

Globus for Data Management

Rachana Ananthakrishnan
(ranantha@uchicago.edu)



Data Management Challenges

- “Transfers often take longer than expected based on available network capacities”
- “Lack of an easy to use interface to some of the high-performance tools”
- “Tools [are] too difficult to install and use”
- “Time and interruption to other work required to supervise large data transfers”
- “Need data transfer tools that are easy to use, well-supported, and permitted by site and facility cybersecurity organizations”
- “No easy mechanism to share selected data with collaborators”



We envisage a world where data ...

... flows rapidly, reliably, and securely
among:

experimental facilities,
online and archival storage,
computing facilities, and
remote institutions



We envisage a world where data ...

... is **easily and securely shared** with collaborators and partners at other institutions



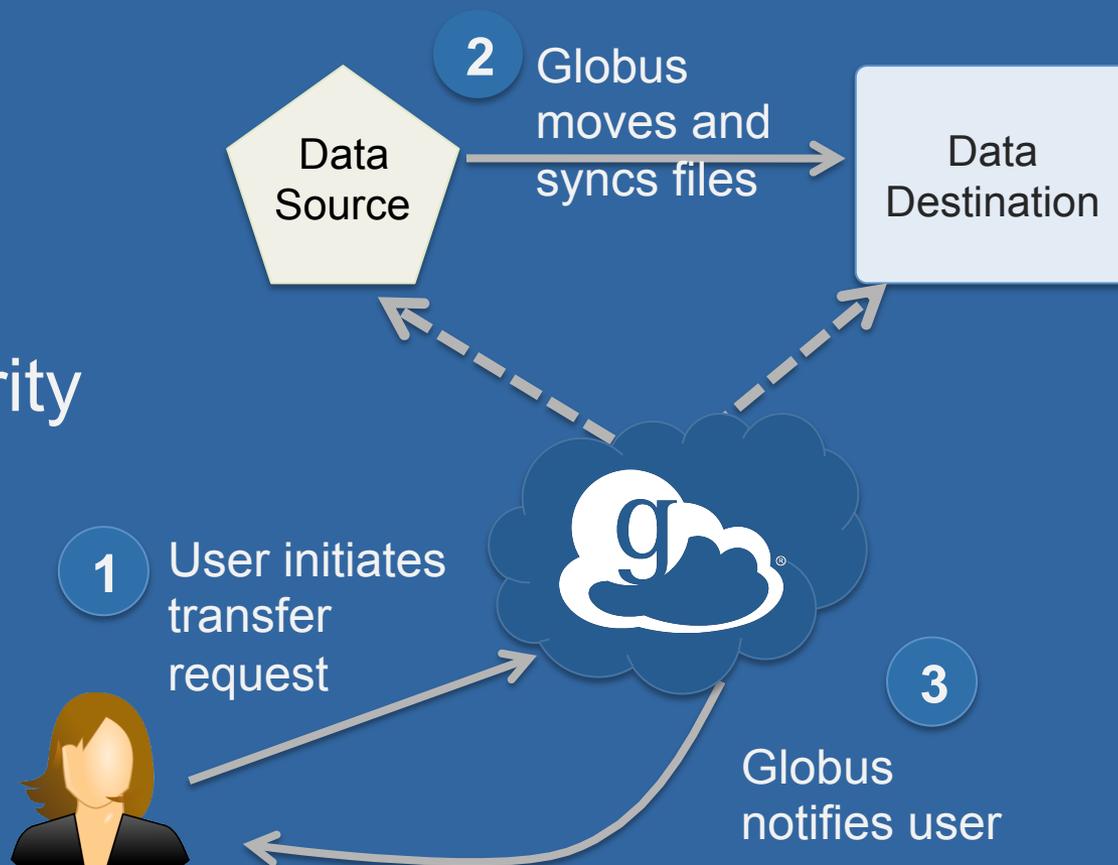
We envisage a world where data ...

... is readily **discoverable and accessible** to collaborators, regardless of their and the data's location



Reliable, secure, high-performance *file transfer* and *synchronization*

- “Fire-and-forget” transfers
- Automatic fault recovery
- Seamless security integration
- Powerful GUI and APIs



Transfer Files

[Manage Data](#)[Groups](#)[Support](#)[ranantha](#)[Transfer Files](#) | [Activity](#) | [Manage Endpoints](#) | [Dashboard](#) | [Flight Control](#)

Transfer Files

Get Globus Connect Personal
Turn your computer into an endpoint.

