



Introduction to Globus: SaaS for Research Data Management

Vas Vasiliadis
vas@uchicago.edu

Jason Williams
jaswilli@globus.org

NYU Center for Data Science — May 31, 2017





Slides and useful links:
globusworld.org/tutorials



Globus delivers...

Big data transfer, sharing,
publication, and discovery...

...directly from your own
storage systems...

...via software-as-a-service



Bridge to campus HPC

Move datasets to campus research computing center



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



Bridge to national cyberinfrastructure

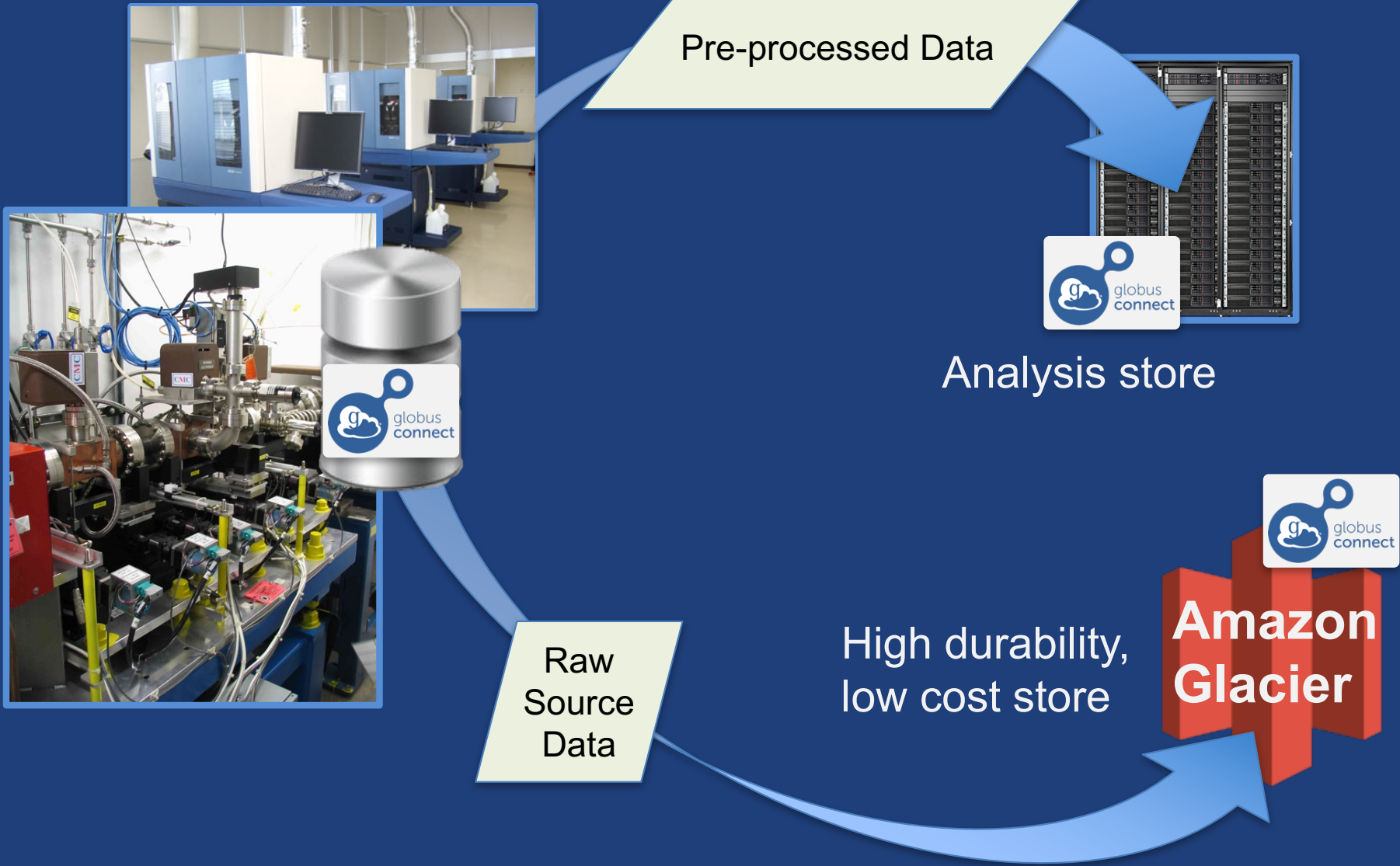
Move datasets to supercomputer,
national facility



Move results to campus (...)

