

Monsoon and Globus

Utilizing Globus tools for fast, easy, & powerful data sharing

Presentation media, including this slide-stack:

rcdata.nau.edu/hpcpub/globus

Presentation topics

- General introduction to Globus' capabilities and benefits
- Understanding its user-facing architecture
- Applicable use-cases (or alternative approaches)
- Usage walk-throughs (web client and command line)
- Personal end-points for workstations
- Group collaboration with both edu AND unaffiliated users

Globus: multi-use data sharing utility

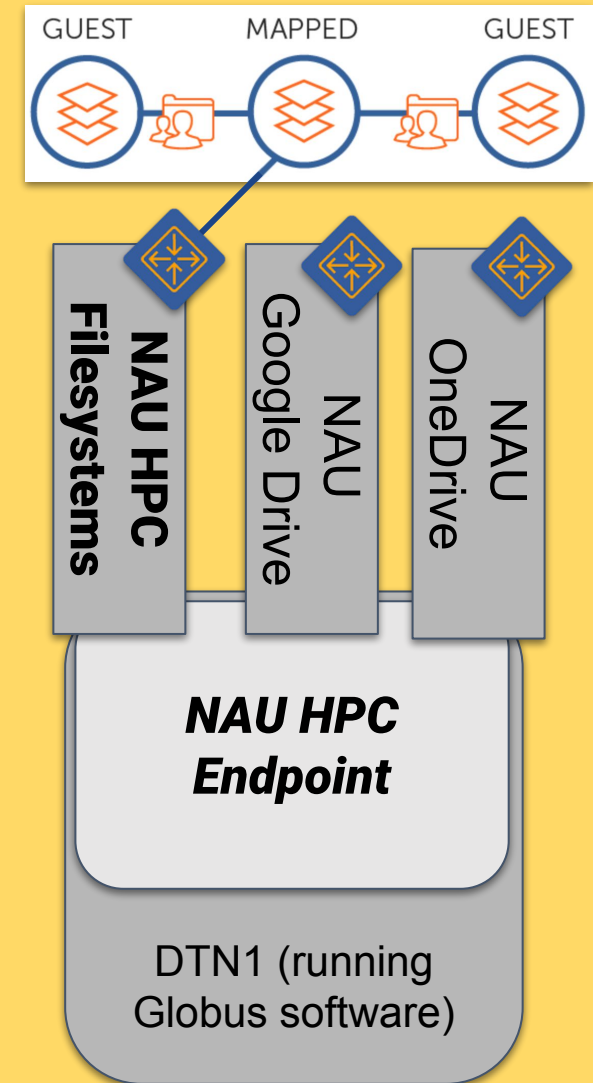
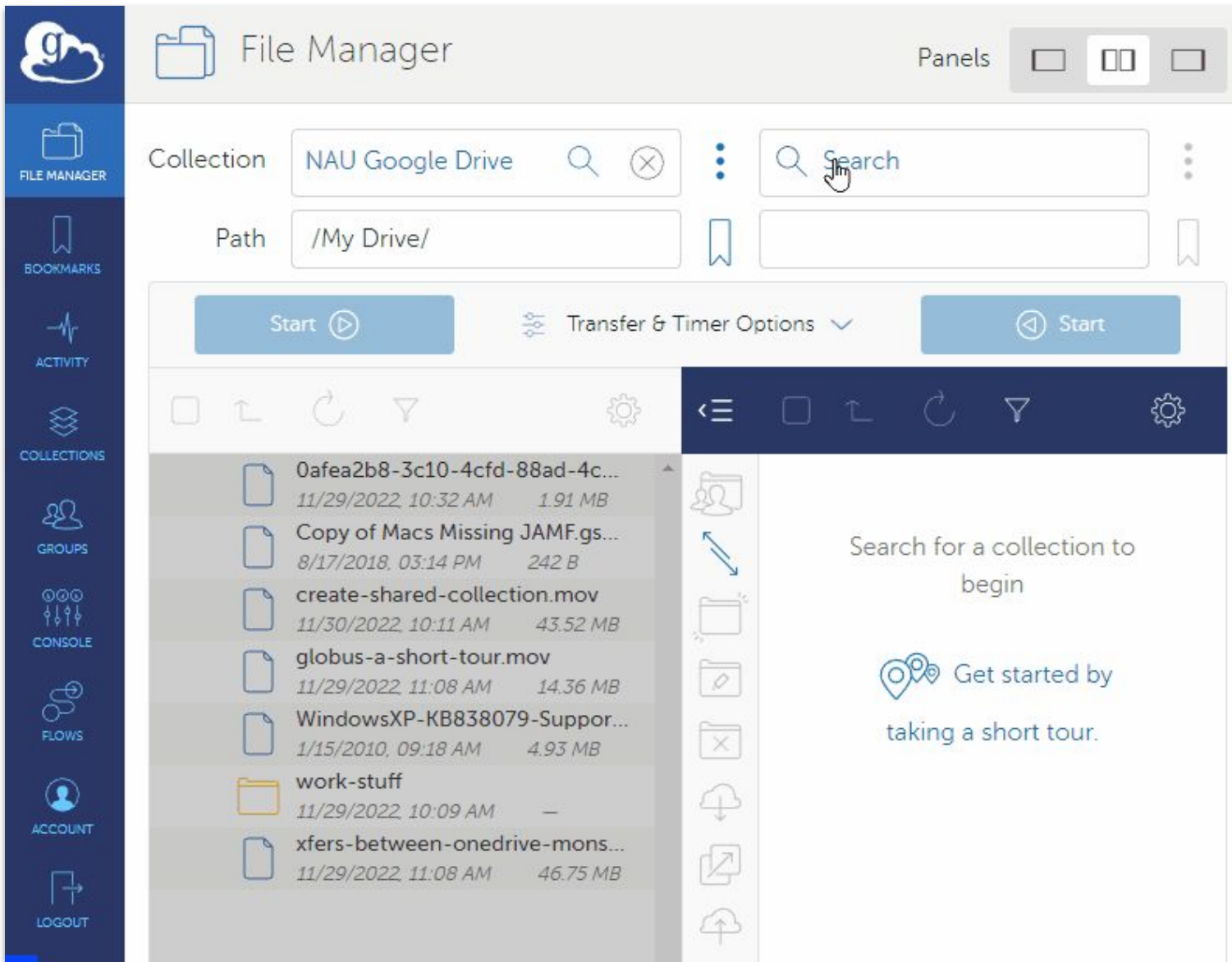
What: Globus is a great solution for transferring data to/from Monsoon, as well as to internal filesystems of other Universities and research facilities.

Using its simple web interface *like a virtual remote control*, users may submit instructions for data-transfers, or may schedule data-syncs... and Globus takes it from there! It can even update you via email about status or exceptions.

How: Your data is transferred directly between the source and destination systems. Meanwhile, the service tunes performance parameters, maintains security, monitors progress, and validates correctness. You can check the transfer status at any time via a dashboard.

When: If a network/hardware issue is encountered, Globus will automatically resume when all components are ready. Even if an issue requires user intervention, Globus will resume after you've remedied the problem.