...

select all | none | up one folder | refresh list

- scratch-midway *Folder*
- share *Folder*

...

select all | none | up one folder | refresh list

- user2 *Folder*
- vytas *Folder*
- CASC Registration Form - April 2014-Rachana.docx *13.16 kB*
- NAMD_2.9_Linux-x86_64-multicore.tar.gz *2.75 MB*
- test2.txt *28 b*
- test3.txt *32 b*

▶ more options

Label This Transfer

This will be displayed in your transfer activity.



Transfer Options

▲ less options

Label This Transfer

This will be displayed in your transfer activity.

Transfer Settings

- only transfer new or changed files where the destination is newer
- delete files on destination that do not exist on source
- preserve source file modification times ?
- verify file integrity after transfer ?
- encrypt transfer ?

- checksum is different ?
- file does not exist on destination
- file size is different
- modification time is newer



Command line and scripting

Interactive login to command line interface:

```
$ ssh tuecke@cli.globusonline.org
```

Running commands remotely:

```
$ ssh tuecke@cli.globusonline.org <command>
```

```
$ ssh tuecke@cli.globusonline.org scp -r -s 3 -D \  
  nersc#dtn:~/myfile* mylaptop:~/projects/p1  
Task ID: 4a3c471e-edef-11df-aa30-1231350018b1  
$ _
```

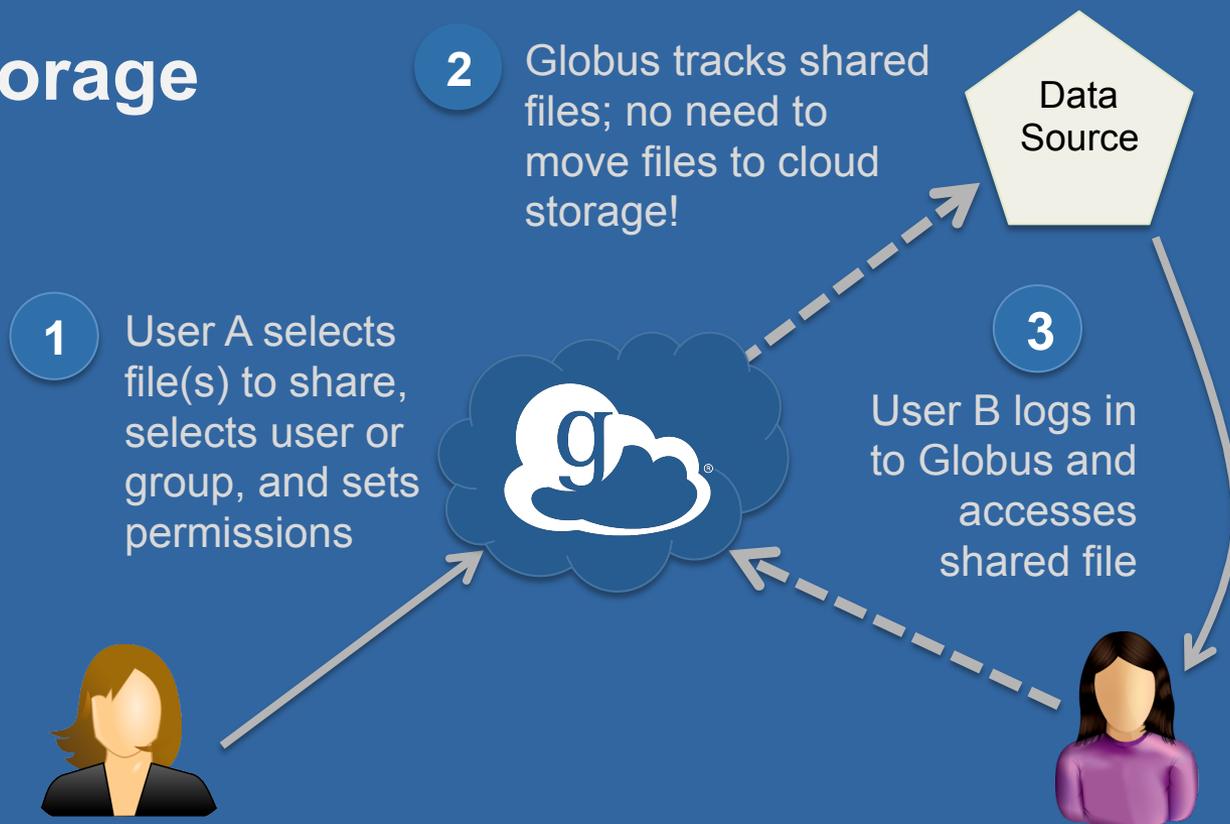
Using CLI with gsissh:

```
$ gsissh tuecke@cli.globusonline.org <command>
```



Simple, secure *sharing* off existing storage systems

- Easily share large data with any user or group
- No cloud storage required





Share files/folders



Manage Data

Groups

Support

ranantha

Transfer Files

Activity

Manage Endpoints

Dashboard

Flight Control

Transfer Files

Get Globus Connect Personal
Turn your computer into an endpoint.