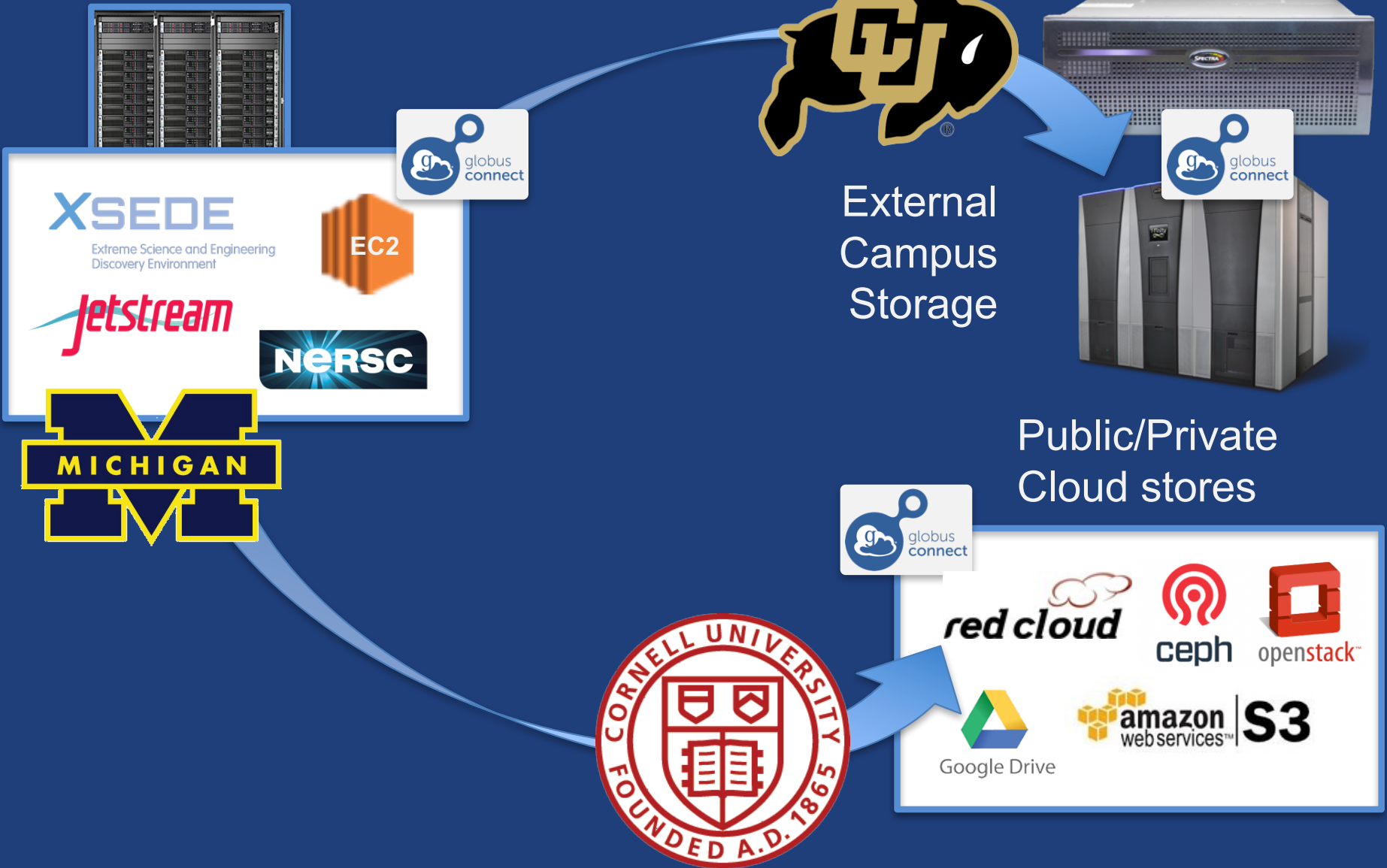


Bridge to instruments





Bridge to collaborators





Bridge to community/public



Project Repositories,
Replication Stores

globus connect

XSEDE
Extreme Science and Engineering
Discovery Environment

Jetstream

EC2

NERSC

Public Repositories

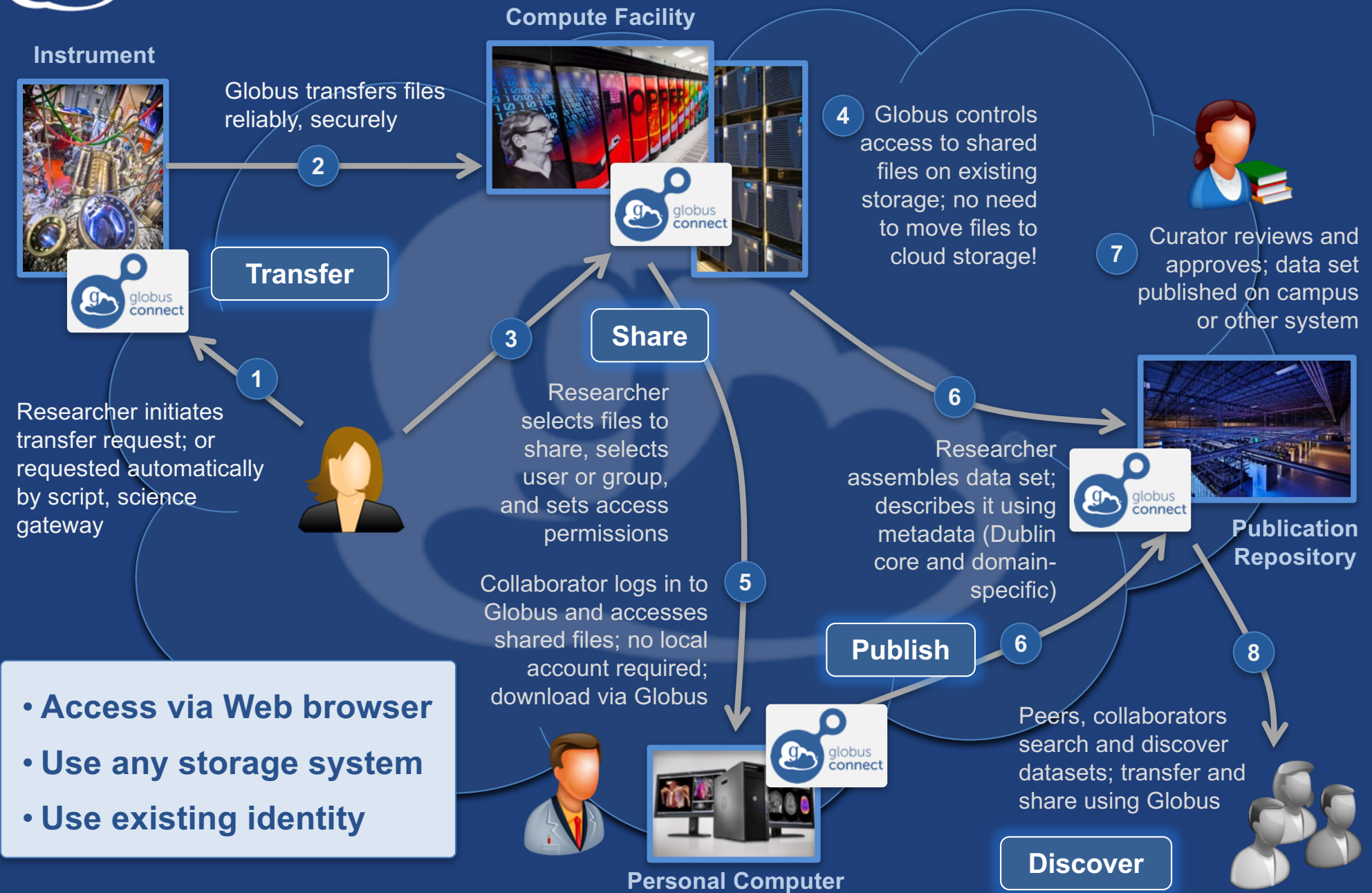
globus connect

globus connect

globus connect



Globus and the research data lifecycle





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**



File Transfer and Sharing Demonstration



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 text "globus". To the right are "Log In" and "Sign Up" links. A blue banner below the navigation bar contains the text: "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 graphic of the "MATERIALS DATA FACILITY" logo, which consists of a cluster of colorful circles (blue, green, yellow, orange, red) surrounding the text. Below the logo, there is a paragraph of text: "The Materials Data Facility (MDF) is a scalable repository where materials scientists can publish, preserve, and share research data. The repository provides a focal point for the materials community, enabling publication and discovery of materials data of all sizes." This is followed by another paragraph: "MDF is a pilot project funded by NIST, and serves as the first pilot community of the [National Data Service](#)." Below this is a link: "Contact Ben Blaiszik (blaiszik@uchicago.edu) to begin publishing your data". At the bottom of the screenshot is a "Browse" section with four buttons labeled "Issue Date", "Author", "Title", and "Subject".

<https://publish.globus.org>

Publish

globus Publish Manage Data Groups Support

Browse & Discover Data Publication Dashboard Communities & Collections

Submit: Describe this Dataset

Please fill in the requested information about this submission below. In most browsers, you can use the tab key to move the cursor to the next input box or button. To save you having to use the mouse each time, click on the save icon to save.

Title*

Authors*

Publication Year* 2010

Publisher*

© 2010-2015 Computation Institute, University of Chicago, Argonne National Laboratory legal

Discover

globus Publish Login Sign Up

Search Results

Collection results (2 results) advanced search Discover

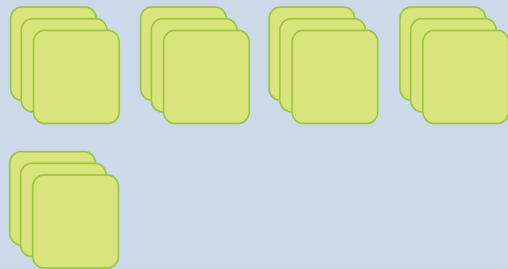
Date	Title	Author	Subject
21-Jul-2014	Sample Cathil Data	NCSL, NCSL Materials Science and Engineering, NCSL, Jozsef	Materials Science
1990	Coming glass ceramic #9006 Thermal Conductivity	Cornell University, Cornell University, Materials Science and Engineering, Cornell University, Cornell University, Cornell University	Thermal Conductivity
1990	Thermal Conductivity Amorphous SiO2	Cornell University, Cornell University, Materials Science and Engineering, Cornell University, Cornell University	Thermal Conductivity
1990	Coming Pyrex #7743 Thermal Conductivity	Cornell University, Cornell University, Materials Science and Engineering, Cornell University, Cornell University	Thermal Conductivity
1991	Sapphire Thermal Conductivity	University of Illinois at Urbana-Champaign, Materials Science and Engineering, Materials Science and Engineering, Materials Science and Engineering, Cornell University, Cornell University, Cornell University	Thermal Conductivity



Globus Authentication

Globus Data Publication

Medical Imaging Collection



Configuration and Policies



Materials Collection



Configuration and Policies



Globus Identity

Users and Groups

Handle

Persistent Identifiers

DOI

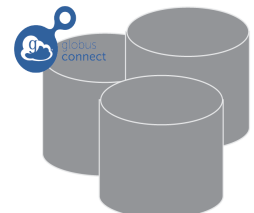
Metadata Catalog

UChicago

Argonne

NCSA

Amazon S3





Peer reviewed paper data

(Re)format...

- PDF/A
- HDF
- ...

Describe...

- Dublin core metadata
- Domain metadata
- Provenance info
- ...

