:  
**`globus-connect-server-setup`**
  - “Rinse and repeat”



# Common configuration options

- **Manage Endpoints page**
  - Display Name
  - Visibility
  - Encryption
- **DTN configuration file – common options:**
  - RestrictPaths
  - IdentityMethod (CILogon, OAuth)
  - Sharing
  - SharingRestrictPaths



# Exercise: Make your endpoint visible

- **Edit endpoint attributes**
  - Change the name to something useful, e.g. <your\_name> EC2 Endpoint
  - For the “Visible To” attribute select “Public - Visible to all users”
- **Find your neighbor’s endpoint**
  - You can access it too 😊



# Path Restriction

- **Default configuration:**
  - All paths allowed, access control handled by the OS
- **Use RestrictPaths to customize**
  - Specifies a comma separated list of full paths that clients may access
  - Each path may be prefixed by R (read) and/or W (write), or N (none) to explicitly deny access to a path
  - '~' for authenticated user's home directory, and \* may be used for simple wildcard matching.
- **e.g. Full access to home directory, read access to /data:**
  - RestrictPaths = RW~,R/data
- **e.g. Full access to home directory, deny hidden files:**
  - RestrictPaths = RW~,N~/.\*

## Exercise: Restrict access

- **Set RestrictPaths=RW~,N~/archive**
- **Run globus-connect-server-setup**
- **Access your endpoint as 'researcher'**
- **What's changed?**



# Enabling sharing on an endpoint

- In config file, set `Sharing=True`
- Run `globus-connect-server-setup`
- Use the CLI to flag as managed endpoint (also configurable via the web app)

\* Note: Creation of shared endpoints requires a Globus subscription for the managed endpoint

# Limit sharing to specific accounts

- `SharingUsersAllow =`
- `SharingGroupsAllow =`
- `SharingUsersDeny =`
- `SharingGroupsDeny =`



# Sharing Path Restriction

- **Restrict paths where users can create shared endpoints**
- **Use `SharingRestrictPaths` to customize**
  - Same syntax as `RestrictPaths`
- **e.g. Full access to home directory, deny hidden files:**
  - `SharingRestrictPaths = RW~,N~/.*`
- **e.g. Full access to public folder under home directory:**
  - `SharingRestrictPaths = RW~/public`
- **e.g. Full access to `/proj`, read access to `/scratch`:**
  - `SharingRestrictPaths = RW/proj,R/scratch`



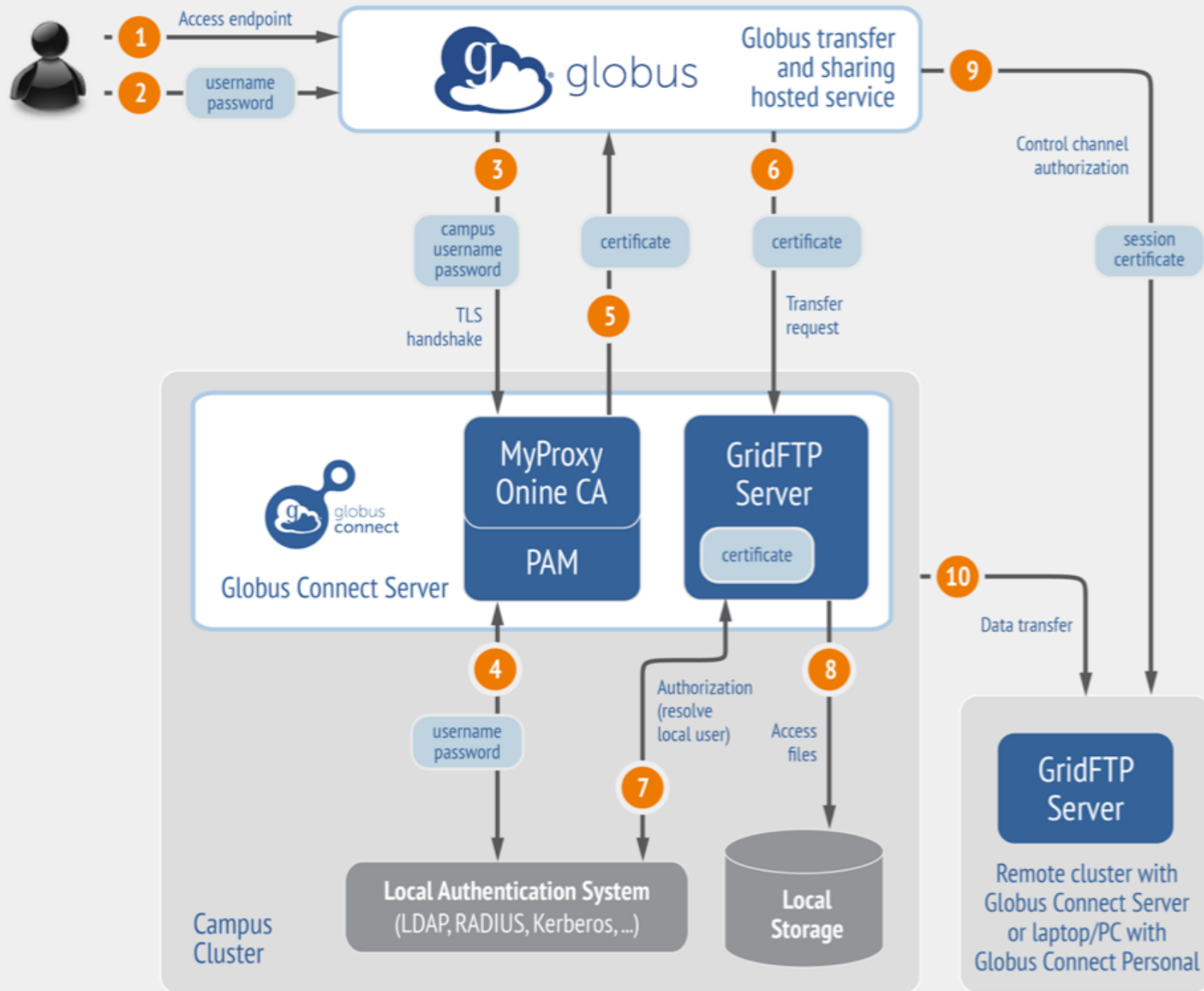
# Accessing Endpoints

# Ports needed for Globus

- **Inbound: 2811 (control channel)**
- **Inbound: 7512 (MyProxy), 443 (OAuth)**
- **Inbound: 50000-51000 (data channel)**
- **If restricting outbound connections, allow connections on:**
  - 80, 2223 (used during install/config)
  - 50000-51000 (GridFTP data channel)