Endpoint

Go

Path

Go



Endpoint

Go

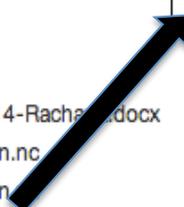
Path

Go

select all | none | up one folder | refresh list

CSSEF	
SharingDocScreenshot	
ec2	
gpntest	
mghdemo	
09-035_OWS-6_Security_ER.pdf	1.08 MB
CASC Registration Form - April 2014-Rache.docx	13.16 kB
fcn-ne120b.cam2.h0.0005-11-latlon.nc	0 b
fcn-ne120b.cam2.h0.0005-12-latlon.nc	0 b
mediastack_dbg_output.txt	0 b
test2.txt	28 b
test3.txt	32 b

- new folder
- show hidden files
- delete selected files
- share



Please select an endpoint above.



Manage permissions

Manage Shared Endpoint

« shared endpoints list

Manage Permissions For ranantha#testshare

Host: ucrcc#midway:/~/share/CSSEF/

name	read	write
▼ Path:/ view link for sharing		
Rachana Ananthakrishnan (ranantha)	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Ian Foster (ian)	<input checked="" type="checkbox"/>	<input type="checkbox"/> ✕
Globus Team	<input checked="" type="checkbox"/>	<input type="checkbox"/> ✕

✔ A group added successfully

ID (User or Group) [search »](#)

Path

NOTE: All paths are assumed to be folders

Permissions read write

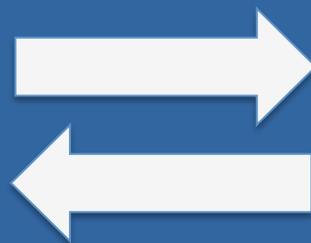


Endpoint Administrator Controls

- Restricted paths
 - Files and folders that can be used to transfer
- Enable or disable ability to share from an endpoint
- Sharing restricted paths
 - Files and folders that can be used to share
- Whitelist of users who are allowed to share from an endpoint
- Only allow read permissions on shared endpoints created by users



Amazon S3 Endpoints





Deploying Globus...



8,000

active endpoints

(in the past year)



Institution	Endpoints	Features
ALCF	alcf#*	
JGI	jgi#*	
LANL	Globus Connect Personal	Updating to DTNs
LBNL	lbnl#*	Sharing enabled
NERSC	nersc#*	Custom site
OLCF	olcf#*	
PNNL	pic#*	
SDSC	sdsc#*	Sharing enabled
XSEDE	xsede#*	In process of enabling Sharing



85

U.S. campuses



Globus Connect Personal

Create Globus Connect Personal Endpoint

Step 1 Get Your Globus Connect Personal Setup Key

Please enter a unique name for your Globus Connect Personal endpoint to help you identify it.

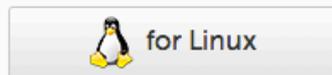
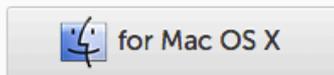
Endpoint Name: ranantha#

Setup Key

Copy the key above to your clipboard.

Step 2 Download & Install Globus Connect Personal

Click one of the buttons below to download and install Globus Connect Personal for your operating system.



Once downloaded, run the installer. When prompted, paste in the Setup Key to complete the installation.