Interface & Architecture



app.globus.org/file-manager?origin_id=4ce35914-620b-11e9-bf44...

File Manager

Collection: NAU HPC Filesystems

Path: /home/jtb49/ondemand/

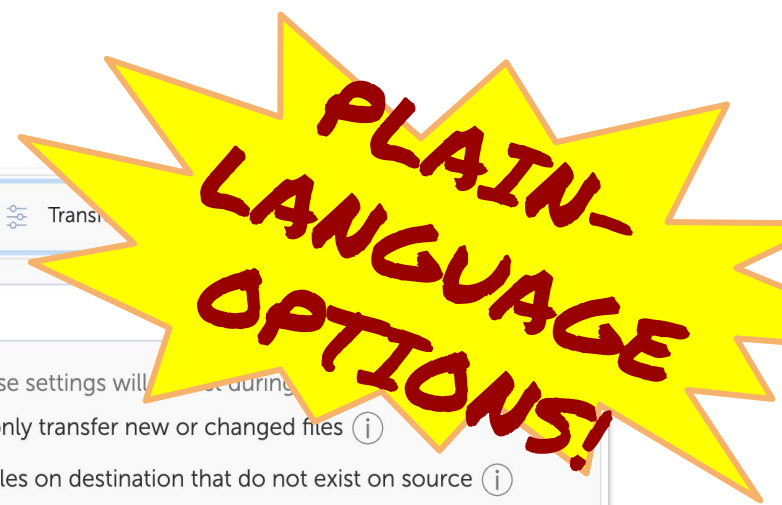
select all up one folder refresh list filter view

NAME ^	LAST MODIFIED	SIZE	TYPE
RStudio.tar	12/17/2020, 05:36 PM	133.12 KB	file
dev-retired2	1/27/2022, 12:47 PM	—	folder
dev-retired	12/3/2020, 02:00 PM	—	folder
dev	8/23/2022, 04:36 PM	—	folder
data	12/2/2020, 12:25 PM	—	folder

Transfer or Sync to...

- RStudio.tar
- dev-retired2
- dev-retired
- dev
- data

- Rename
- Delete Selected
- Download
- Open
- Get Link



Start

Transfer

Label This Transfer

Transfer Settings

NOTE: These settings will be used during the transfer.

- sync - only transfer new or changed files
- delete files on destination that do not exist on source
- preserve source file modification times
- do NOT verify file integrity after transfer
- encrypt transfer
- Skip files on source with errors
- Fail on quota errors

Notification Settings

- Disable success notification
- Disable failure notification
- Disable inactive notification

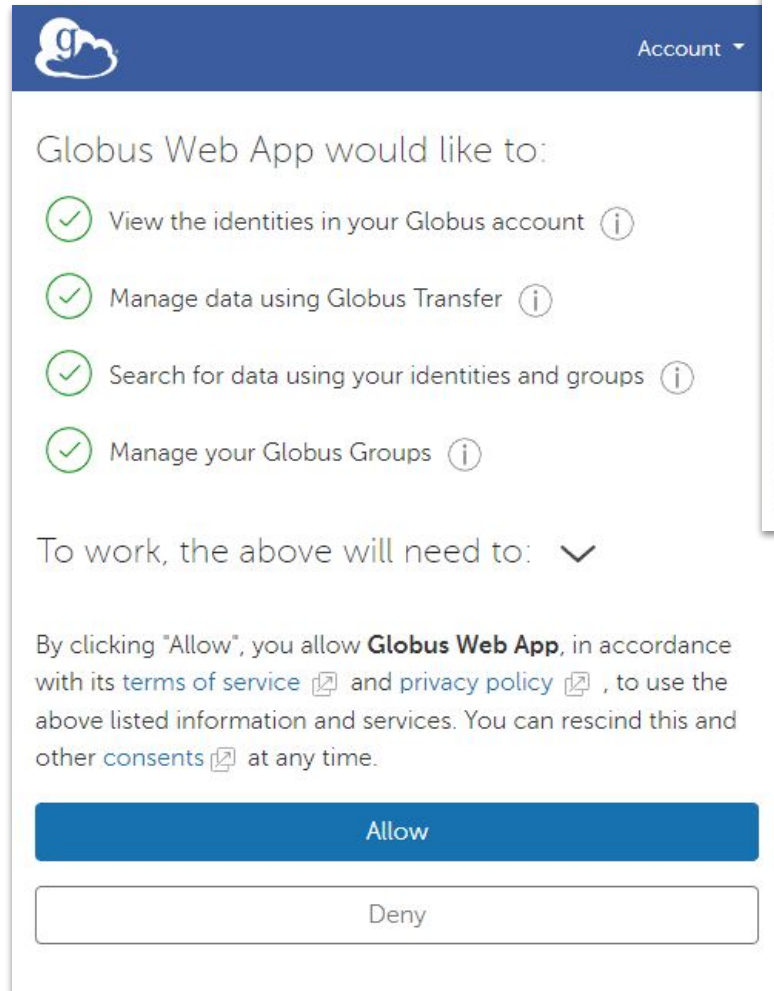
Schedule Start: mm/dd/yyyy, --:-- --

Repeat: does not repeat

Demo 'A': Quick Globus interface tour

About audit 'Consents'

The Globus service has many indiv. **capabilities** ...each one is individually authorized via these consent pop-ups



Account ▾

Globus Web App would like to:

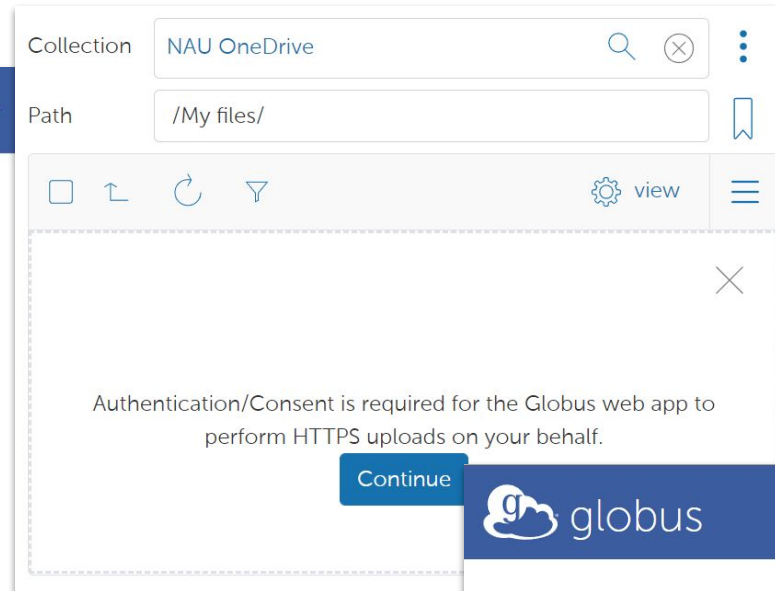
- ✓ View the identities in your Globus account ⓘ
- ✓ Manage data using Globus Transfer ⓘ
- ✓ Search for data using your identities and groups ⓘ
- ✓ Manage your Globus Groups ⓘ

To work, the above will need to: ▾

By clicking "Allow", you allow **Globus Web App**, in accordance with its [terms of service](#) and [privacy policy](#), to use the above listed information and services. You can rescind this and other consents [at any time](#).