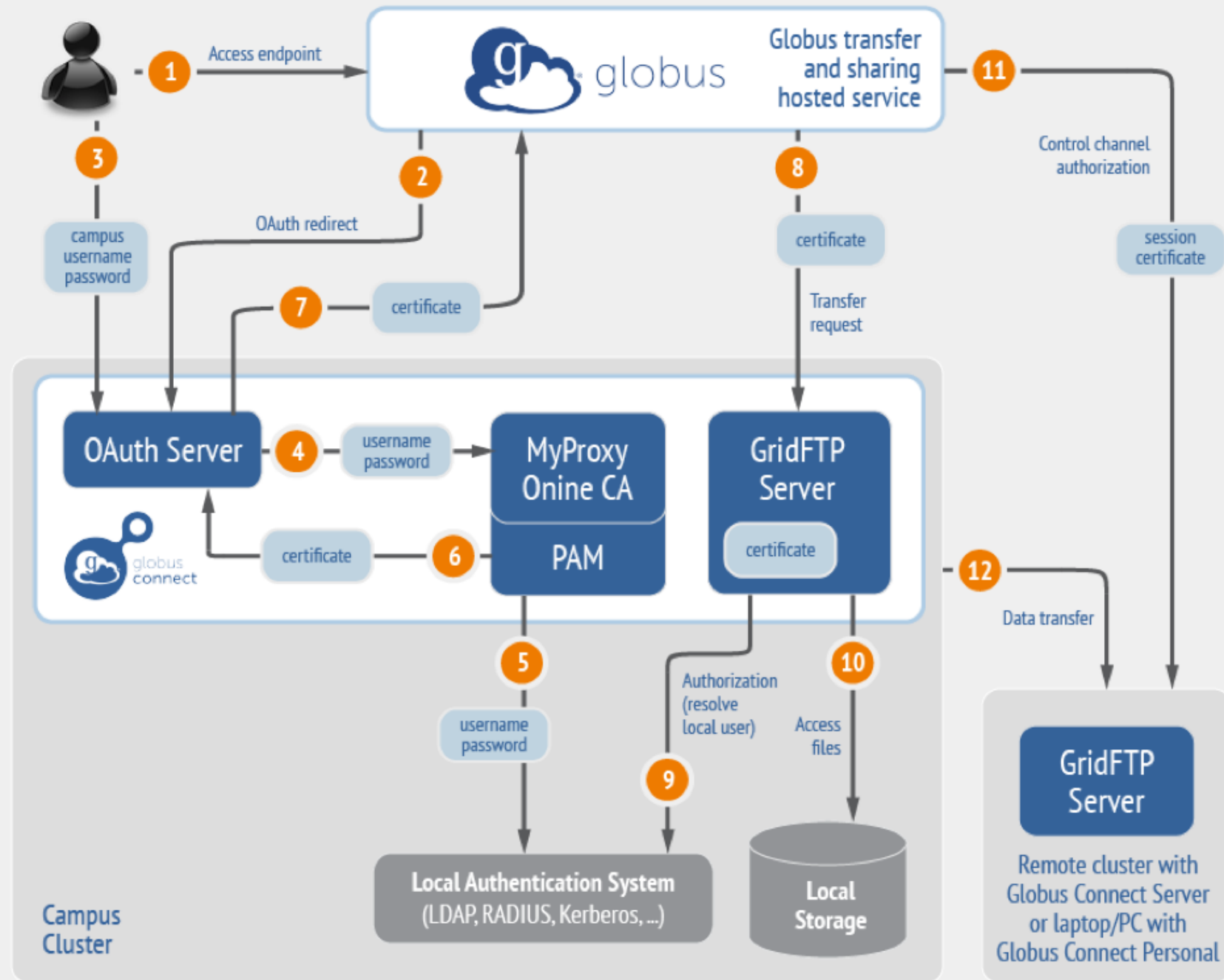
# Endpoint activation using MyProxy



**DON'T  
LEAVE IT  
LIKE THIS!**



# Endpoint activation using MyProxy OAuth



Yes, please do this!

# Single Sign-On with InCommon/CILogon

- **Your Shibboleth server must release R&S attributes to CILogon—especially the ePPN attribute**
- **Local resource account names must match your institutional ID (InCommon ID)**
- **In `/etc/globus-connect-server.conf` set:**

```
AuthorizationMethod = CILogon
```

```
CILogonIdentityProvider =  
<institution_listed_in_CILogon_IdP_list>
```





# Managed endpoints and subscriptions



# Subscription configuration

- **Subscription manager**
  - Create/upgrade managed endpoints
  - Requires Globus ID linked to Globus account
- **Management console permissions**
  - Independent of subscription manager
  - Map managed endpoint to Globus ID
- **Globus Plus group**
  - Subscription Manager is admin
  - Can grant admin rights to other members



# Creating managed endpoints

- **Required for sharing, management console, reporting, ...**
- **Convert existing endpoint to managed via CLI (or web):**  
`globus endpoint update --managed <endpt_uuid>`
- **Must be run by subscription manager**
- **Important: Re-run endpoint update after deleting/re-creating endpoint**



# Monitoring and managing Globus endpoint activity

# Management console

- **Monitor all transfers**
- **Pause/resume specific transfers**
- **Add pause conditions with various options**
- **Resume specific tasks overriding pause conditions**
- **Cancel tasks**
- **View sharing ACLs**



# Endpoint Roles

- **Administrator:** define endpoint and roles
- **Access Manager:** manage permissions
- **Activity Manager:** perform control tasks
- **Activity Monitor:** view activity



Demonstration:

**Management console**

**Endpoint Roles**

**Usage Reporting**



**...on performance**





# Balance: performance - reliability

- **Network use parameters: concurrency, parallelism**
- **Maximum, Preferred values for each**
- **Transfer considers source and destination endpoint settings**

```
min(  
    max(preferred src, preferred dest),  
    max src,  
    max dest  
)
```

- **Service limits, e.g. concurrent requests**

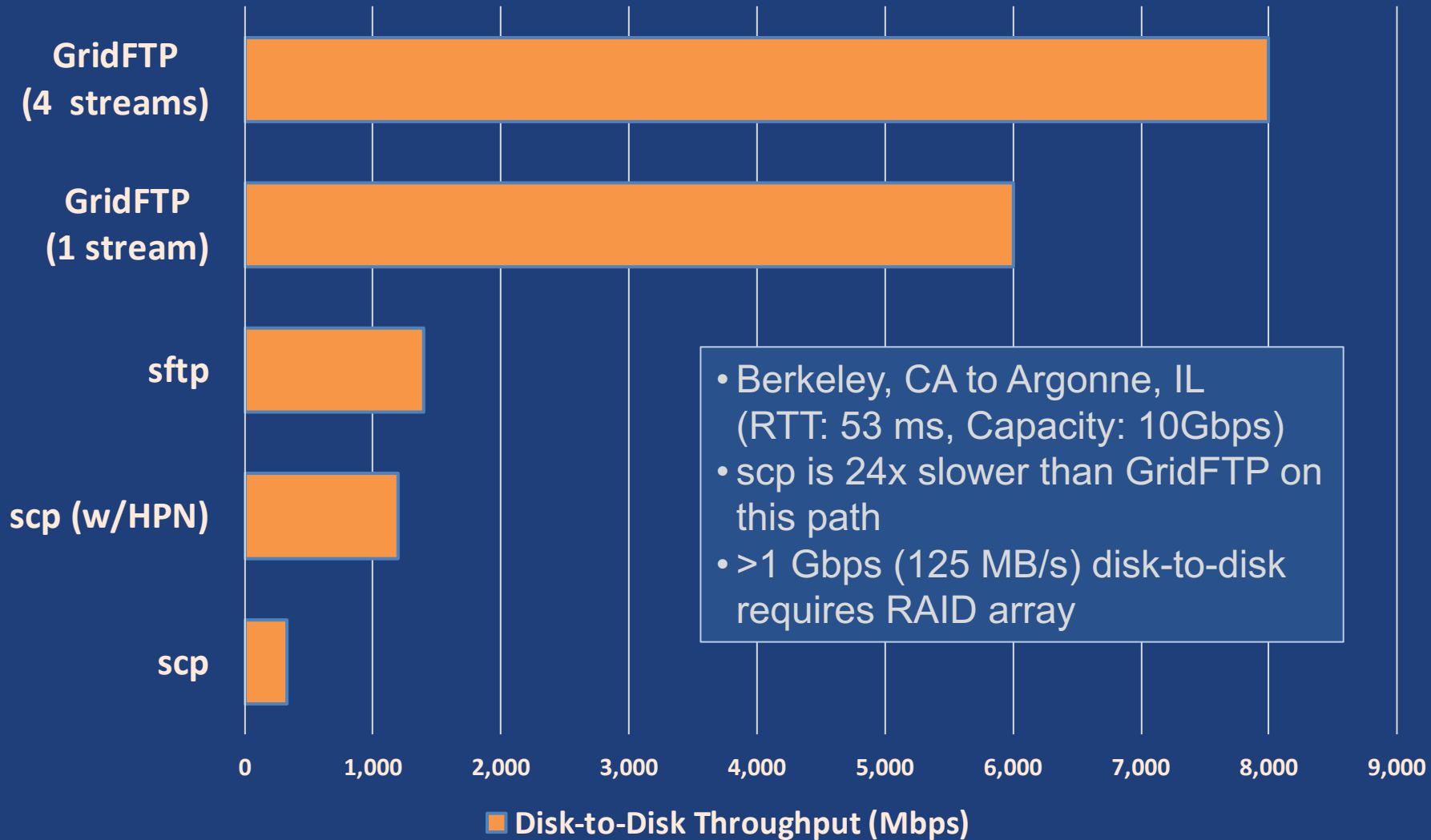


# Illustrative performance

- **20x scp throughput (typical)**
  - >100x demonstrated
- **On par/faster than UDP based tools (NASA JPL study and anecdotal)**
- **Capable of saturating “any” WAN link**
  - Demonstrated 85Gbps sustained disk-to-disk
  - Typically require throttling for QoS