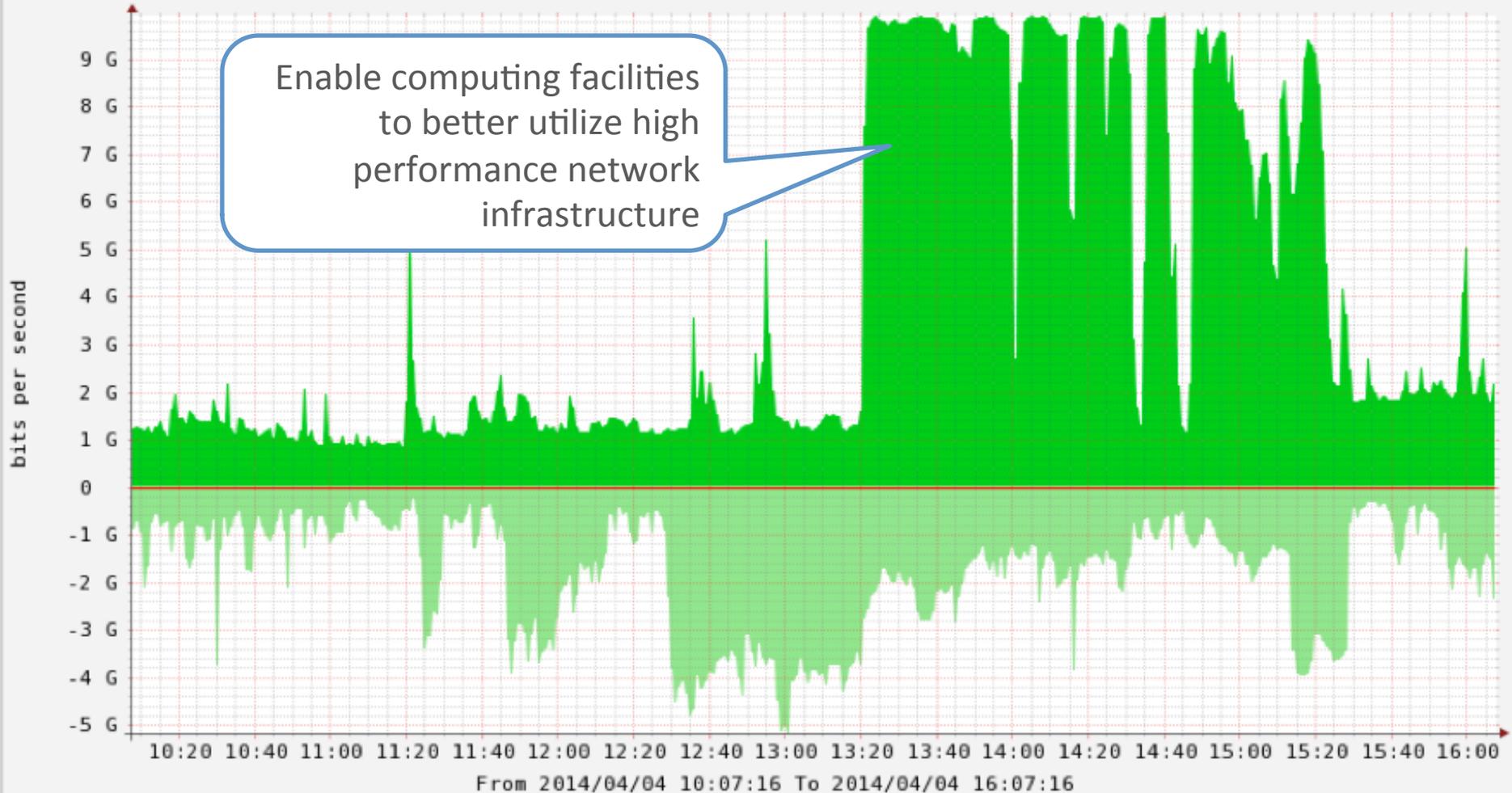




Globus increasingly used to better utilize network

hcc-schorr-mlxe - Traffic - ethernet7/1

Enable computing facilities to better utilize high performance network infrastructure



Inbound	Current:	2.16G	Average:	3.90G	Maximum:	9.90G
Outbound	Current:	2.36G	Average:	1.83G	Maximum:	5.18G

Source: University of Nebraska
Holland Computing Center



Using Globus...



Dan Kozak (Caltech)
replicates 1 PB LIGO
astronomy data for
resilience



Erin Miller (PNNL)
collects data at
Advanced Photon
Source, renders at
PNNL, and views at
ANL





Users upload data to their archives at University of Exeter: Globus API used for integration into their services

UNIVERSITY OF EXETER

THE SUNDAY TIMES UNIVERSITY OF THE YEAR 2012-13 UNIVERSITY OF EXETER

Start Transfer View Transfer Activity Groups Manage Endpoints Flight Control ranantha's account Log Out

Transfer Files View Transfer Activity Get Globus Connect Turn your computer into an endpoint.

Endpoint enter endpoint name ... Go Path Go

Please select an endpoint above.