Allow

Deny



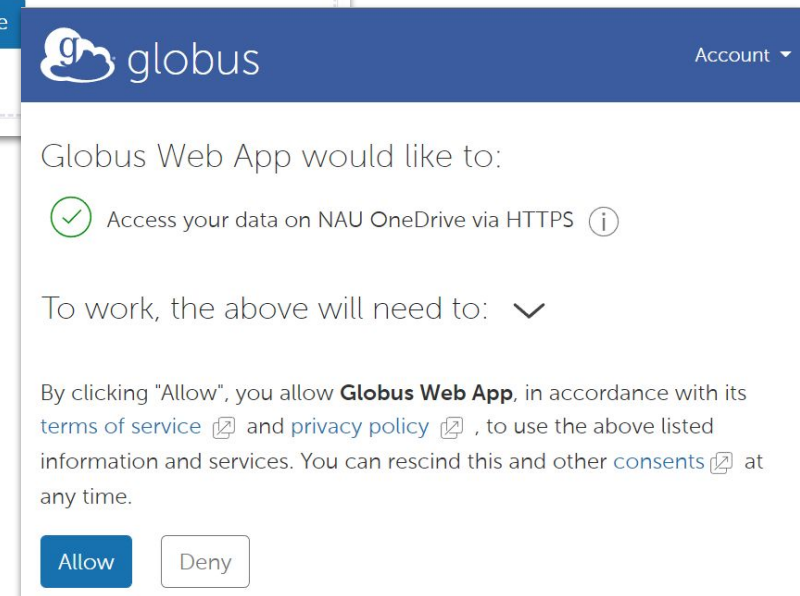
Collection NAU OneDrive 🔍 ⊗ ⋮

Path /My files/ 📌

📁 ↶ ↷ 🔍 ⚙️ view ≡

Authentication/Consent is required for the Globus web app to perform HTTPS uploads on your behalf.

Continue



globus Account ▾

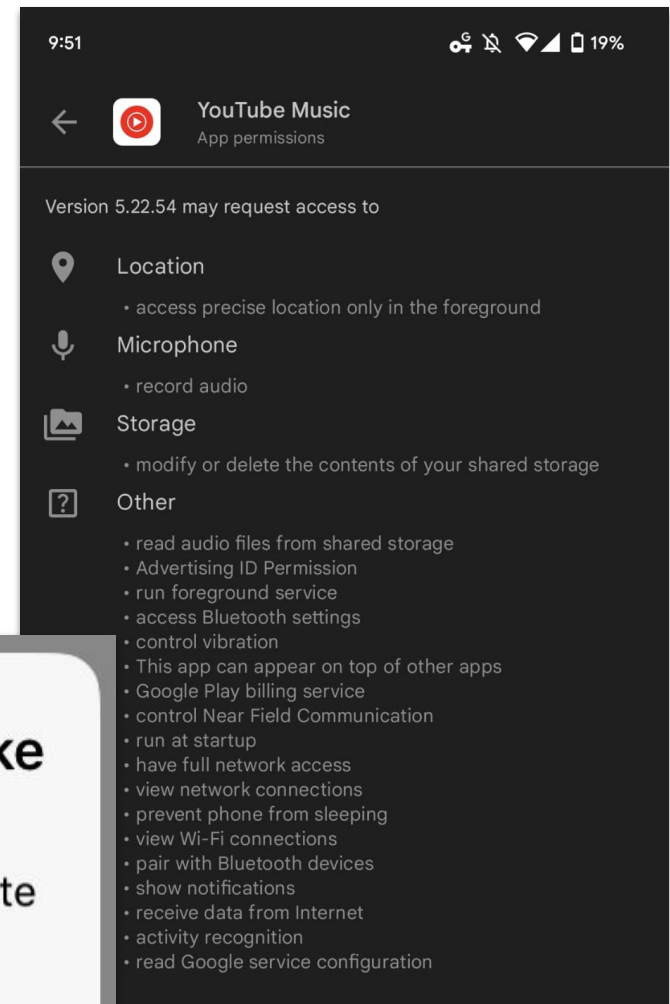
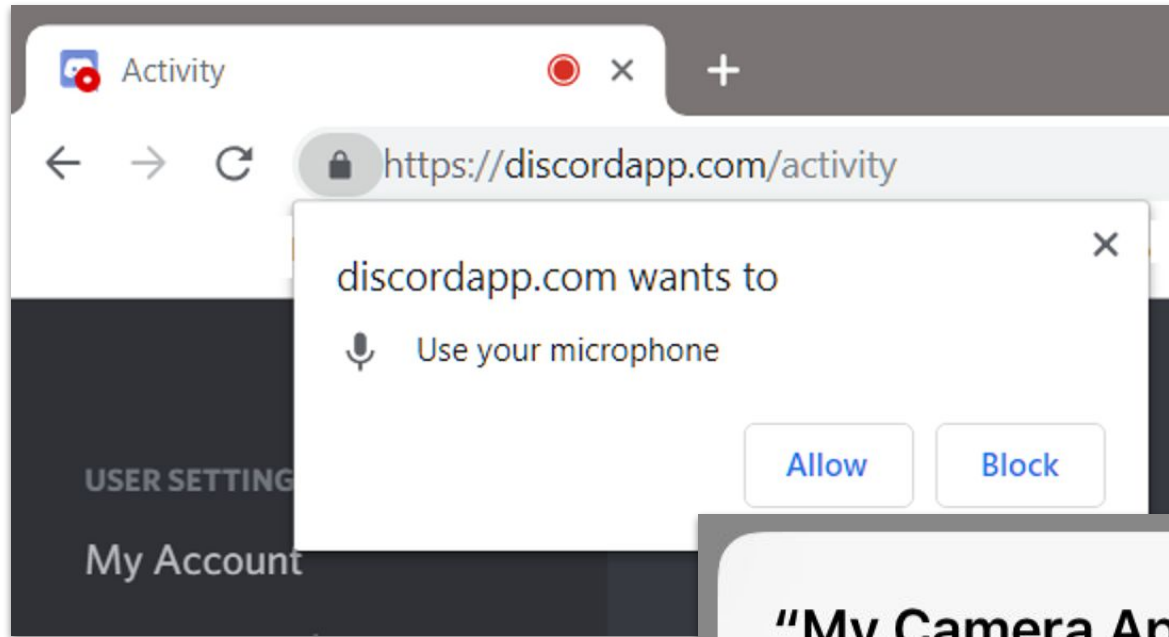
Globus Web App would like to:

- ✓ Access your data on NAU OneDrive via HTTPS ⓘ

To work, the above will need to: ▾

By clicking "Allow", you allow **Globus Web App**, in accordance with its [terms of service](#) and [privacy policy](#), to use the above listed information and services. You can rescind this and other consents [at any time](#).