# Disk-to-Disk Throughput: ESnet Testing

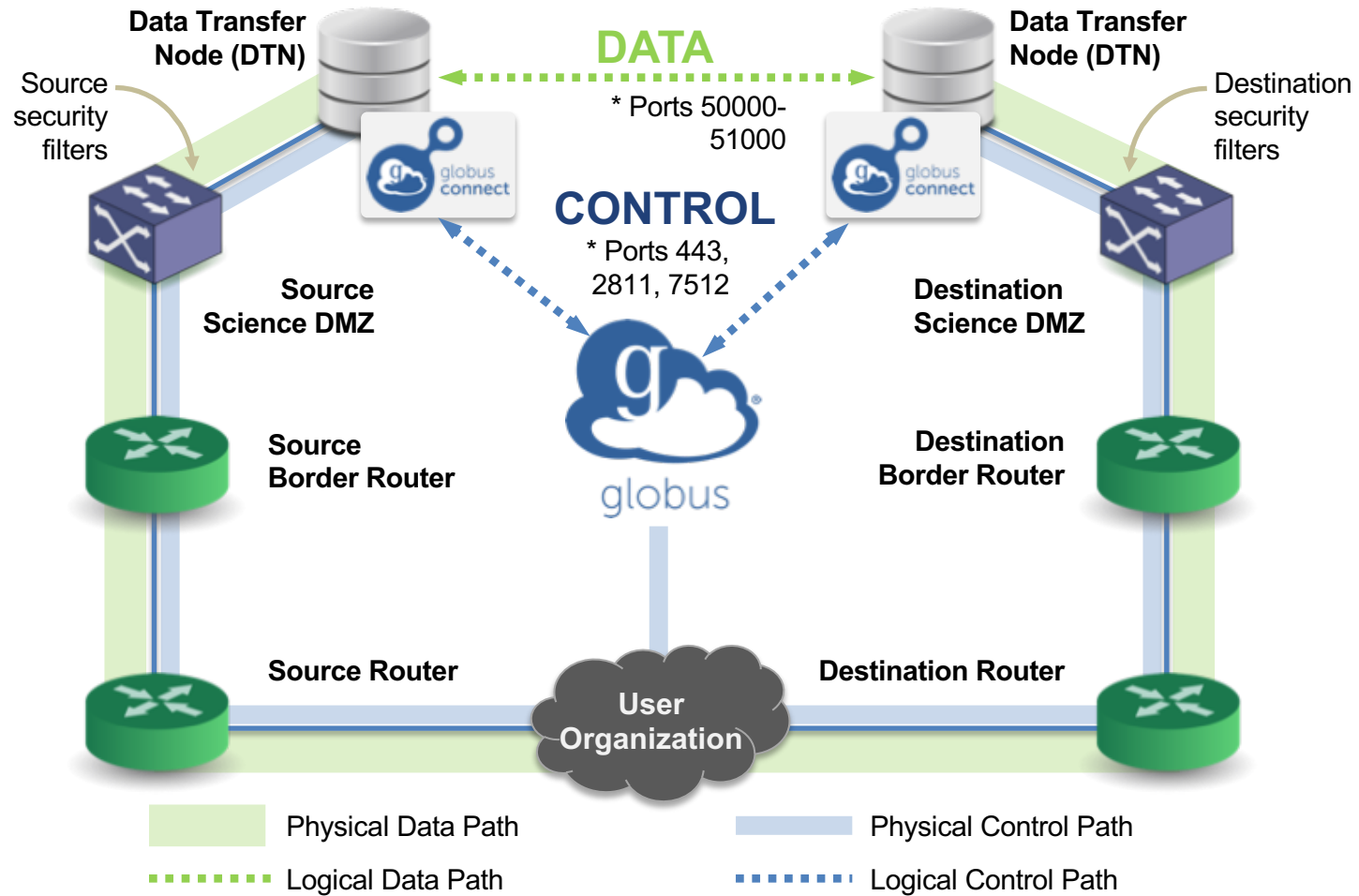




# Deployment Scenarios



# Best practice network configuration



\* Please see TCP ports reference: [https://docs.globus.org/resource-provider-guide/#open-tcp-ports\\_section](https://docs.globus.org/resource-provider-guide/#open-tcp-ports_section)



# The Data Transfer Node

On-premises  
Data Transfer  
Node (DTN)



Data Storage Interface  
(DSI) for  
POSIX-compliant  
filesystems

## Non-POSIX DSI

- Google Drive
- Amazon S3 (native)
- Spectra BlackPearl
- Ceph S3 RadosGW
- HPSS

Cloud-hosted  
DTN



AWS  
EBS  
Volume

AWS S3  
Bucket



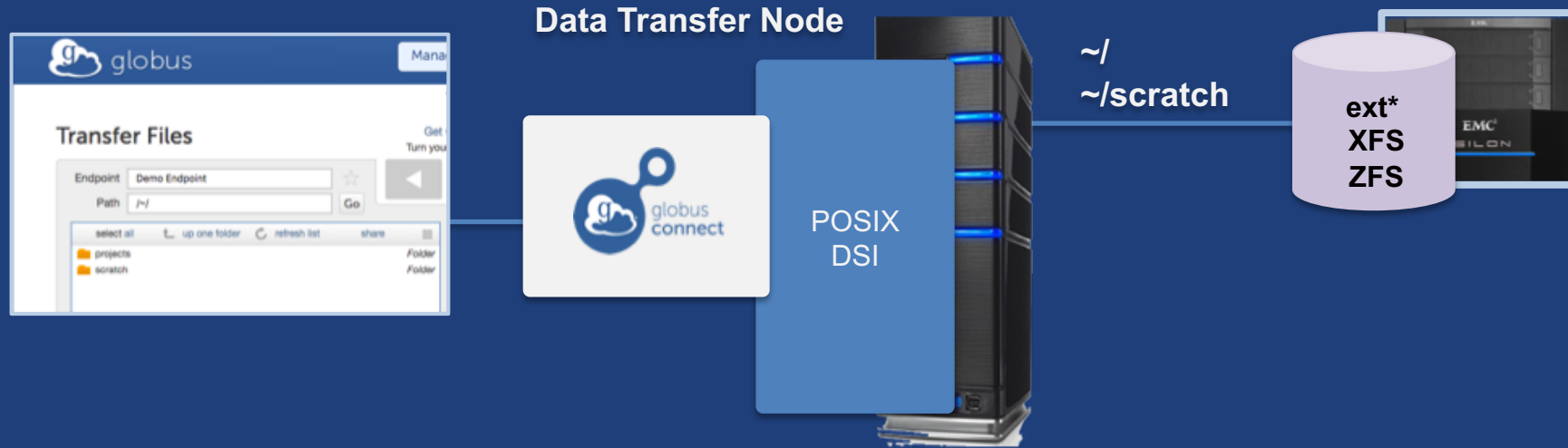
Data  
Storage  
Interface  
(DSI) for  
POSIX-  
compliant  
filesystems

## Non-POSIX DSI

- Google Drive
- Amazon S3 (native)
- Spectra BlackPearl
- Ceph S3 RadosGW
- HPSS