Endpoint enter endpoint name ... Go Path Go

Please select an endpoint above.

more options Label This Transfer This will be displayed in your transfer activity.

Using our site | Freedom of Information | Data Protection | Copyright & disclaimer | Privacy & Cookies | powered by globus online



Earth System Grid Federation leverages Globus for data download: including the user interfaces

The screenshot displays the ESGF search results page. At the top, there are logos for ESGF and Argonne National Laboratory. A navigation bar includes links for Home, Search, Tools, Account, and Logout. The main content area shows search filters and results. On the left, there are sections for 'Current Selections' and 'Search Categories'. The search results section shows a single result: 'anl.cssef.homme.v1|esg.anl.gov' with 12 files. A red circle highlights the 'Globus Online' link in the 'Expand' options.

ESGF
Earth System Grid Federation

Argonne
NATIONAL
LABORATORY

Home Search Tools Account Logout

Current Selections
[remove all](#)
[\(x\) text:HOMME](#)
[\(x\) time frequency:day](#)

Search Categories
Project
Institute
Model
Instrument
Experiment Family
Time Frequency
Product
Realm
Variable

Search

Examples: *temperature*, *"surface temperature"*, *climate AND project:CMIP5 AND variable:hus*.
To download data: add datasets to your Data Cart, then click on *Expand* or *wget*.

Search All Sites Show All Replicas Show All Versions

< 1 > displaying 1 to 1 of 1 search results
Display 10 datasets per page
[Add All Displayed to Datacart](#) [Remove All Displayed from Datacart](#)

Results Data Cart

Show all Filter over search constraints Show initial 10 files [Remove All](#) [WGGET All Selected](#)

anl.cssef.homme.v1|esg.anl.gov (Total Number of Files: 12) [Expand](#) | [WGGET](#) | [Globus Online](#) | [Remove](#)



Using Globus web pages

globus online Manage Data | Support | rachanacssef

start transfer | view transfer activity | manage endpoints | dashboard

Select Destination

Get Globus Connect
Turn your computer into an endpoint.

Your Globus Connect endpoint has been selected by default. If you would prefer to transfer to a different endpoint, you can select it below.

Endpoint: rachanacssef@laptop Go

Path: /~/work/ Go

Folder	Creation Date	Type
coDas	04/20/2012 12:18 PM	Folder
cog-1.8	04/07/2010 10:31 AM	Folder
coyote	08/04/2011 08:56 AM	Folder
crux	05/28/2010 04:27 PM	Folder
csesf	10/31/2012 11:26 AM	Folder
documents	10/03/2011 07:56 AM	Folder
doeCollaboratories2012	04/09/2012 01:00 PM	Folder
elephantHost	09/08/2008 09:08 AM	Folder
esg	07/12/2012 01:43 PM	Folder
globusOnline	06/27/2011 10:27 PM	Folder
globusOrg	04/15/2010 01:39 PM	Folder
globusToolkit	05/12/2010 08:35 PM	Folder
glDocs	02/21/2007 01:44 PM	Folder
bl	09/17/2010 10:43 AM	Folder
lglobus2.0	03/04/2011 10:46 AM	Folder
lglobusReview	05/03/2011 01:11 PM	Folder
nhinDemo	06/10/2010 10:11 PM	Folder
notes	04/11/2012 08:47 AM	Folder
nsfSI2Panel	11/07/2011 03:04 PM	Folder
performanceEvaluations	04/17/2012 03:37 PM	Folder

Label This Transfer

This will be displayed in your transfer activity.

ESGF Earth System Grid Federation

Argonne NATIONAL LABORATORY

Home Search Tools Account Logout

Globus Online Data Transfer

GlobusOnline Transfer Status

Your Globus Online Transfer has been submitted
(The transfer has been accepted and a task has been created and queued for execution.)

Globus Online TaskID: 55b8c034-243b-11e2-bf56-1231380b8963

[You can monitor your transfer here.](#)

You will receive email notifications from Globus Online when your transfer is complete.