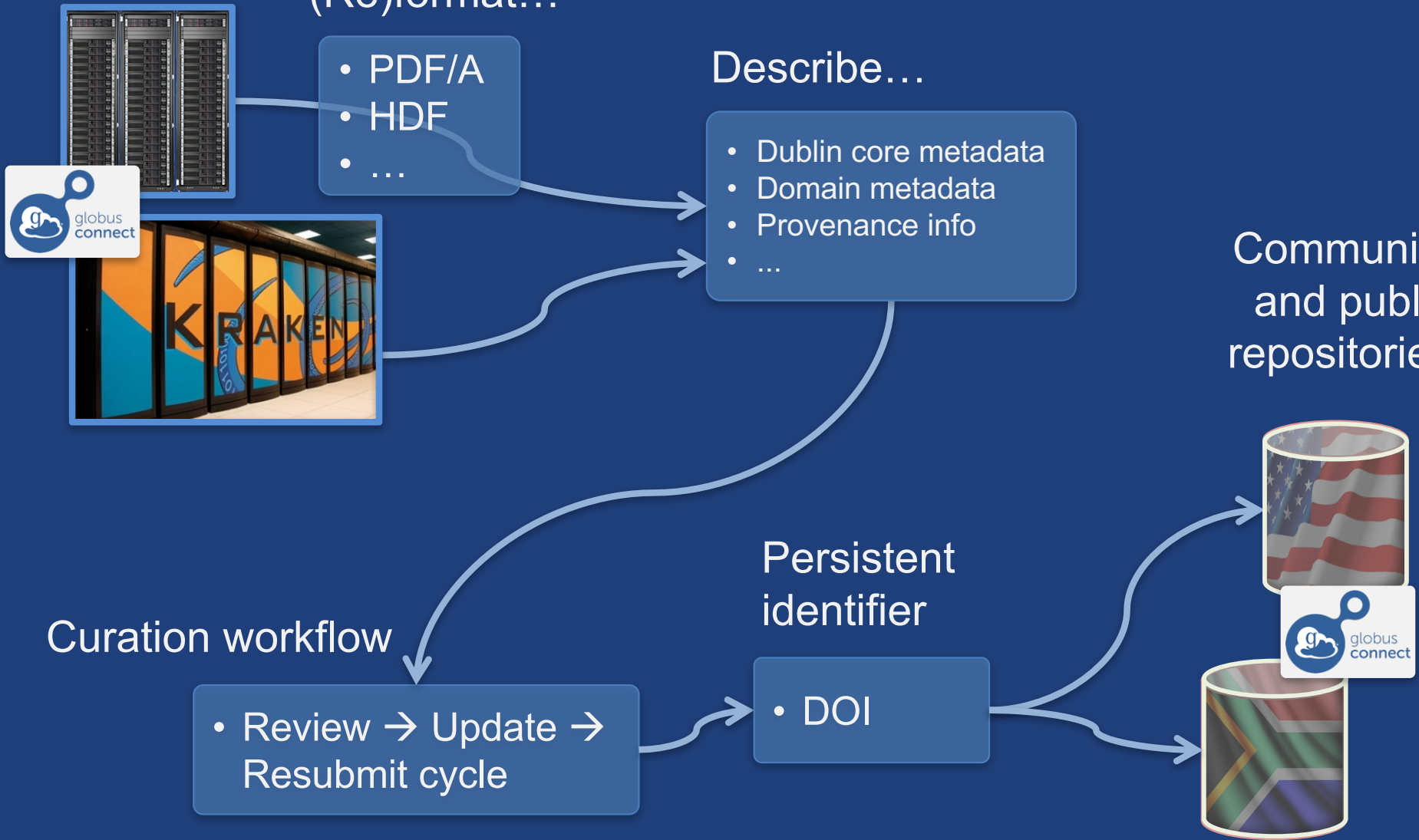
Community and public repositories

Curation workflow

- Review → Update → Resubmit cycle

Persistent identifier

- DOI





Data Publication Demonstration



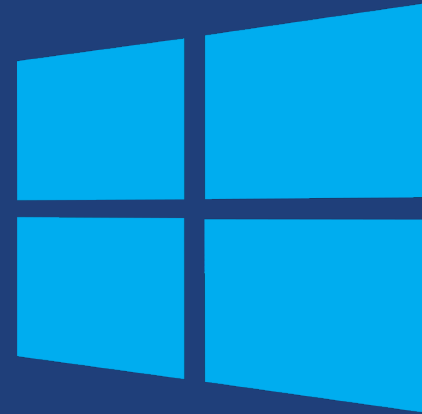
How can I use Globus
on my system?



Globus Connect...
Makes your storage
system a Globus
endpoint



Globus Connect Personal



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



Moving data between
your laptop and
another system



Exercise: Log in & transfer files

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 an ESnet test endpoint and your laptop**



Sharing Data



Share files

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



How can I integrate
Globus into my
research workflows?

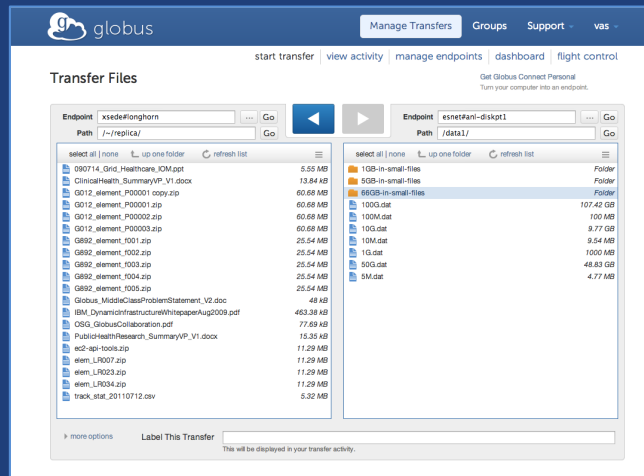


Globus serves as...

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



Use(r)-appropriate interfaces



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.
```

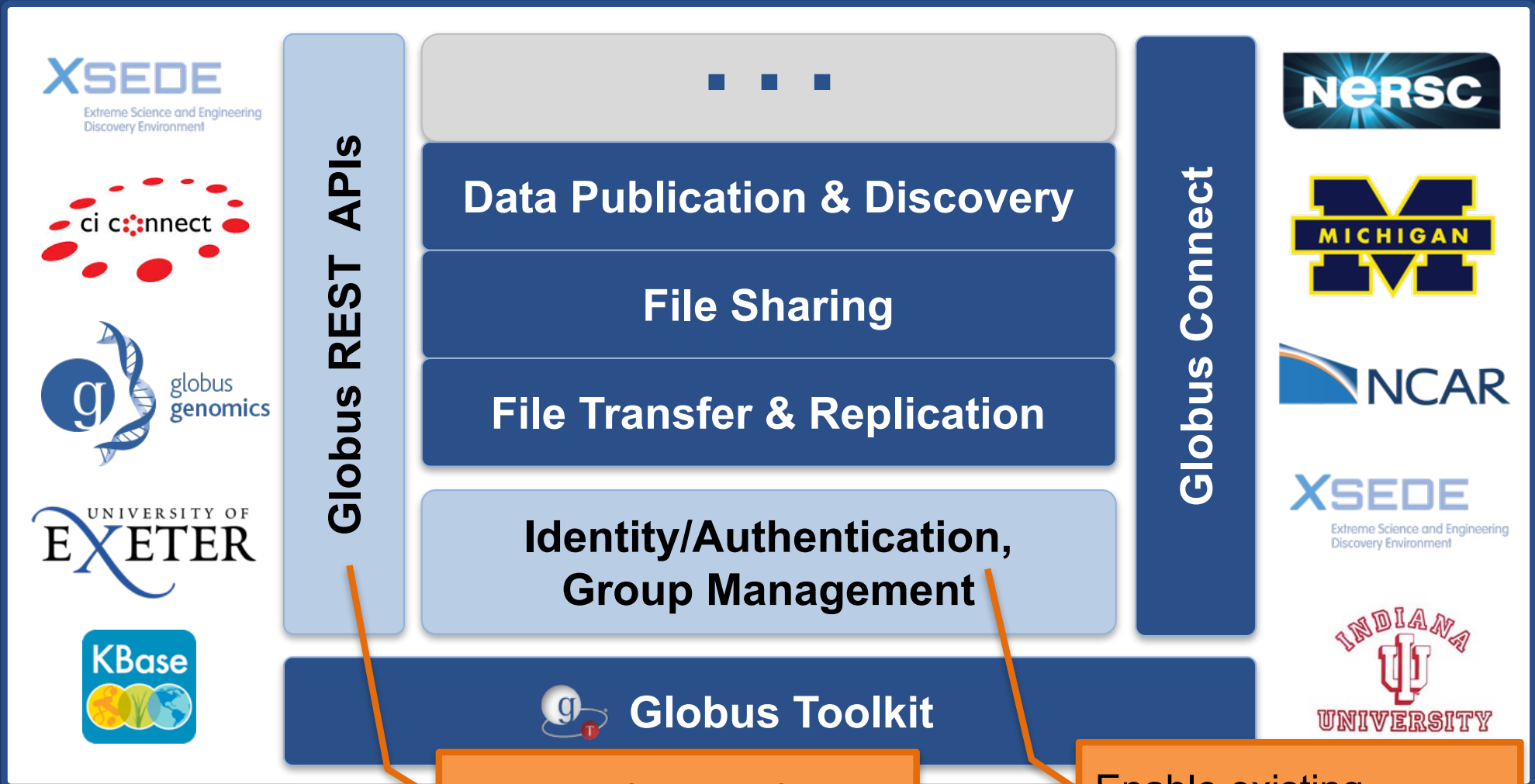
CLI