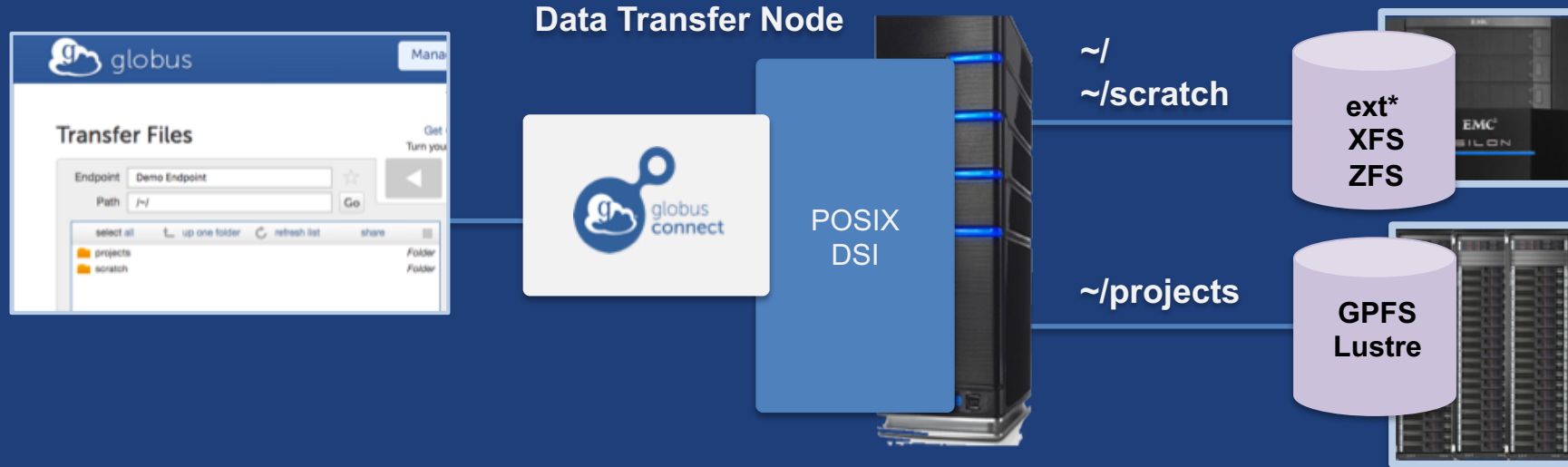


# Multi-endpoint configuration





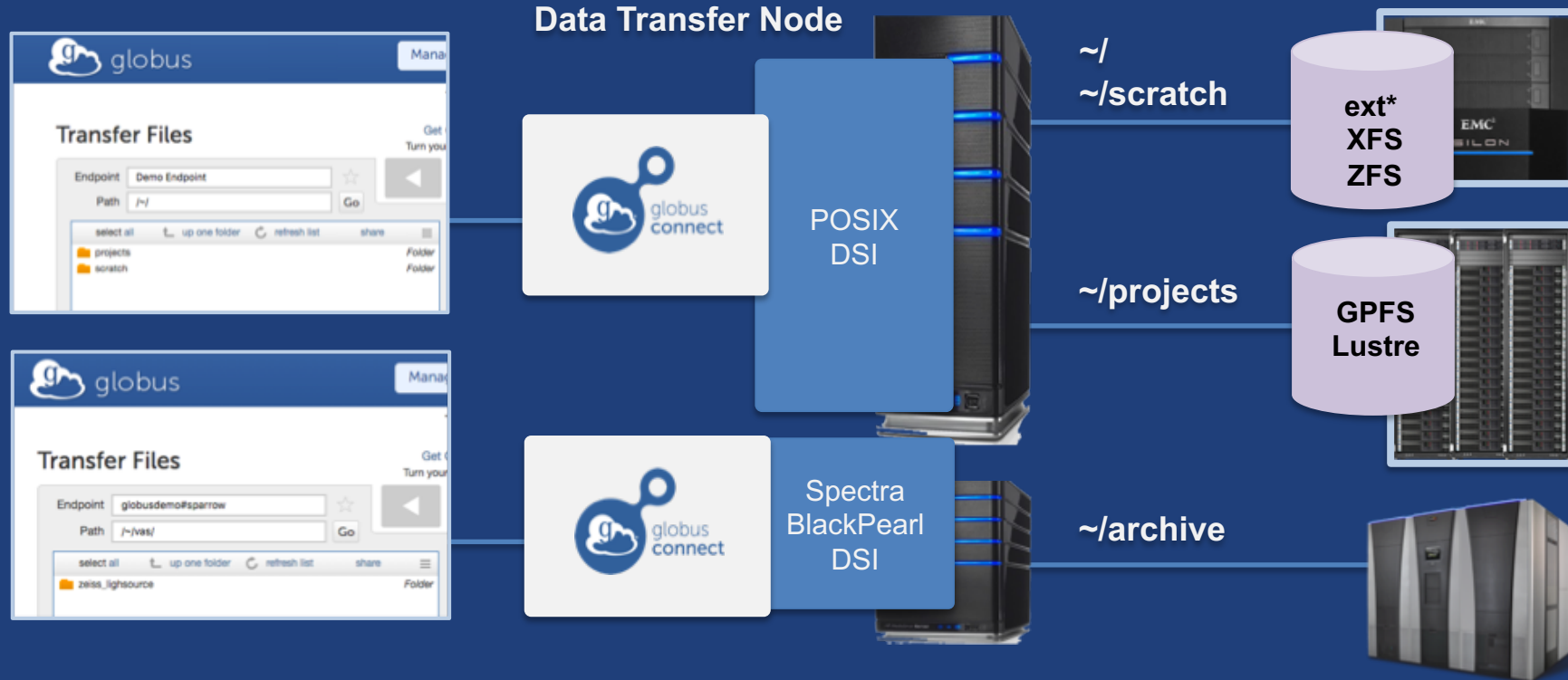
# Multi-endpoint configuration





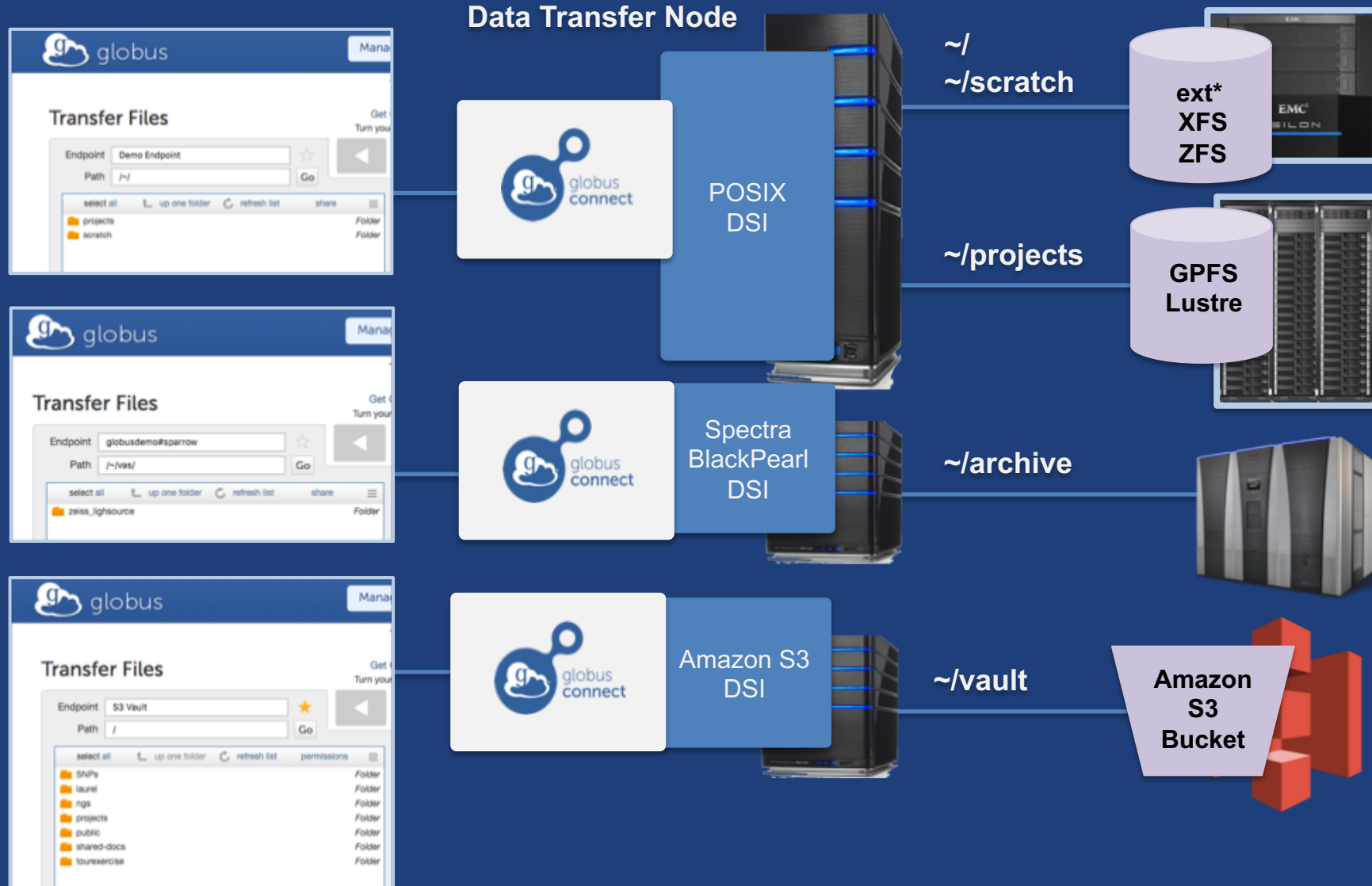


# Multi-endpoint configuration

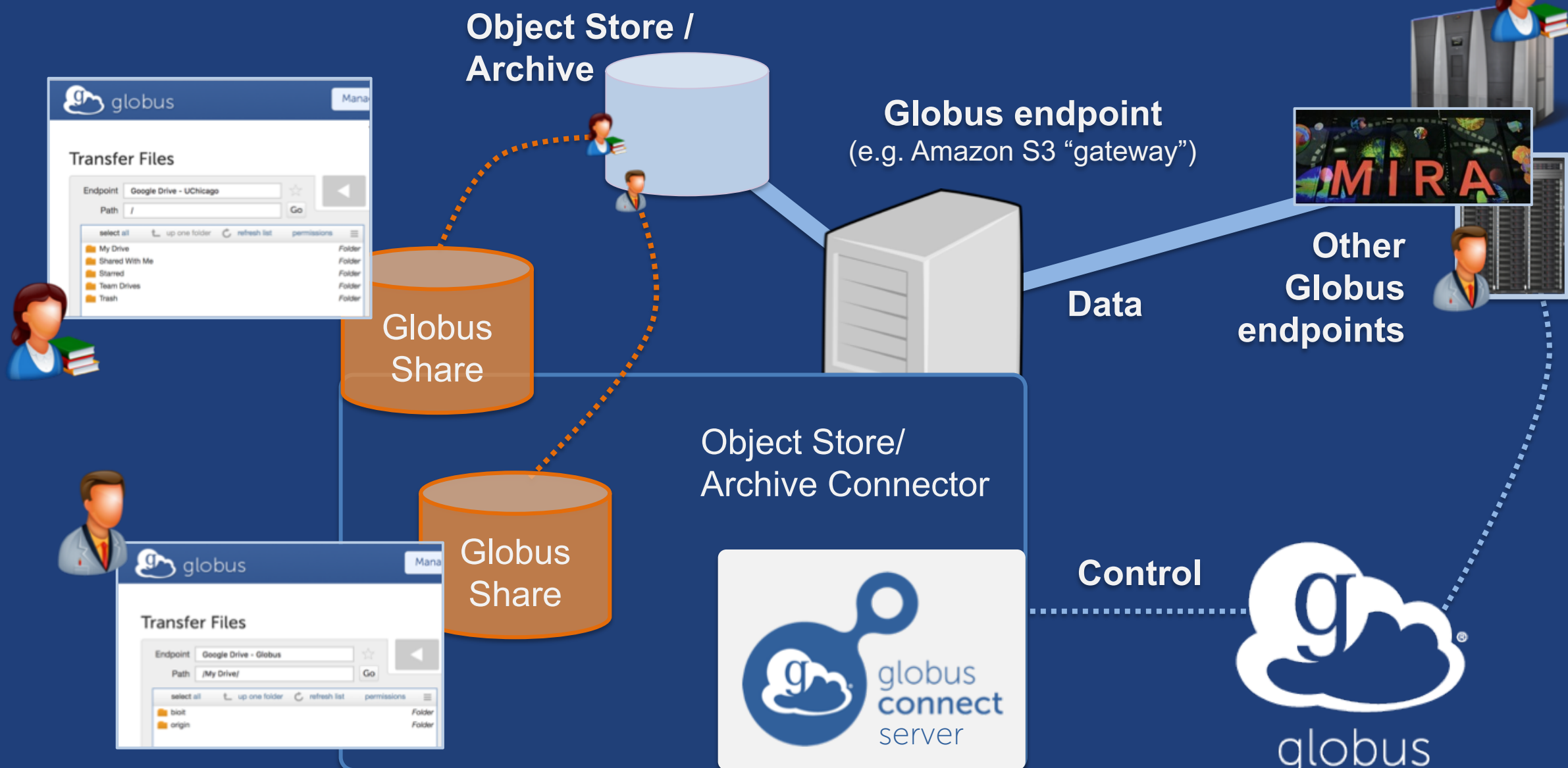




# Multi-endpoint configuration



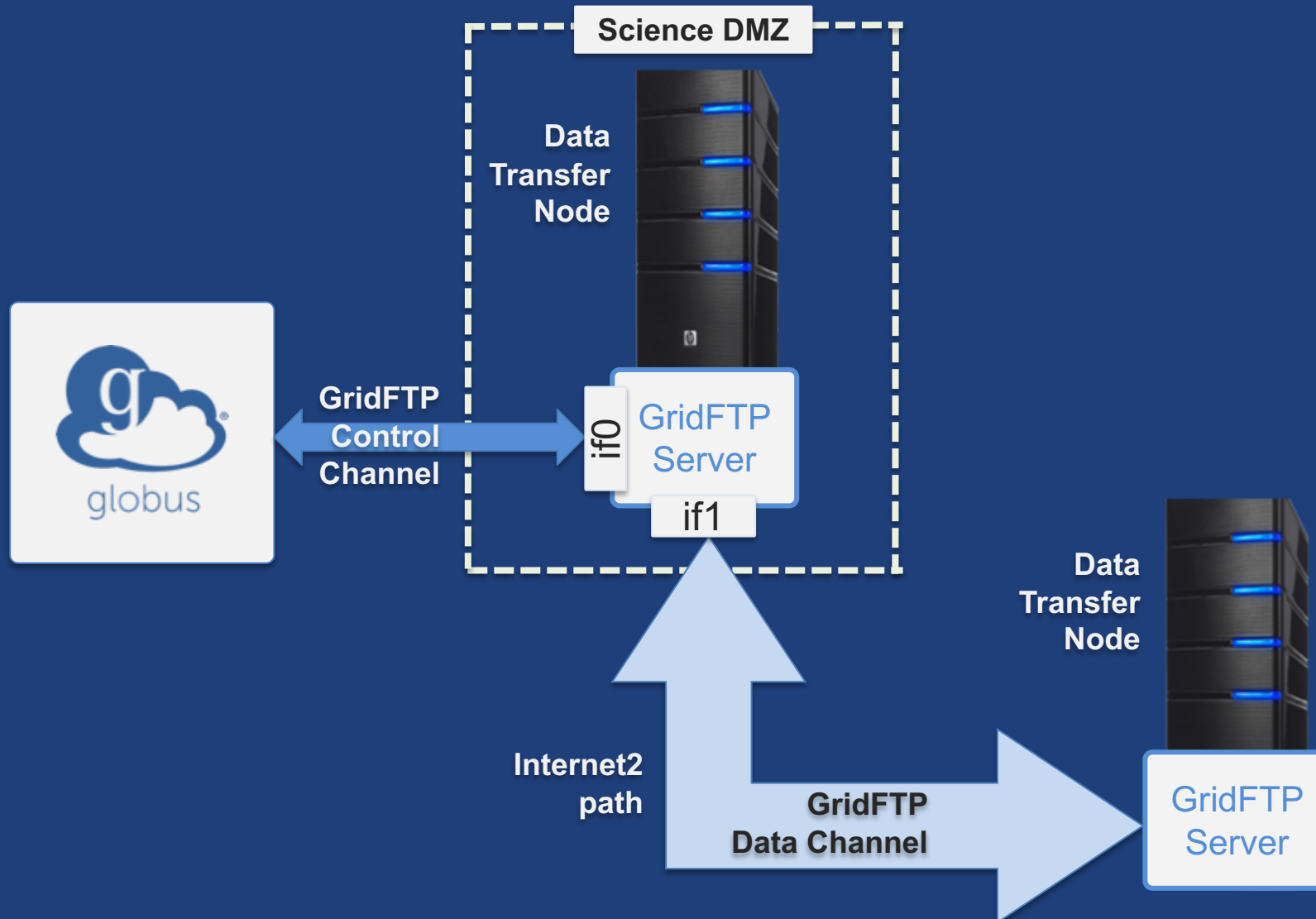