Allow Deny



"My Camera App" Would Like to Access the Camera
This app uses the camera to take cute pictures of cats.

Don't Allow **OK**

Globus sign-in options

The screenshot displays the Globus sign-in interface. At the top, the Globus logo is visible. The main heading is "Log in to use Globus Web App". Below this, users are prompted to "Use your existing organizational login" with examples like "university, national lab, facility, project". A search box labeled "Look-up your organization..." is shown with a dropdown menu. A red arrow points from the search box to a magnified view of the dropdown results, which lists "arizona", "The University of Arizona", "Arizona State University", and "Northern Arizona University". Below the search box, a "Continue" button is present. A red arrow points from the "Continue" button to the "OR" separator. Below the separator are two buttons: "Sign in with Google" and "Sign in with ORCID iD". A red arrow points from the "Sign in with Google" button to the left. At the bottom, a link reads "Didn't find your organization? Then use Globus ID to sign in. (What's this?)".

globus

Log in to use Globus Web App

Use your existing organizational login
e.g., university, national lab, facility, project

Look-up your organization...

By selecting Continue, you agree to Globus [terms of service](#) and [privacy policy](#).

Continue

OR

Sign in with Google

Sign in with ORCID iD

Didn't find your organization? Then use [Globus ID to sign in](#). (What's this?)