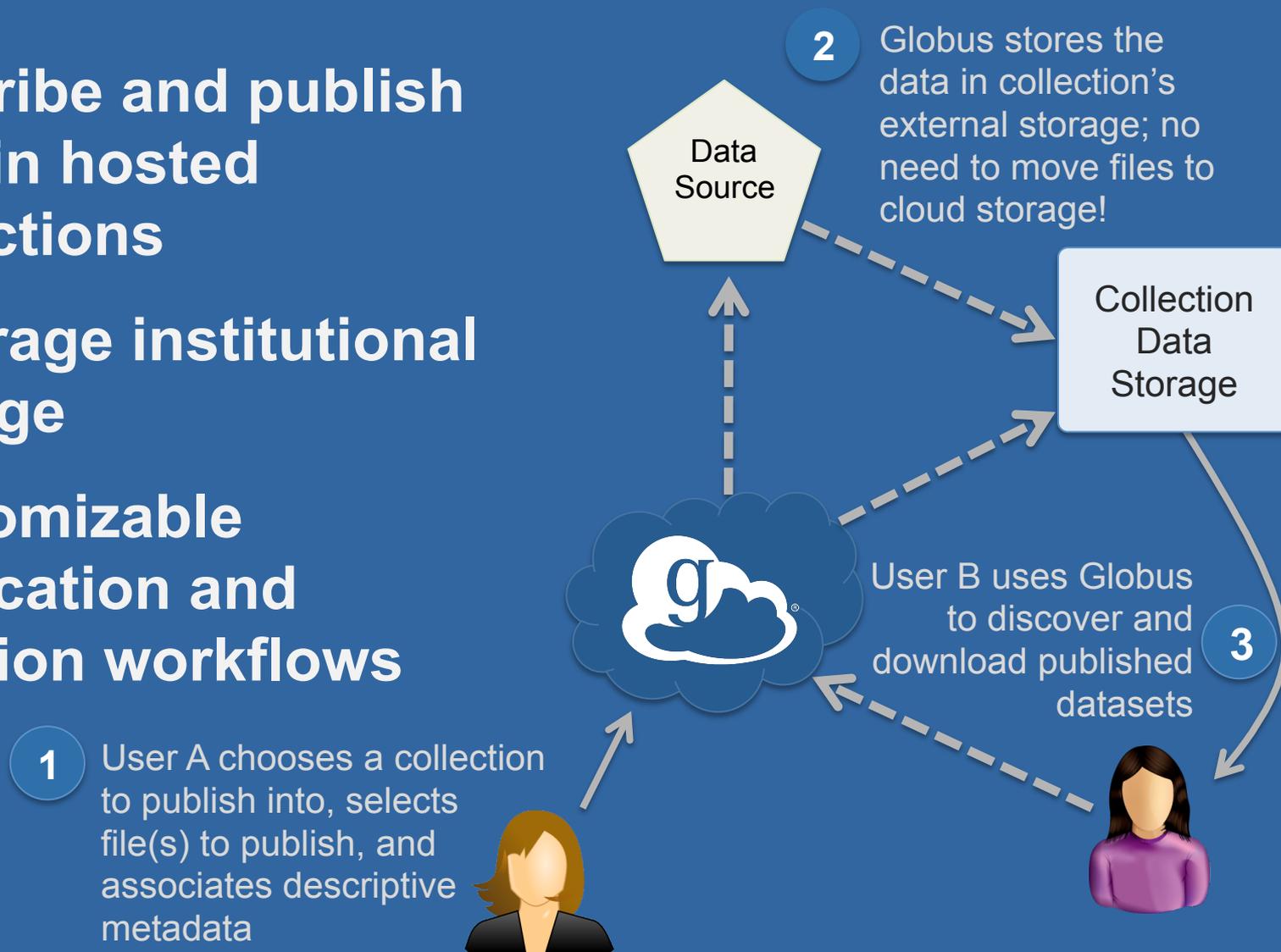


Globus is moving beyond
transfer and sharing to
**data publication and
discovery**



Curated *publishing* and rich *discovery* of research data.

- Describe and publish data in hosted collections
- Leverage institutional storage
- Customizable publication and curation workflows





Globus Data Publication

- SaaS for publishing large research data
- Bring your own storage
- Extensible metadata
- Publication and curation workflows
- Public and restricted collections
- Rich discovery model



We are a non-profit, delivering a production-grade service to the non-profit research community

Our challenge:
Sustainability



Globus Provider Subscriptions

- Managed Endpoints
 - Priority support
 - Management console
 - Usage reports
 - Mass Storage System optimization
 - Host shared endpoints
 - Integration support
- Branded Web Site
- Alternate Identity Provider (InCommon is standard)



globus.org/provider-plans

 Thank you to our sponsors!



U.S. DEPARTMENT OF
ENERGY



THE UNIVERSITY OF
CHICAGO

Argonne
NATIONAL LABORATORY



powered by
amazon
web services