# Deploying a premium connector gateway



# Network paths

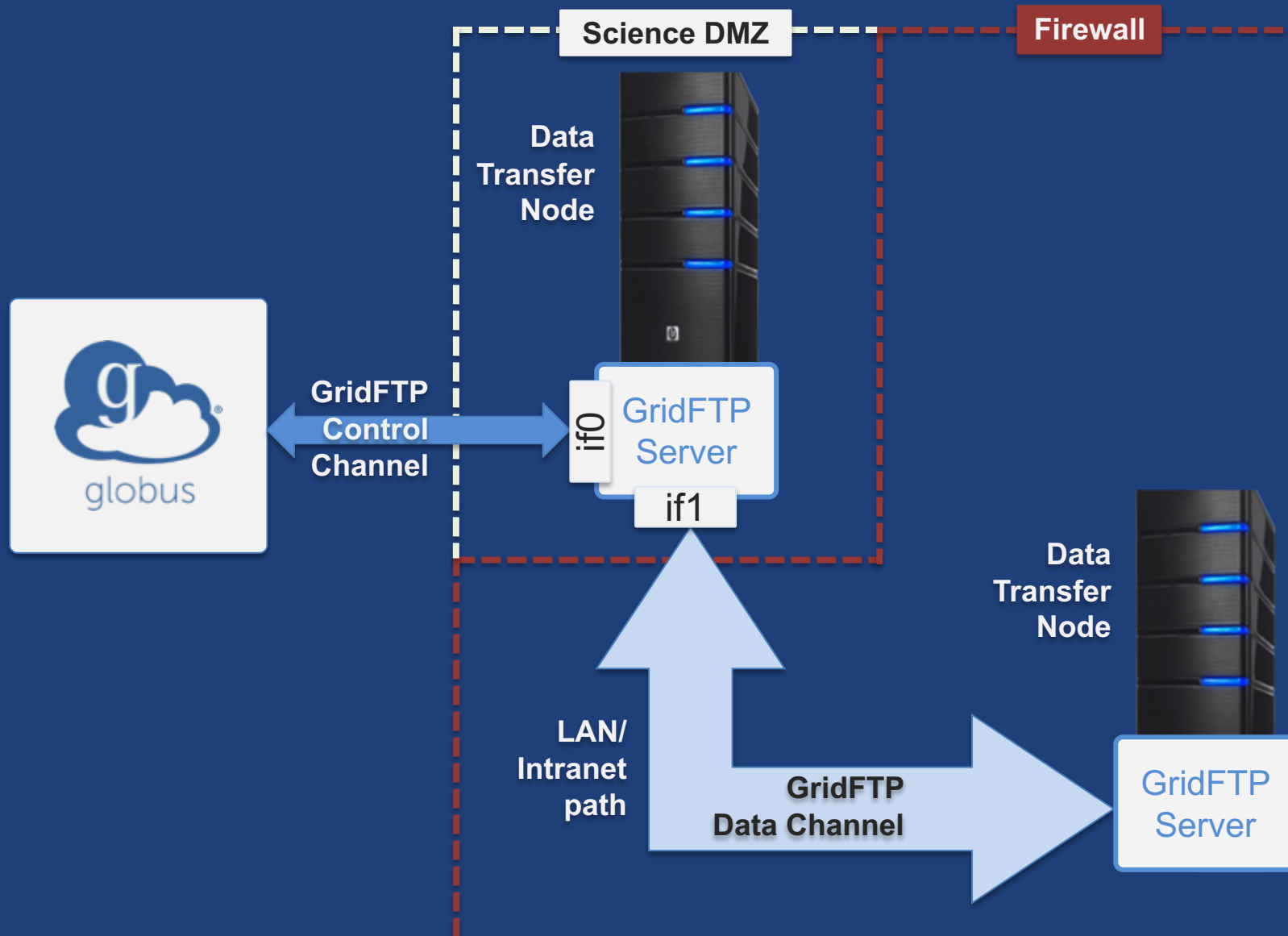
- **Separate control and data interfaces**
- **"DataInterface =" option in globus-connect-server-conf**
- **Common scenario: route data flows over Science DMZ link**

# Dual-homed DTN – high speed data path





# Dual-homed DTN – private network data path





# Other Deployment Options



# Encryption

- **Requiring encryption on an endpoint**
  - User cannot override
  - Useful for “sensitive” data
- **Globus uses OpenSSL cipher stack as currently configured on your DTN**
- **FIPS 140-2 compliance: ensure use of FIPS capable OpenSSL libraries on DTN**
  - <https://www.openssl.org/docs/fips/UserGuide-2.0.pdf>