```
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



Integrate file transfer and sharing capabilities into scientific web apps, portals, gateways, etc.


Enable existing institutional ID systems to be used in external web applications



Data App: NCAR RDA


UCAR NCAR Closures/Emergencies Locations/Directions Find Pe

Hello [trocke@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	Globus Transfer Service (GridFTP)	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



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.

What is this [?](#)

→ 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...](#)



Globus PaaS: National Resource Access

XSEDE
Extreme Science and Engineering
Discovery Environment

globus Account ▾

Jetstream Web App would like to:


✓ Access all Jetstream resources

By clicking "Allow", you allow **Jetstream Web** information and services. You can rescind this

Allow

Deny

globus Globus Account Log In



compute | **calcul**
canada | canada

Compute Canada has partnered with Globus to offer this high performance file transfer service.

Calcul Canada s'est associé à Globus pour vous offrir ce service de transfert de fichier à haute performance.

Log in to use Compute Canada Globus Web App

Use your existing organizational login
e.g. university, national lab, facility, project, Google or [Globus ID](#)
(Your Globus username and password used prior to February 13, 2016 is now Globus ID)

WestGrid ▾

Continue

Didn't find your organization? Then use Globus ID to [sign up](#).



Globus PaaS: Identity Management

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

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



Latest News

[KBase at International Plant and Animal Genome XXI](#)

Posted by salazar Jan 09, 2013

[KBase Team at Argonne for November Build](#)

Posted by nharris Nov 30, 2012

[November Build at Argonne](#)

Posted by salazar Nov 23, 2012

[view news](#)

Upcoming Events

2013-01-12

[International Plant and Animal Genome XXI \(PAG 2013\)](#)

2013-02-18

[BERAC Presentations](#)

2013-02-24

[DOE/NIFA Plant Feedstocks Genomics for Bioenergy](#)

2013-02-25

[Proposed: Genomic Science Contractors-Grantees Meeting](#)



Globus PaaS developer resources



globus.github.io/globus-sdk-python/

globus-sdk-python 0.2.5 documentation »

Table Of Contents

- Globus SDK for Python (Beta)
- Installation
- Basic Usage
- API Documentation
- License

Python SDK

Installation

The Globus SDK requires Python 2.6

```
pip install globus-sdk
```

This will install the Globus SDK and it Bleeding edge versions of the Globus

```
git checkout https://github.com/globus/globus-sdk-python
cd globus-sdk-python
python setup.py install
```

Basic Usage

Globus Transfer API

API reference for transfer and sharing functions

Sample Application

Requirements

- You need to be in the tutorial
- Installed Globus Python SD

Jupyter Notebook

```
In [15]: from __future__ import print_function # for python 2
tutorial_endpoint_1 = "ddb59aef-6d04-11e5-ba46-22000b92c6ec" # endpoint "Glo
tutorial_endpoint_2 = "ddb59af0-6d04-11e5-ba46-22000b92c6ec" # endpoint "Glo
tutorial_users_group = "50b6a29c-63ac-11e4-8062-22000ab68755" # group "Tutori
```