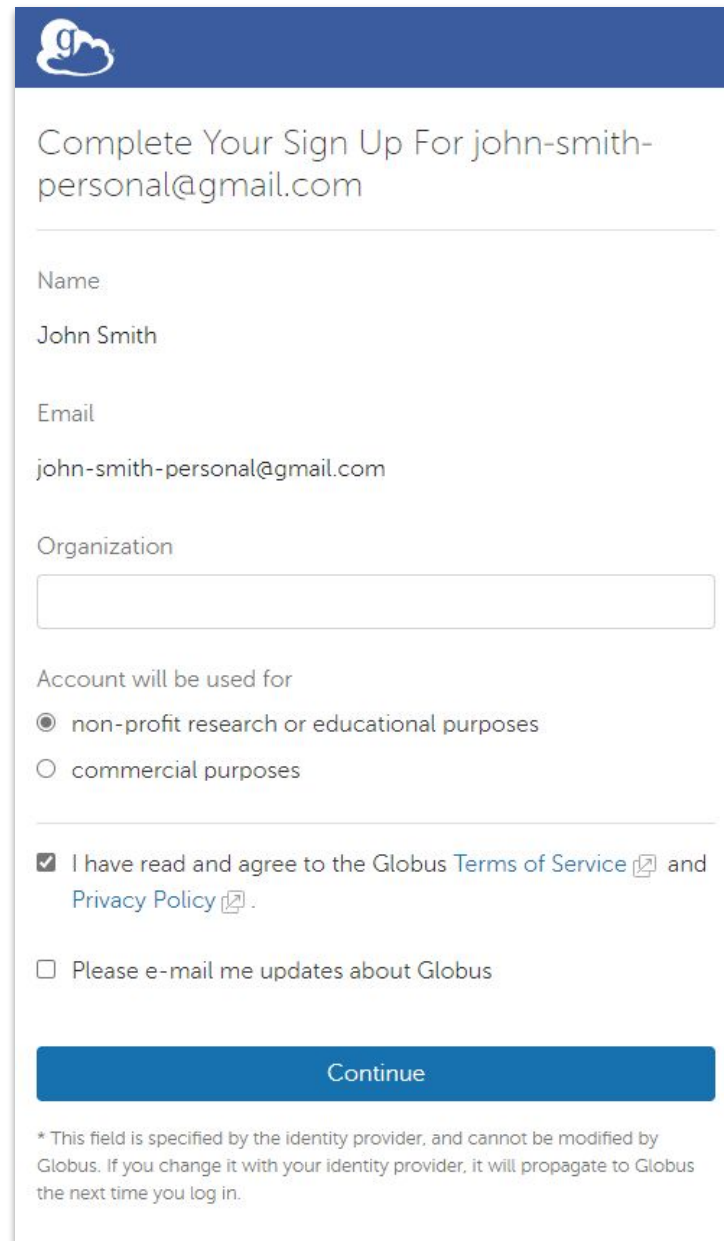
arizona

The University of Arizona

Arizona State University

Northern Arizona University

Non-NAU login with Globus



Complete Your Sign Up For john-smith-personal@gmail.com

Name
John Smith

Email
john-smith-personal@gmail.com

Organization

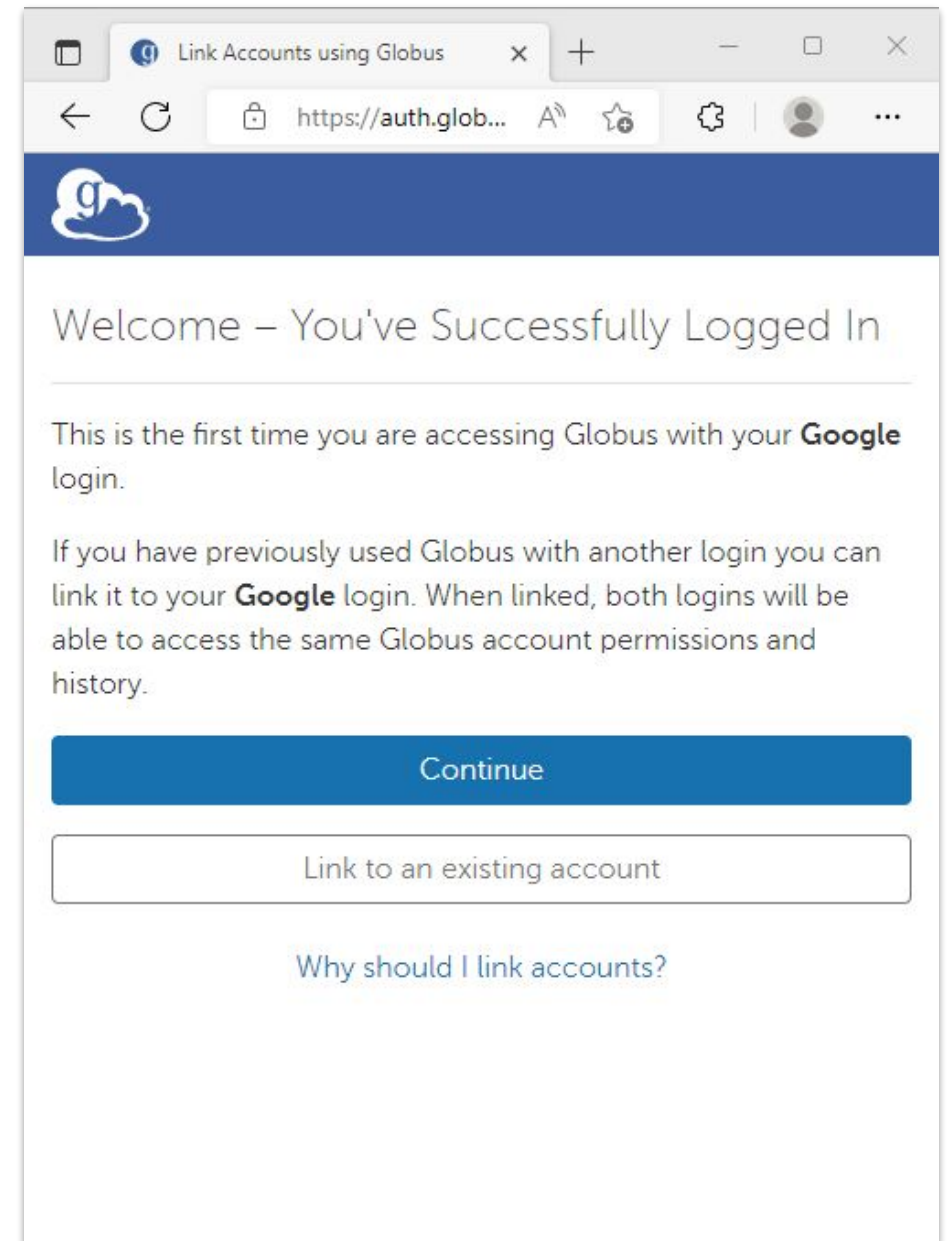
Account will be used for

- non-profit research or educational purposes
- commercial purposes

I have read and agree to the Globus [Terms of Service](#) and [Privacy Policy](#).

Please e-mail me updates about Globus

* This field is specified by the identity provider, and cannot be modified by Globus. If you change it with your identity provider, it will propagate to Globus the next time you log in.



Welcome – You've Successfully Logged In

This is the first time you are accessing Globus with your **Google** login.

If you have previously used Globus with another login you can link it to your **Google** login. When linked, both logins will be able to access the same Globus account permissions and history.

[Why should I link accounts?](#)

Exercise #1: login & browse your files

1. Visit app.globus.org in a browser
2. Select NAU under **Use your existing org login**, then **Continue**
3. Type “nau” in the **Collection** field to filter for Monsoon-based
4. Click on the **NAU HPC Filesystems** title *...and explore!*
5. Read and authorize all “consents” you are presented with

NOTE: Non-HPC users can

- Browse OneDrive & G-Drive,
- but not NAU HPC Filesystems

Cool stuff:

- You can *manually* change the Path field
- You can “bookmark” a folder within a collection
- What else caught your attention?

Transfer options (1/2)

- sync - only transfer new or changed files (i)

Selecting this option gives the ability to choose how files will be overwritten on the destination file system.

where
the

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

- delete files on destination that do not exist on source (i)

Files which do not exist on the source endpoint, but do exist on the destination endpoint will be deleted from the destination endpoint.

- preserve source file modification times (i)

After transfer of a file, the modification time on the destination will reflect the modification time on the source instead of the time the transfer completed.

- do NOT verify file integrity after transfer (i)

After transfer, verify that the source and destination file checksums match. If they don't, retransfer the entire file and keep trying until it succeeds. If a collection is configured to disable verification of file integrity, this option is ignored.

- encrypt transfer (i)

Encrypt the data while it is being transferred. If encryption is requested but is not supported by an endpoint, Globus will signal an error. Please visit <https://docs.globus.org/faq/security/> for more information on our encryption approach.

- Skip files on source with errors (i)

Files on source with "file not found" and "permission denied" errors will be skipped rather than cause the transfer to fail.

Transfer options (2/2)

Fail on quota errors ⓘ

If the storage quota is exceeded during the transfer, the transfer will fail rather than retry.

Disable success notification ⓘ

An email will not be sent when this transfer succeeds.

Disable failure notification ⓘ

An email will not be sent when this transfer fails.

Disable inactive notification ⓘ

An email will not be sent when this transfer becomes inactive.

Schedule Start

mm/dd/yyyy --:--:--



Repeat

does not repeat

does not repeat

days

hours

minutes

Globus use-cases

Globus is well suited for transferring data sets to and from an HPC resource like Monsoon, without the need for a command line and extensive technical knowledge.

This also comes in especially handy when a user wants to make a quick file transfer, and may not have access to a terminal: only a web browser is needed in most cases.

Globus is the best solution by far for transferring data between different collaborators, especially when the size of the data gets relatively large.

- ✨ Monsoon ↔ personal PCs*
- ✨ Monsoon ↔ other HPCs/Edus**
- ✨ My Monsoon ↔ *your* Monsoon
- ✨ $PC_A \leftrightarrow PC_B^*$
- ✨ Remote collaborator(s) ↔ Monsoon
 - NAU-affiliated, or not
 - Different slurm or enterprise group
- ✨ Time consuming & complex xfers
- ✨ Advanced
 - CLI tools (via conda) and APIs

Globus' capabilities and benefits

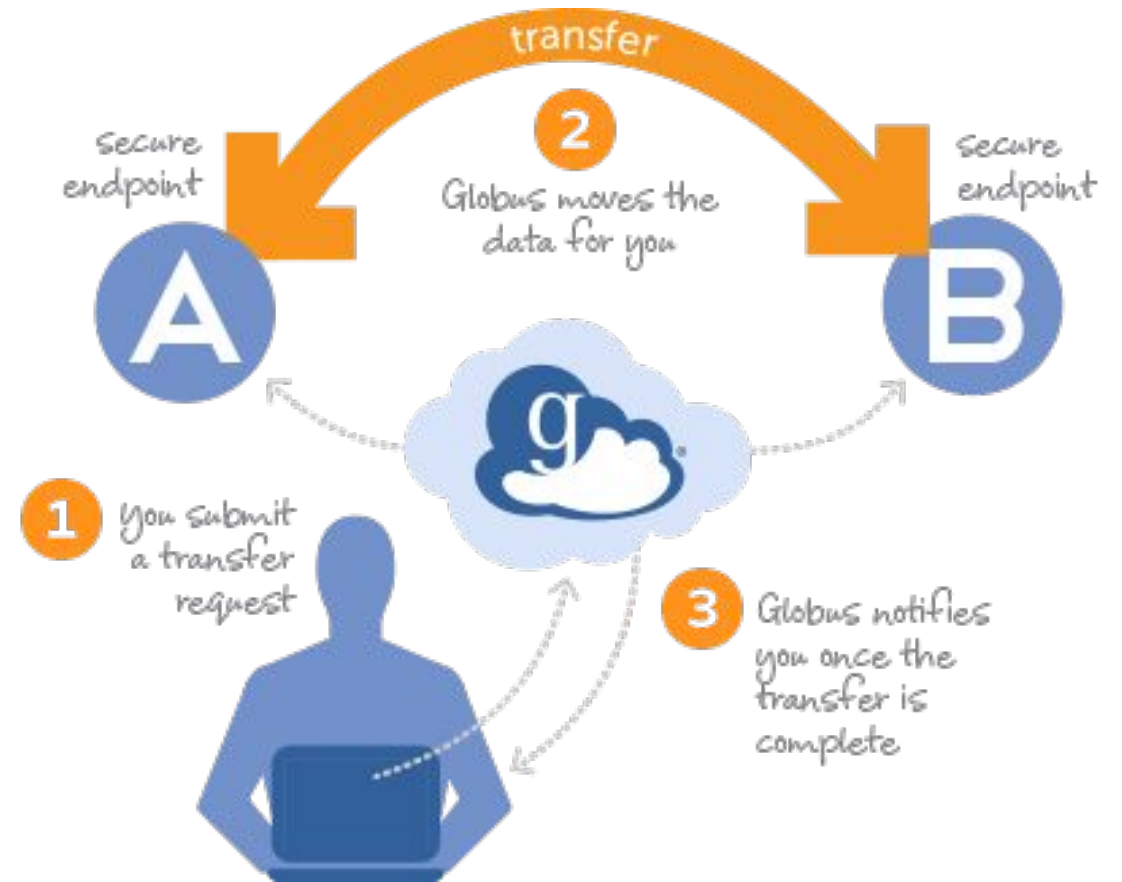
- Secure!
 - Highly customizable, granular permissions
 - Share as read-only or full-access
 - Require logins, or allow anonymous access
- Direct transfers between Monsoon ↔ *other institutions'* filesystems
 - Networked-filesystem to networked-filesystem!
 - Simple copies, or (schedulable!) sync tasks
- Also between Monsoon ↔ personal PCs, G-Drive, OneDrive...
- Web-based interface – no user-software required
 - CLI tools also available, as alternative to web