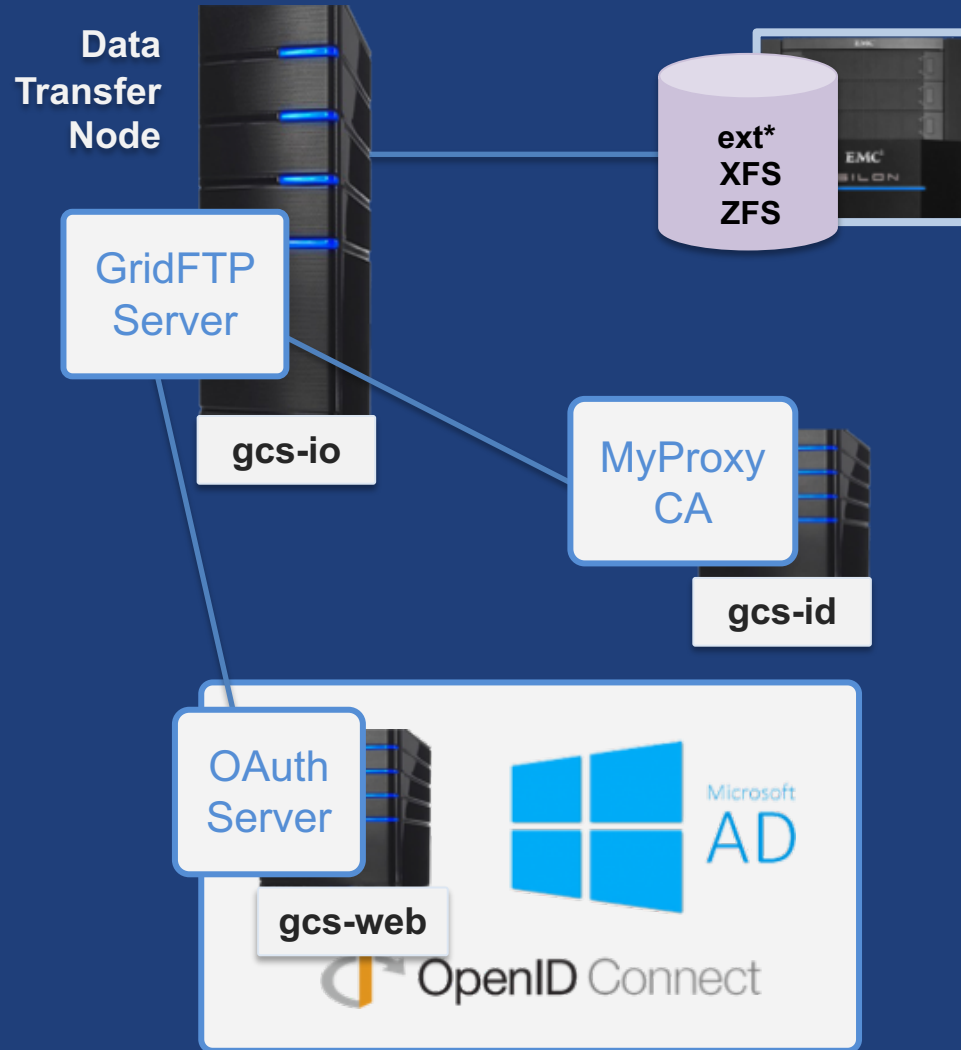


# Distributing Globus Connect Server components

- **Globus Connect Server components**
  - globus-connect-server-io, -id, -web
- **Default: -io, -id and -web on single server**
- **Common options**
  - Multiple -io servers for load balancing, failover, and performance
  - No -id server, e.g. third-party IdP
  - -id on separate server, e.g. non-DTN nodes
  - -web on either -id server or separate server for OAuth interface

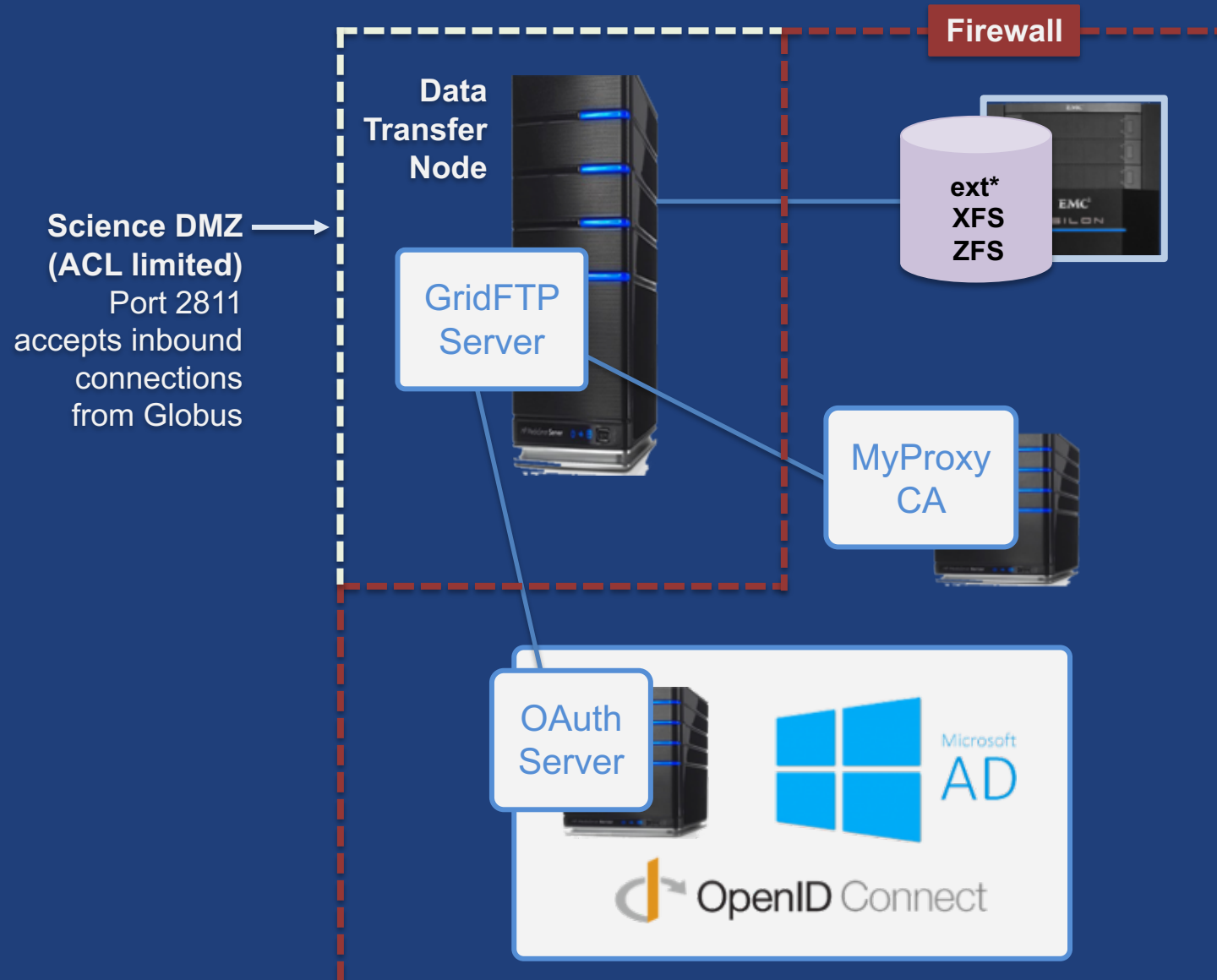


# Globus Connect Server Deployment





# Globus Connect Server Deployment





# Setting up multiple `-io` servers

- **Guidelines**
  - Use the same `.conf` file on all servers
  - First install on the server running the `-id` component, then all others
- **Install Globus Connect Server on all servers**
- **Edit `.conf` file on one of the servers and set [MyProxy] Server to the hostname of the server you want the `-id` component installed on**
- **Copy the configuration file to all servers**
  - `/etc/globus-connect-server.conf`
- **Run `globus-connect-server-setup` on the server running the `-id` component**
- **Run `globus-connect-server-setup` on all other servers**
- **Repeat steps 2-5 as necessary to update configurations**



# Example: Two-node DTN

-id  
-io



On “primary” DTN node (34.20.29.57):  
/etc/globus-connect-server.conf  
[Endpoint] Name = **globus\_dtn**  
[MyProxy] Server = **34.20.29.57**

-io



On other DTN nodes:  
/etc/globus-connect-server.conf  
[Endpoint] Name = **globus\_dtn**  
[MyProxy] Server = **34.20.29.57**



# Globus Network Manager

For environments with super duper  
special network constraints...