Configuration

First you will need to configure the client with an OAuth2 access token. For the purpose of this tutorial, you can use 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
```

docs.globus.org/api

github.com/globus



HTTPS support (coming soon)

- **Synchronous alternative to GridFTP**
- **Same fine-grained access control model**
- **Greatly simplified sharing/transfer of “small” datasets**
- **Standard browser behaviors**
- **Integration with clients and web apps to further leverage existing research storage systems**



Globus sustainability model

- **Standard Subscription**

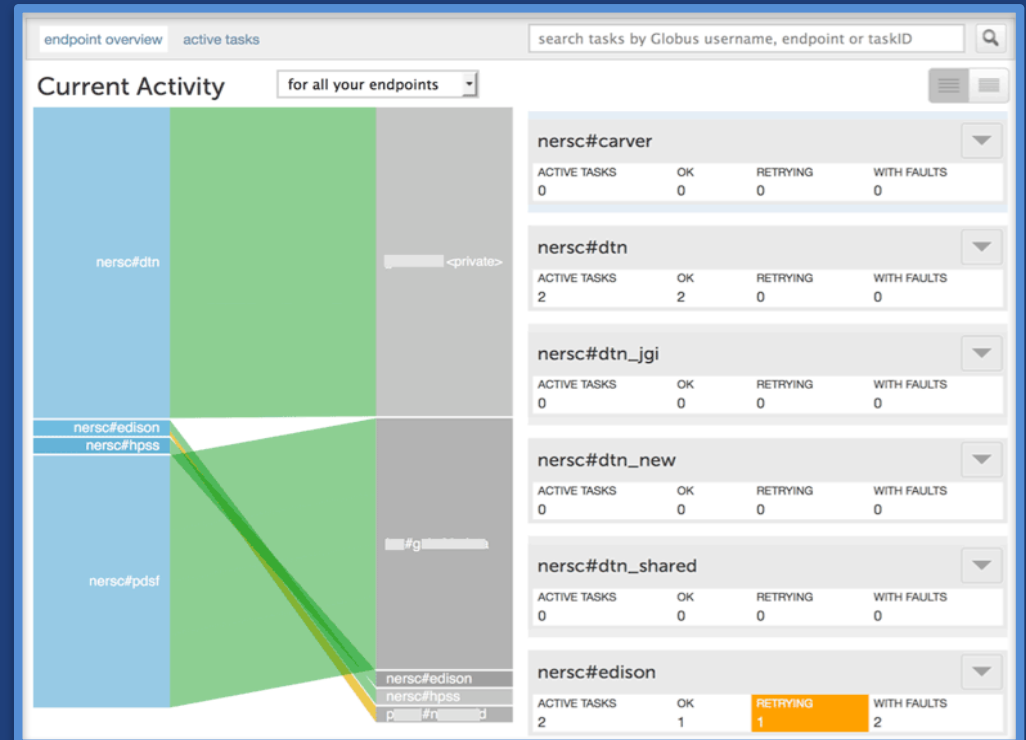
- Shared endpoints
- Data publication
- HTTPS support*
- Management console
- Usage reporting
- Priority support
- Application integration

- **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 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



Thank you to our users...

5

major services

270 PB

transferred

45 Bn

files processed

58,000

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!



NEW YORK UNIVERSITY



Berkeley
UNIVERSITY OF CALIFORNIA



JOHNS HOPKINS
UNIVERSITY

CORNELL
UNIVERSITY



NCAR

UF | UNIVERSITY of
FLORIDA



MICHIGAN STATE
UNIVERSITY



NCSA

Yale



THE UNIVERSITY OF
CHICAGO



syngenta

NIST



VirginiaTech
Invent the Future





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)**



Globus for System Administrators

Vas Vasiliadis
vas@uchicago.edu

Jason Williams
jaswilli@globus.org

NYU Center for Data Science — May 31, 2017





Slides and useful links:
globusworld.org/tutorials

All passwords: **g1obus2017**

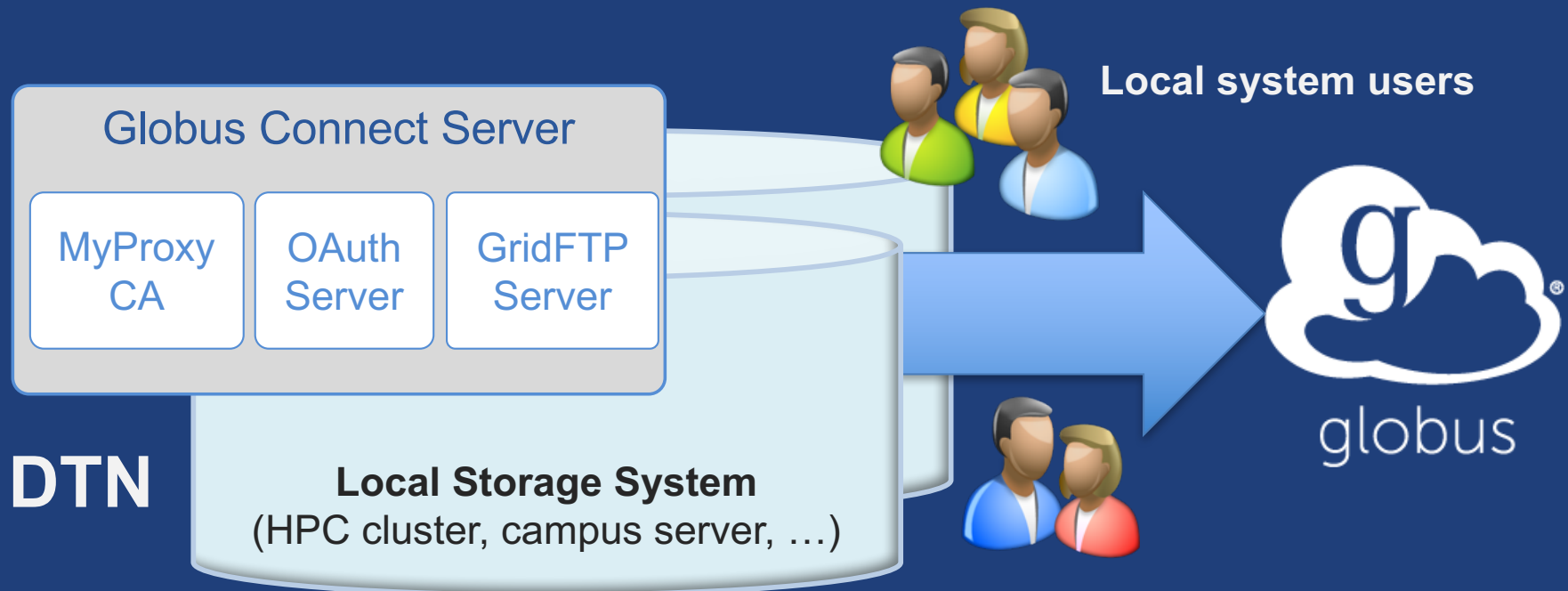


Enabling your storage system:

Globus Connect Server



Globus Connect Server



- **Create endpoint on practically any filesystem**
- **Enable access for all users with local accounts**
- **Native packages: RPMs and DEBs**



Storage connectors

- **Standard storage connectors (POSIX)**
 - Linux, Windows, MacOS
 - Lustre, GPFS, OrangeFS, etc.
- **Premium storage connectors**
 - Amazon S3
 - Google Drive
 - Spectra Logic BlackPearl
 - HPSS
 - Ceph RadosGW (S3 API)
 - iRODS
 - HDFS
- **Planned: Box, Azure, et al**



Demonstration

- **Creating a Globus endpoint on your storage system**
- **In this example, storage system = Amazon EC2 server**



Step 0: Create a Globus ID

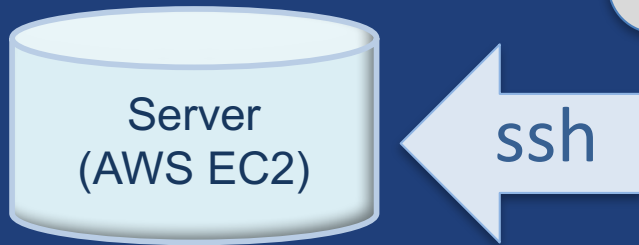
- **Installation and configuration of Globus Connect Server requires a Globus ID**
- **Go to `globusid.org`**
- **Click “create a Globus ID”**



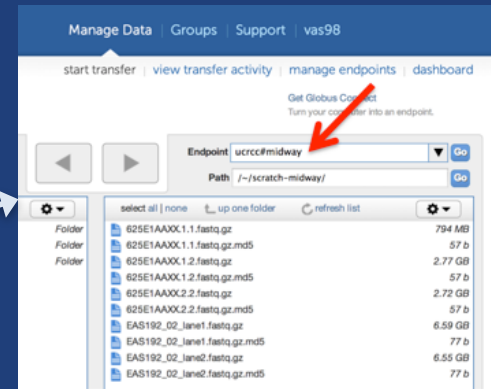
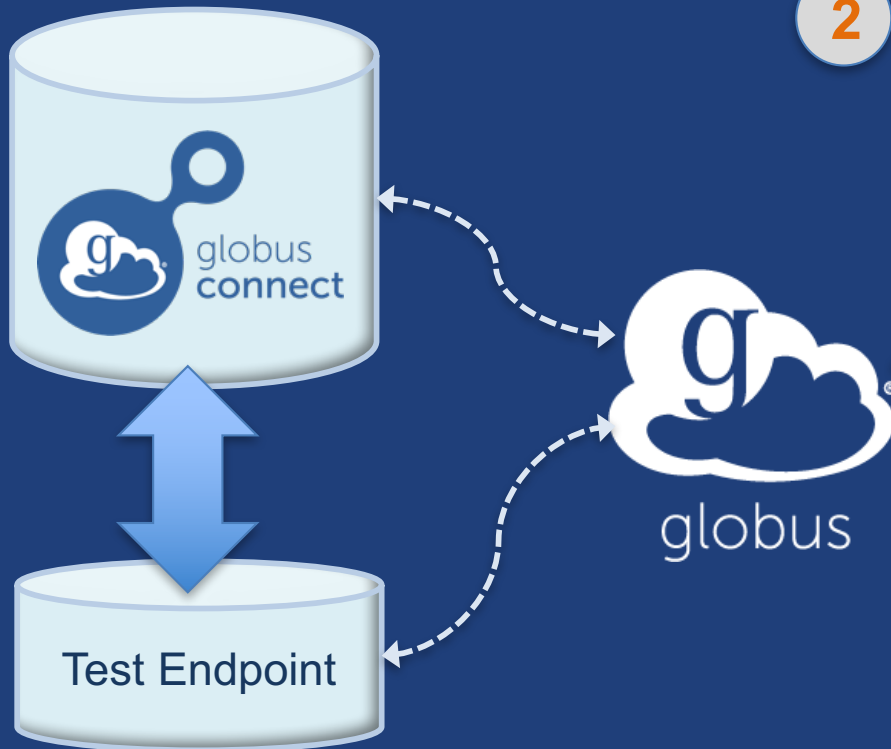
What we are going to do:

1 Install Globus Connect Server

- Access server as user “campusadmin”
- Update repo
- Install package
- Setup Globus Connect Server



2 Log into Globus



3 Access the newly created endpoint (as user 'researcher')

4 Transfer a file



Access your host

- **Create a Globus ID**
 - Optional: associate it with your Globus account
- **Get the DNS for your EC2 server**
- **Log in as user 'campusadmin':**

```
ssh campusadmin@<EC2_instance_IP_address>
```

(password: **globus2017**)
- **NB: Please sudo su before continuing**
 - User 'campusadmin' has sudo privileges



Step 3: Install Globus Connect Server

Cheatsheet: globusworld.org/tutorial