Globus 'collections'

A Globus collection is simply a server hosting Globus' filesharing software.

For example, Monsoon hosts the **NAU HPC Filesystems** collection by running some enterprise Globus software on our DTN1 login-node.

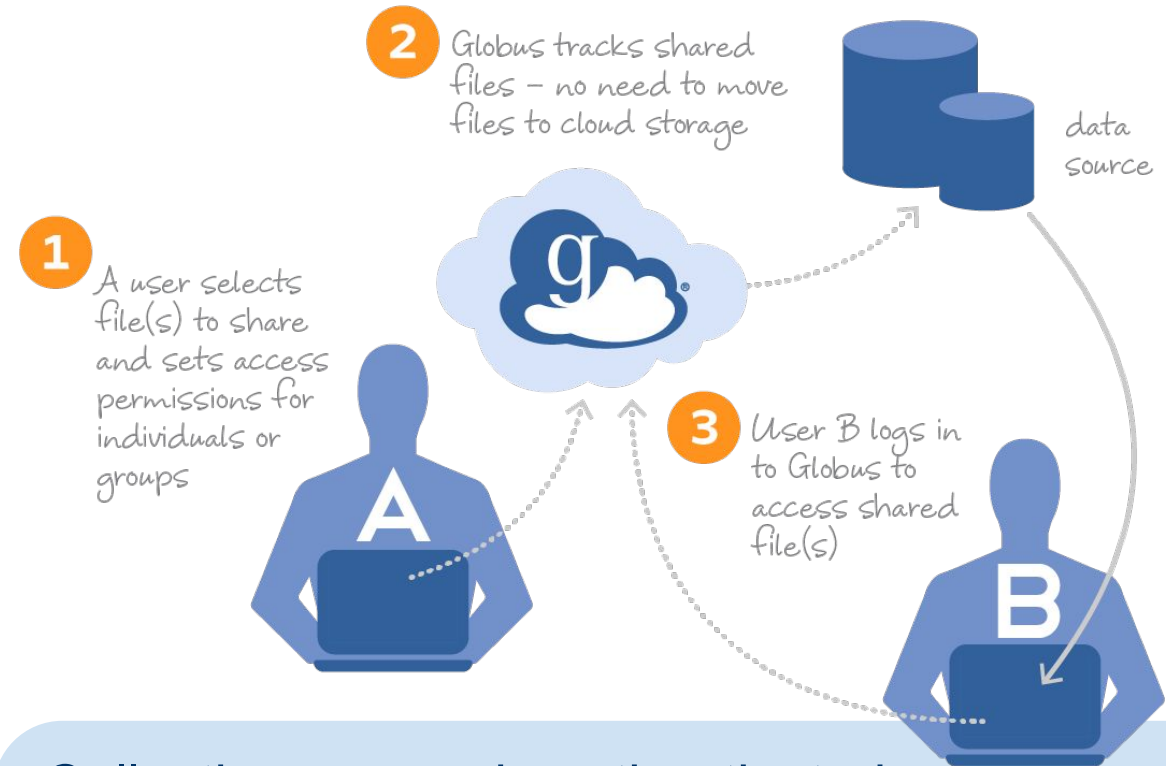
Individual users also can install/run Globus Connect Personal on their own desktop/laptop to enable transfers between PC ↔ any other endpoint.



What is a Globus 'collection'?

A collection is a named set of files which are hierarchically organized in folders, and associated with a specific storage gateway.

- **Mapped collection:** Access requires a user to have a local account. Their Globus identity is *mapped* to their local account and its local filesystem.
- **Guest collection:** Access is based on permissions explicitly granted by an authenticated collection owner.



Collection access is authenticated:

1. Users: Globus Auth-issued OAuth2 access tokens, and audit-consents
2. Files: with data access policies defined in the collection itself (ACLs)

Guest collection via Globus Personal

The image displays a macOS file manager window showing the 'Downloads' folder. The files listed are:

Name	Date Modified	Size
globusconnectpersonal-latest.dmg	Today at 1:07 PM	23.3 MB
PXL_20230212_212726211.mp4	Feb 13, 2023 at 9:00 AM	20.3 MB
PXL_20230215_...232186.PANO.jpg	Feb 14, 2023 at 7:29 PM	1.2 MB

Overlaid on this is the Globus Personal File Manager web interface. The browser address bar shows `app.globus.org/file-manager?origin_id=e4def596-ae35-...`. The interface shows a 'Collection' named 'jtb49 macbook downloads' and a 'Path' of '/Users/jtb49/Downloads/'. A table of files is displayed:

NAME	LAST MODIFIED	SIZE
globusconnectpersonal-lates...	2/16/2023,...	23.32 ME
PXL_20230212_212726211.m...	2/13/2023,...	20.32 ME
PXL_20230215_012232186.P...	2/14/2023,...	1.16 MB

The Globus interface includes a sidebar with navigation options: FILE MANAGER, BOOKMARKS, ACTIVITY, COLLECTIONS, GROUPS, and CONSOLE. The main area has a 'view' button and a list of file management icons on the right.