(a.k.a. "for the very brave")



# Globus Network Manager

- **Information from GridFTP to facilitate dynamic network changes**
- **Callbacks during GridFTP execution on local DTN**
- **Supplements information available via Globus transfer API**



# Globus Network Manager Callbacks

- **Pre-listen (binding of socket)**
- **Post-listen**
- **Pre-accept/Pre-connect (no Data yet)**
- **Post-accept/Post-connect (data in flight)**
- **Pre-close**
- **Post-close**





# Network manager use cases

- **Science DMZ Traffic Engineering**
  - Use SDN to dynamically route data path
  - Control path uses traditional route
- **Automated WAN bandwidth reservation**
  - OSCARS, AL2S
- **Note: All this requires custom code**



# Future directions

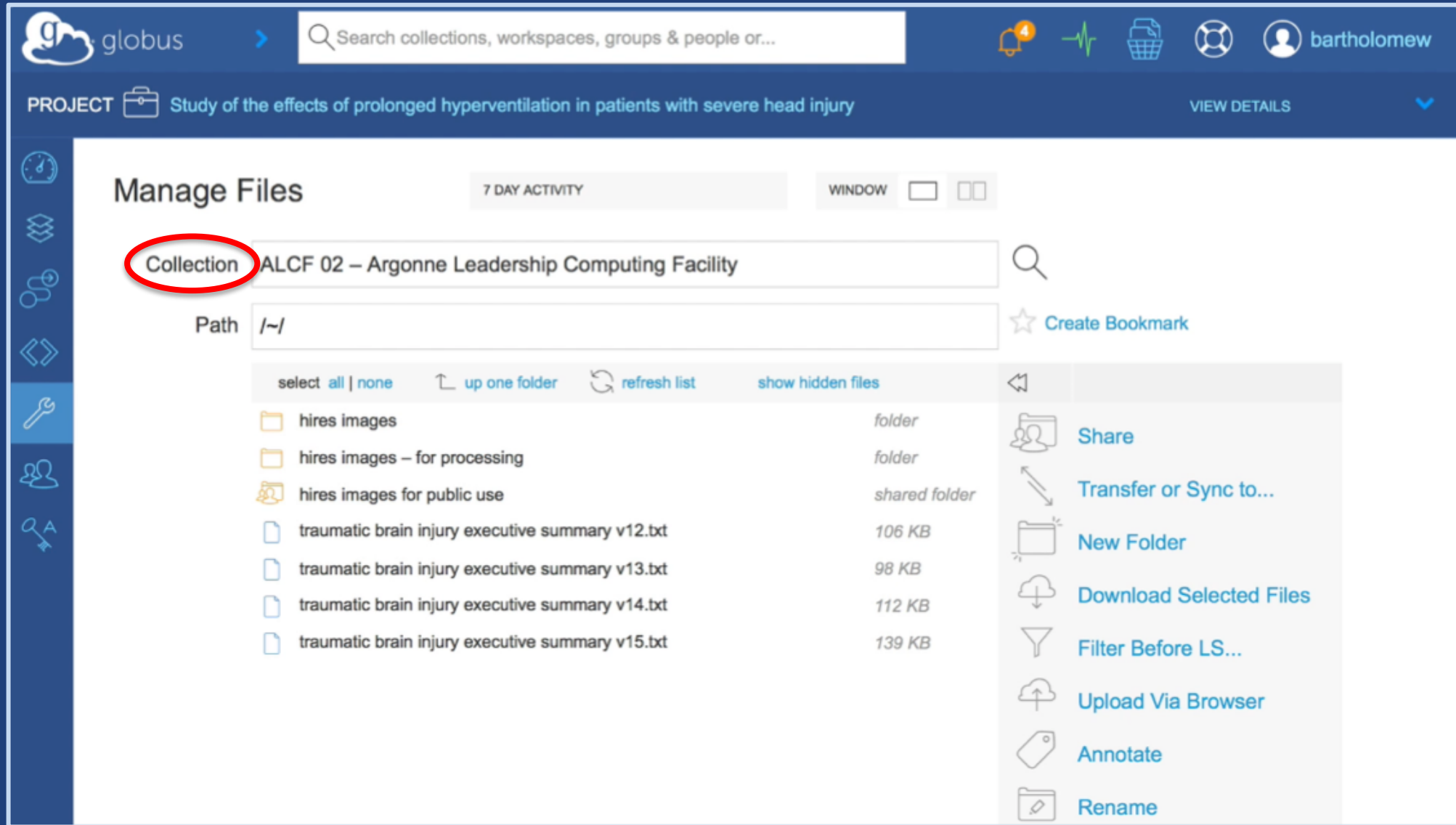
# Motivations for Globus Connect Server v5

- **Facilitate automation of installation and upgrades**
- **Allow scale out deployment**
  - Across DTNs
  - Across multiple file system connectors
- **Reduce number of ports required**
- **Streamline user experience with use of Globus sharing**
- **Enhance user registration of credentials for cloud storage connectors**
- **Prepare foundation for next set of enhanced capabilities**

# New features with Globus Connect Server v5


- **Collection model**
- **HTTPS access to storage**
- **Security improvements**
  - OAuth2 in GridFTP (no more X.509 user certificates or Myproxy!)
  - OpenID Connect identity provider
  - Credential expiration LoA policies
  - User credential management (e.g., for Google Drive, S3, Kerberos)
- **Kerberos protected file systems**
- **Directory listing with path expressions**

# Collections: The evolution of endpoints



The screenshot displays the Globus interface for a project titled "Study of the effects of prolonged hyperventilation in patients with severe head injury". The user is logged in as "bartholomew". The main area is titled "Manage Files" and shows a collection named "ALCF 02 - Argonne Leadership Computing Facility" (highlighted with a red circle). The path is "/~/". The file list includes folders like "hires images" and "hires images for public use", and files such as "traumatic brain injury executive summary v12.txt" through "v15.txt". A right-hand menu offers actions like "Share", "Transfer or Sync to...", "New Folder", "Download Selected Files", "Filter Before LS...", "Upload Via Browser", "Annotate", and "Rename".

globus > Search collections, workspaces, groups & people or...

PROJECT  Study of the effects of prolonged hyperventilation in patients with severe head injury VIEW DETAILS








### Manage Files

7 DAY ACTIVITY WINDOW

**Collection** ALCF 02 – Argonne Leadership Computing Facility

Path /~/ [Create Bookmark](#)

select all | none up one folder refresh list show hidden files

	hires images	folder	
	hires images – for processing	folder	
	hires images for public use	shared folder	
	traumatic brain injury executive summary v12.txt	106 KB	
	traumatic brain injury executive summary v13.txt	98 KB	
	traumatic brain injury executive summary v14.txt	112 KB	
	traumatic brain injury executive summary v15.txt	139 KB	

- Share
- Transfer or Sync to...
- New Folder
- Download Selected Files
- Filter Before LS...
- Upload Via Browser
- Annotate
- Rename



# Collection properties

- **Set of blobs (files), hierarchically named (folders)**
- **Rooted at a unique DNS name**
- **URL referenceable files, folders**
- **Accessible and manageable via:**
  - **HTTPS: client/server file access**
  - **GridFTP: async bulk transfer**
  - **REST API: advanced operations**
- **OAuth2 authentication and authorization via Globus Auth**
- **Collection-specific access policies**
- **Data is stored on a storage system, which determines storage policies such as durability and availability**
- **File change events**



## Installation & configuration enhancements for v5

- **Setup with any identity (GlobusID not required)**
- **Automatable installation and configuration**
- **Configuration API, CLI, GUI**
- **Scale-out deployment without shared file system**
- **Backup / restore configuration to / from the cloud**
- **Multiple storage systems simultaneously**
- **Single port GridFTP (no ephemeral ports)**
- **Distributed as Docker containers**



# Streamlined data sharing with v5

- **Remove friction of sharing**
  - Guest collections where possible, e.g., Google Drive
  - Hybrid collections: Mapped access to home & project folders, else guest access
- **Enhanced sharing permissions**
  - permission expiration
  - permissions on files (not just folders)
  - sharing via URL possession
- **Storage connectors: share from anywhere**





# New capabilities built on collections and v5

- **Data search (early release available now by request)**
  - With access control
  - Schema agnostic
  - Custom indexes domain specific
- **Event driven actions for automation**
  - Replication of data (across storage tiers)
  - Metadata extraction and ingest to search
  - Run analysis pipelines



# 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)

# Support resources

- **Customer engagement team**
- **Globus documentation: [docs.globus.org](https://docs.globus.org)**
- **Helpdesk and issue escalation: [support@globus.org](mailto:support@globus.org)**
- **Globus professional services team**
  - Assist with portal/gateway/app architecture and design
  - Develop custom applications that leverage the Globus platform
  - Advise on customized deployment and intergation scenarios



# Open Discussion