```
$ sudo su
$ curl -LOs 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 DNS name in the Endpoint field
 - Log in as user “researcher” (pwd: **globus2017**); you should see the user’s home directory
- **Transfer files to/from a test endpoint (e.g. Globus Tutorial, ESnet) and your endpoint**



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)
- **Futures: single-port GridFTP**



Configuring Globus Connect Server

- **Globus Connect Server configuration:**
`/etc/globus-connect-server.conf`
- **To enable changes you must run:**
`globus-connect-server-setup`
- **“Rinse and repeat”**
- **Globus service configuration**
 - Temporary: two configs for some options



Configuration file walkthrough

- **Structure based on .ini format**
 - [Section]
 - Option
- **Commonly configured options:**
 - Name
 - Public
 - RestrictedPaths
 - Sharing
 - SharingRestrictedPaths
 - IdentityMethod (CILogon, Oauth)



Exercise: Make your endpoint visible

- **Set `Public = true`**
- **Edit endpoint attributes**
 - Change the name to something useful, e.g. `<your_name> EC2 Endpoint`
- **Find your neighbor's endpoint**
 - You can access it too 😊



Enabling sharing on an endpoint

- **Set Sharing = True**
- **Run `globus-connect-server-setup`**
- **Go to the Transfer Files page**
- **Select the endpoint**
- **Create shared endpoints and grant access to other Globus users***

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



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?**



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`



Advanced Configuration

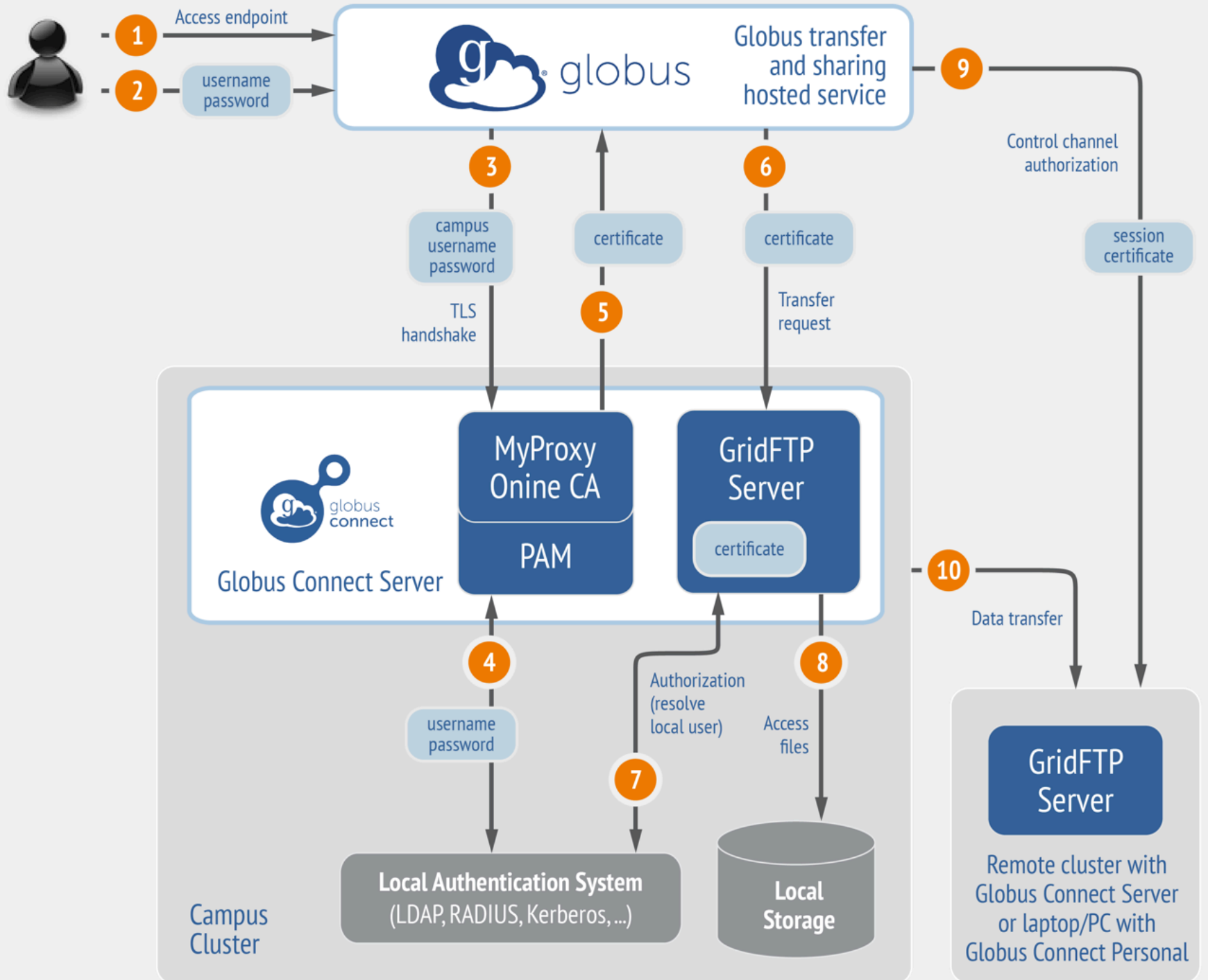


Using MyProxy OAuth server

- **MyProxy without OAuth**
 - Passwords flow via Globus to MyProxy server
 - Globus does not store passwords
 - Still a security concern for many campuses
- **Web-based endpoint activation**
 - Sites run MyProxy OAuth server or use CI Logon
 - Globus gets short-term X.509 credential via MyProxy OAuth protocol