Guest collection via mapped collection

The image shows a terminal window and a web-based file manager interface. The terminal window displays the output of the `ls -l` command in a directory named `/projects/beekman_lab`. The output shows two files: `PROJ_DETAILS` (108 bytes) and `public_data` (a directory).


```
[jtb49@wind /projects/beekman_lab ]  
$ ls -l  
total 5  
-rw-----+ 1 root ITS-HPC-admin 108 Dec  7 15:07 PROJ_DETAILS  
drwxrwsr-x  2 root ITS-HPC-admin   2 Dec  7 15:07 public_data  
[jtb49@wind /projects/beekman_lab ]  
$
```

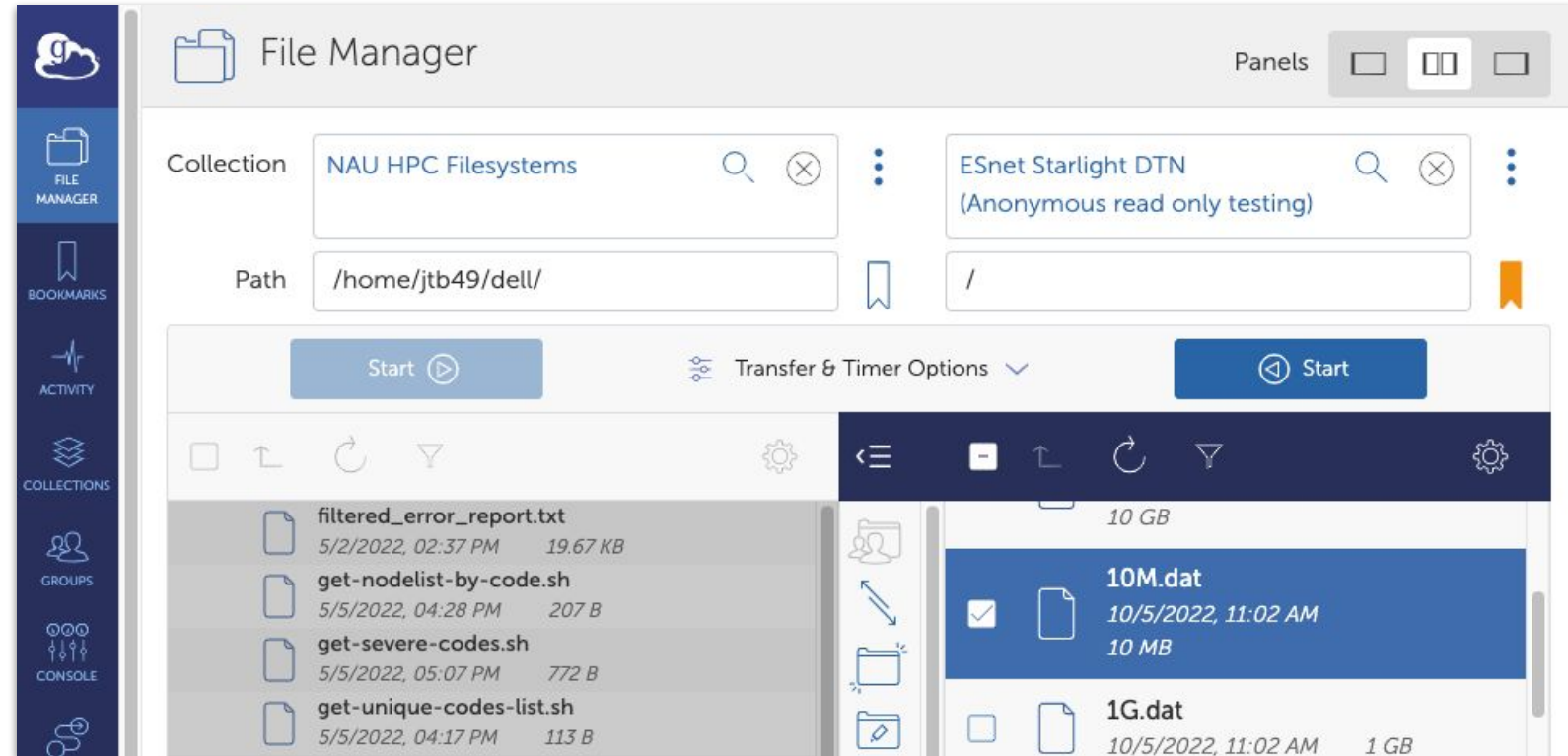
The web-based file manager interface, titled "File Manager | Globus", shows the "Beekman proj collection" selected. The path is `/`. The interface displays a table of files and folders:

NAME	LAST MODIFIED	SIZE
PROJ_DETAILS	12/7/2022, 03:07PM	108 B
public_data	12/7/2022, 03:07PM	-