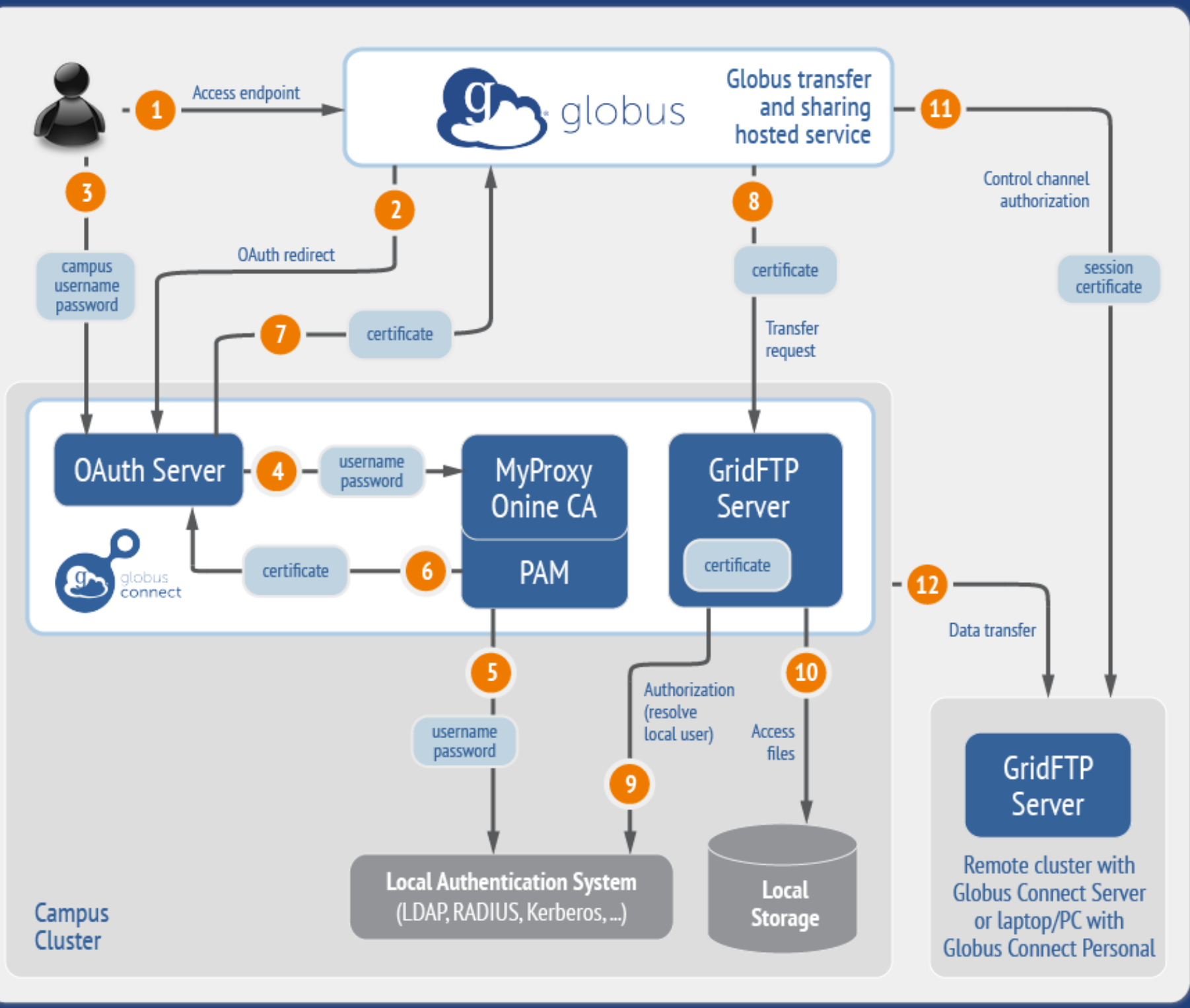


Endpoint activation using MyProxy





Endpoint activation using MyProxy OAuth





Single Sign-On with InCommon/CILogon

- **Your Shibboleth server must release the ePPN attribute to CILogon**
- **Local resource account names must match institutional ID (InCommon ID)**
- **AuthorizationMethod = CILogon**
- **CILogonIdentityProvider =
<institution_listed_in_CILogon_IdP_
list>**



Integrating your IdP

- **InCommon members**

- Must release R&S attributes to CILogon
- Mapping uses ePPN; can use GridMap

AuthorizationMethod = CILogon

CILogonIdentityProvider =

<institution_name_in_CILogon_IdP_list>

- **Non-members**

- IdP must support OpenID Connect
- Requires Alternate IdP subscription

- **Using an existing MyProxy server**



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, etc.

- **Convert existing endpoint to managed:**

```
endpoint-modify --managed-endpoint <endpoint_name>
```

- **Must be run by subscription manager, using the Globus CLI**

- **Important: Re-run endpoint-modify after deleting/re-creating endpoint**

- **Note: The above command requires the hosted (legacy) CLI**

https://docs.globus.org/cli/#hosted_command_line_interface_legacy



Demonstration: Command Line Interface (CLI)



Managed endpoint activity accessible via 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**



Demonstration: Management console



Endpoint Roles

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



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**
 - Limit number of ciphers used by OpenSSL
 - <https://access.redhat.com/solutions/137833>



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 such as CILogon
 - -id on separate server, e.g. non-DTN nodes
 - -web on either -id server or separate server for OAuth interface



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

1. **Install Globus Connect Server on all servers**

2. **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**

3. **Copy the configuration file to all servers**

- `/etc/globus-connect-server.conf`

4. **Run `globus-connect-server-setup` on the server running the `-id` component**

5. **Run `globus-connect-server-setup` on all other servers**

6. **Repeat steps 2-5 as necessary to update configurations**



Example: Two-node DTN

-id

-io



```
/etc/globus-connect-server.conf
```

```
[Endpoint] Name = globus_dtn
```

```
[MyProxy] Server = ec2-34-20-29-57.compute-1.amazonaws.com
```

-io



```
/etc/globus-connect-server.conf
```

```
[Endpoint] Name = globus_dtn
```

```
[MyProxy] Server = ec2-34-20-29-57.compute-1.amazonaws.com
```



Optimizing transfer performance



Balance: performance - reliability

- **In-flight tuning based on transfer profile (#files, sizes)**
- **Request-specific overrides**
 - Concurrency
 - Parallelism
- **Endpoint-specific overrides; especially useful for multi-DTN deployments**
- **Service limits, e.g. concurrent requests**



Network Use Parameters

- **Concurrency and parallelism configuration to tune transfers**
- **Maximum and Preferred**
- **Use values set for source and destination to determine parameters for a given transfer**
- **$\min(\max(\text{preferred src}, \text{preferred dest}), \max \text{src}, \max \text{dest})$**

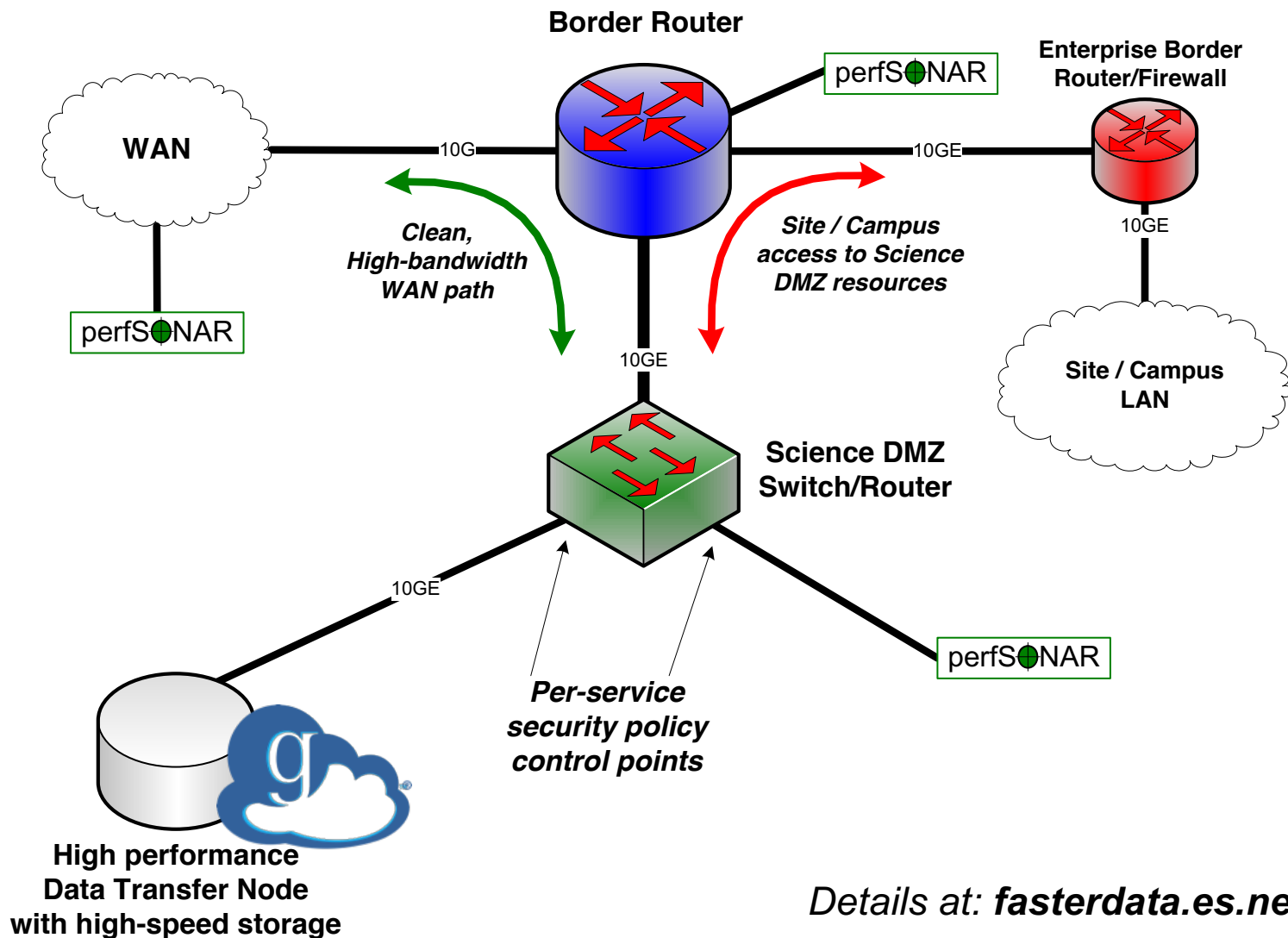


Network paths

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

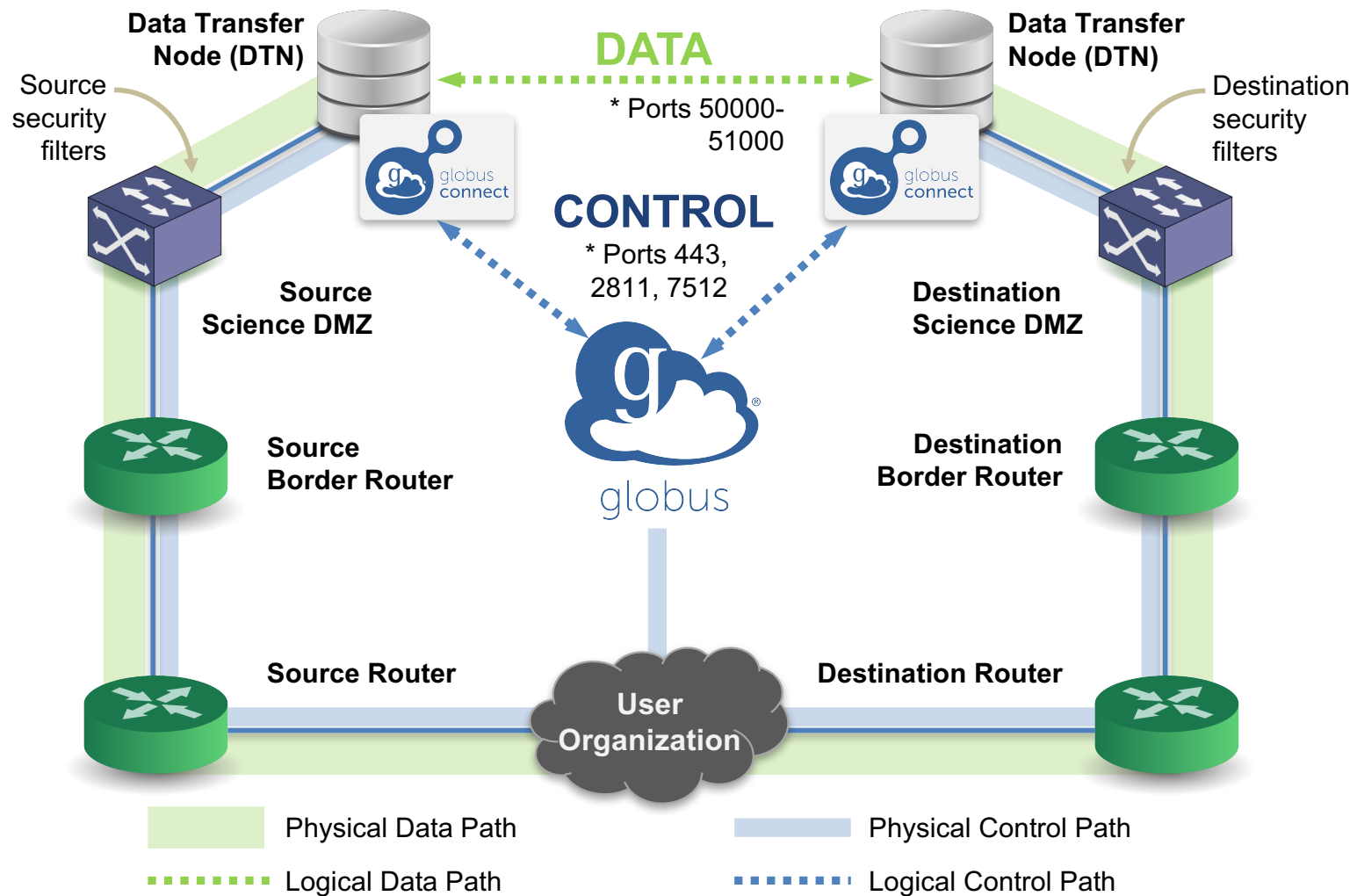


Best-practice deployment





Network Paths - Illustrative



* Please see TCP ports reference: https://docs.globus.org/resource-provider-guide/#open-tcp-ports_section

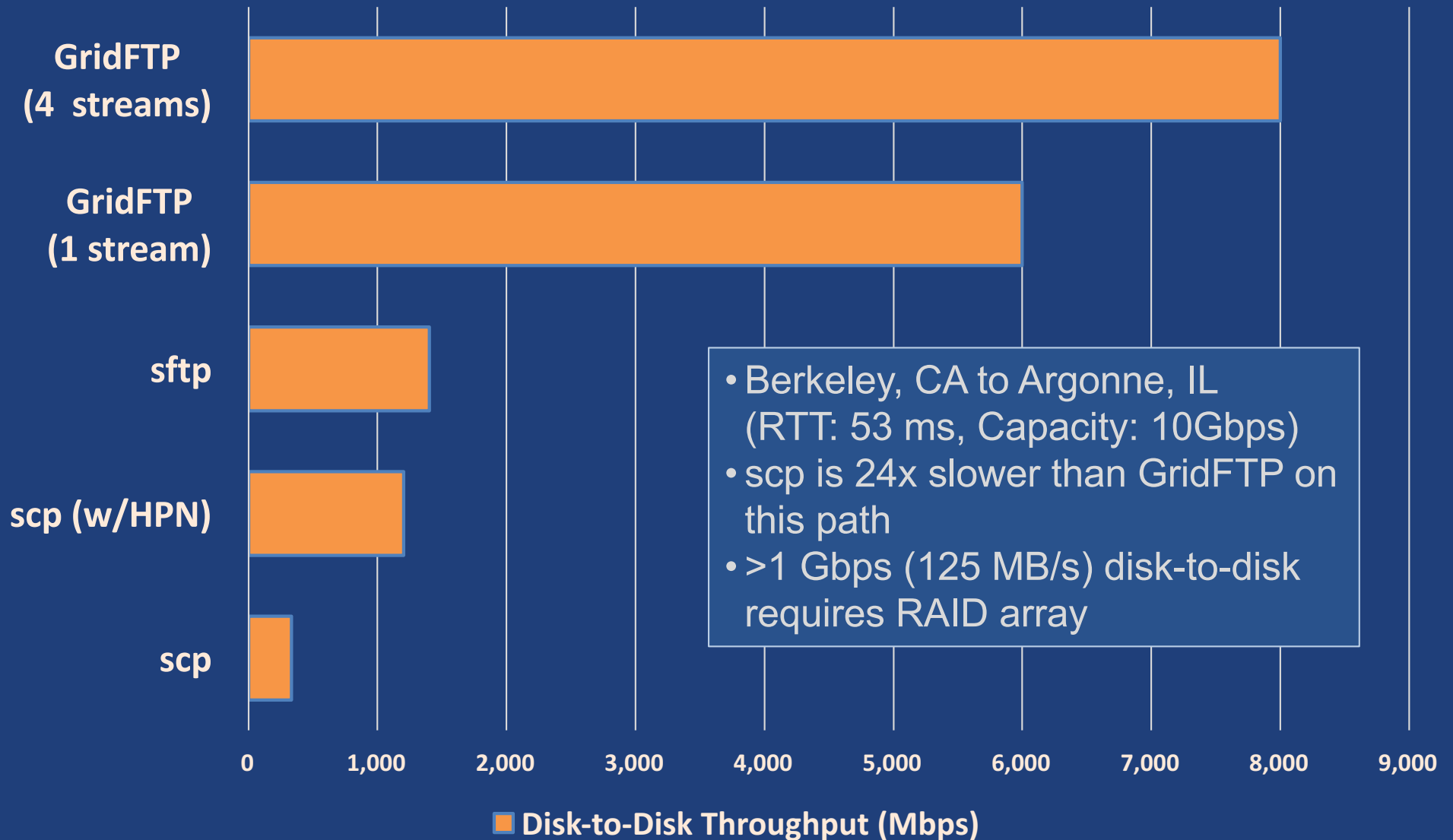


Illustrative performance

- **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





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**



Discussion



Enable your storage system

- Everything you wanted to know: **docs.globus.org**
- Need help? **support.globus.org**
- Mailing Lists: **globus.org/mailing-lists**
- Subscribe to help us make Globus self-sustaining:
globus.org/provider-plans
- Follow us: **[@globusonline](https://twitter.com/globusonline)**