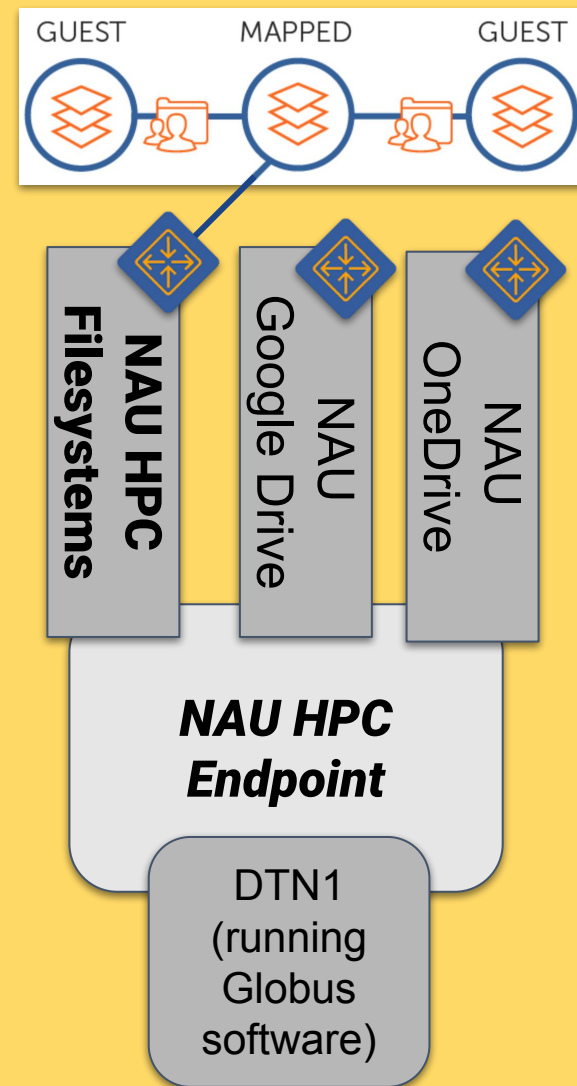
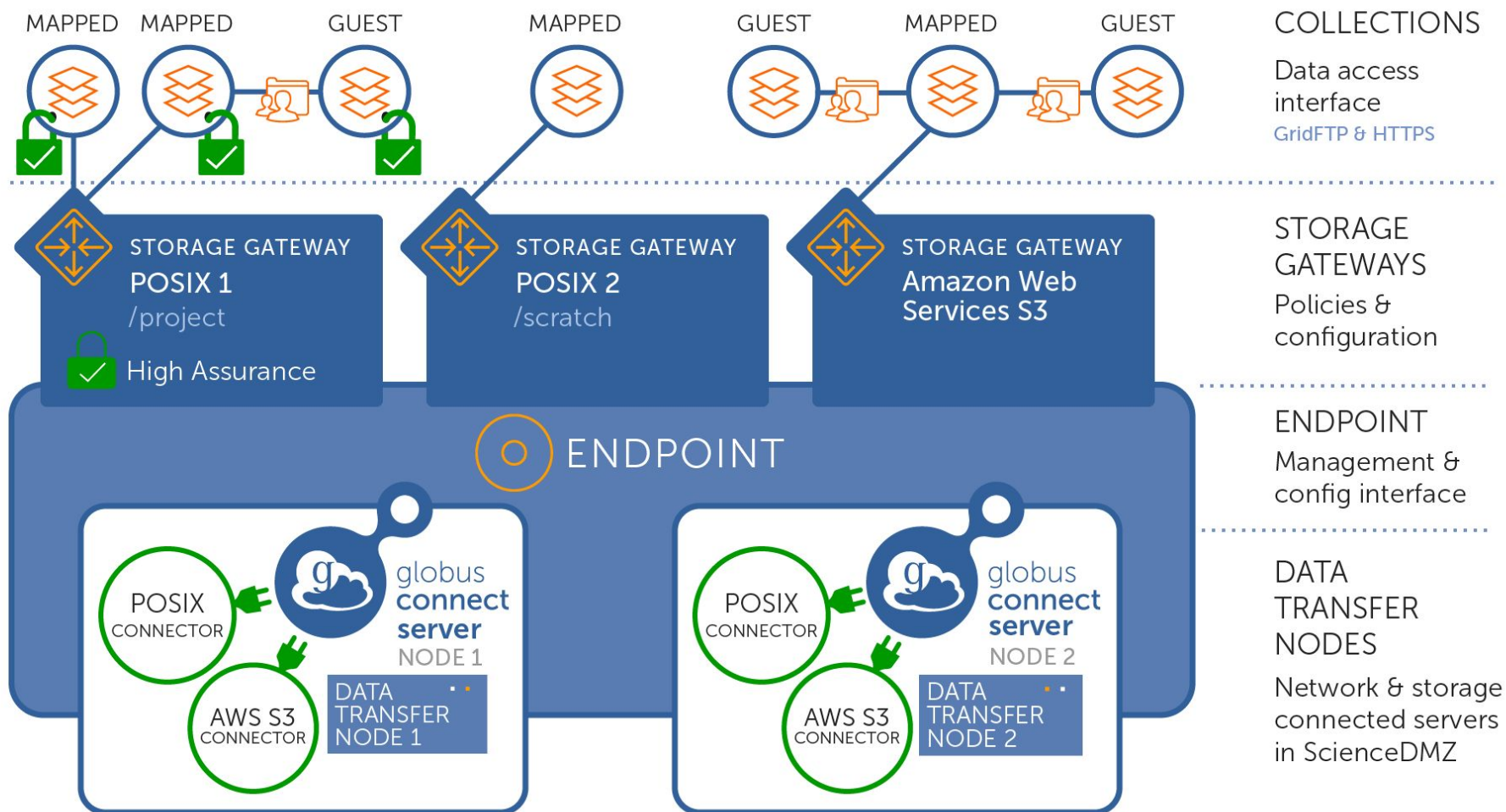
Demo 'B': Transfer between collections, via web

Exercise #2: copy a small file to Monsoon

- Set two-panel view
- In the new panel:
 - Find either your G-/OneDrive or this **ESnet** collection →
 - Select a file
 - Click 
- Check status via *sidebar* > **ACTIVITY**
- Bonus: upload a file from your desktop



Architecture & Interface



Collab/Sharing via Guest Collections


File Manager

Collection: NAU HPC Filesystems


Path: /projects/

select none [up one folder](#) [refresh list](#)

NAME ▾

<input checked="" type="checkbox"/>	 beekman_lab
-------------------------------------	---

- Share
- Transfer or Sync to...
- Rename
- Delete Selected

NAU HPC Filesystems 

[Overview](#) [Collections](#) [Credentials](#)

Guest Collections

[+ Add a Guest Collection](#)

COLLECTION NAME	
asdfasdf	Overview Permissions Roles

Adding a Guest Collection

- Think of a collection like its own pre-defined filesystem
 - “/” in the file-browser equates to the chosen directory →
- **On top of** your Guest Collection, you will define Access Control Lists for files & directories
- Share with
 - Specific Globus users
 - All Globus-auth users
 - Anonymous/web

Create New Guest Collection

You are creating a guest collection on nauhpc#dtn1 to share data

Directory: /projects/beekman_lab/ [Browse](#) [?](#)

Display Name: Beekman proj collection

Description: Base level of project area, for staging deeper permissions-based access

Keywords: genomics, Higgs boson, climate change

Force encryption on transfers to and from this collection

[view more fields](#) ▾

[Create Collection](#) [Cancel](#)

Friendliest UI in the game!

BECKMAN PROJ COLLECTION
Add Permissions - Share With

Path: [Browse](#)

Share With

- user - share with specific individuals
- group - make data accessible to members of a group
- all users - make data accessible to all logged in users of Globus
- public (anonymous) - make data accessible to everyone

User: [Change](#)

Send Email No email will be sent to this user

(Permissions) read write

[Add Permissions](#)

BECKMAN PROJ COLLECTION
Add Permissions - Share With

Path:

Share With

- user - share with specific individuals
- group - make data accessible to members of a group
- all users - make data accessible to all logged in users of Globus
- public (anonymous) - make data accessible to everyone

Who:

(Permissions) read write

[Add Permission](#) [Cancel](#)

Beekman proj collection

Overview Permissions Roles

Shared With [Add Permissions - Share With](#)


USER OR GROUP	CREATED	READ	WRITE	
Path: / Show link for sharing				
NAU HPC Endpoint (9b7bebd2-94da-4b73-8093-78d3d03889dd@clients.auth.globus.org)	-	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Key
Jason Buechler (jtb49@nau.edu)	-	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Key
Path: /bulk_data/ Show link for sharing				
All Users	11/17/2022, 06:29 PM	<input checked="" type="checkbox"/>	<input type="checkbox"/>	Trash
Jason Buechler (buechler@gmail.com)	11/17/2022, 06:35 PM	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Trash

Demo 'C': setup a shared area

Exercise #3: create/share a Guest Collection

- Create a new directory in your home (e.g. ~/jb_collection)
- “Share” it, via side-toolbar or right-click
- Click “Add a guest collection” and fill-out required fields
- Click “Create collection”
 - Note you (and NAU’s “endpoint”) have default access
- Click “Add Permissions – Share With”
 - Give access to another NAU user
 - Give read-only access to a non-nau email address (click **Add** even if not found)

When you forget what path this collection uses

 jtb project on monsoon

[Overview](#) [Permissions](#) [Roles](#)

Display Name	jtb project on monsoon	Manage Consent
Entity Type	Guest Collection (GCS)	
Endpoint	NAU HPC Endpoint	Edit Attributes
Domain		

JTB PROJECT ON MONSOON
Edit Attributes

Display Name

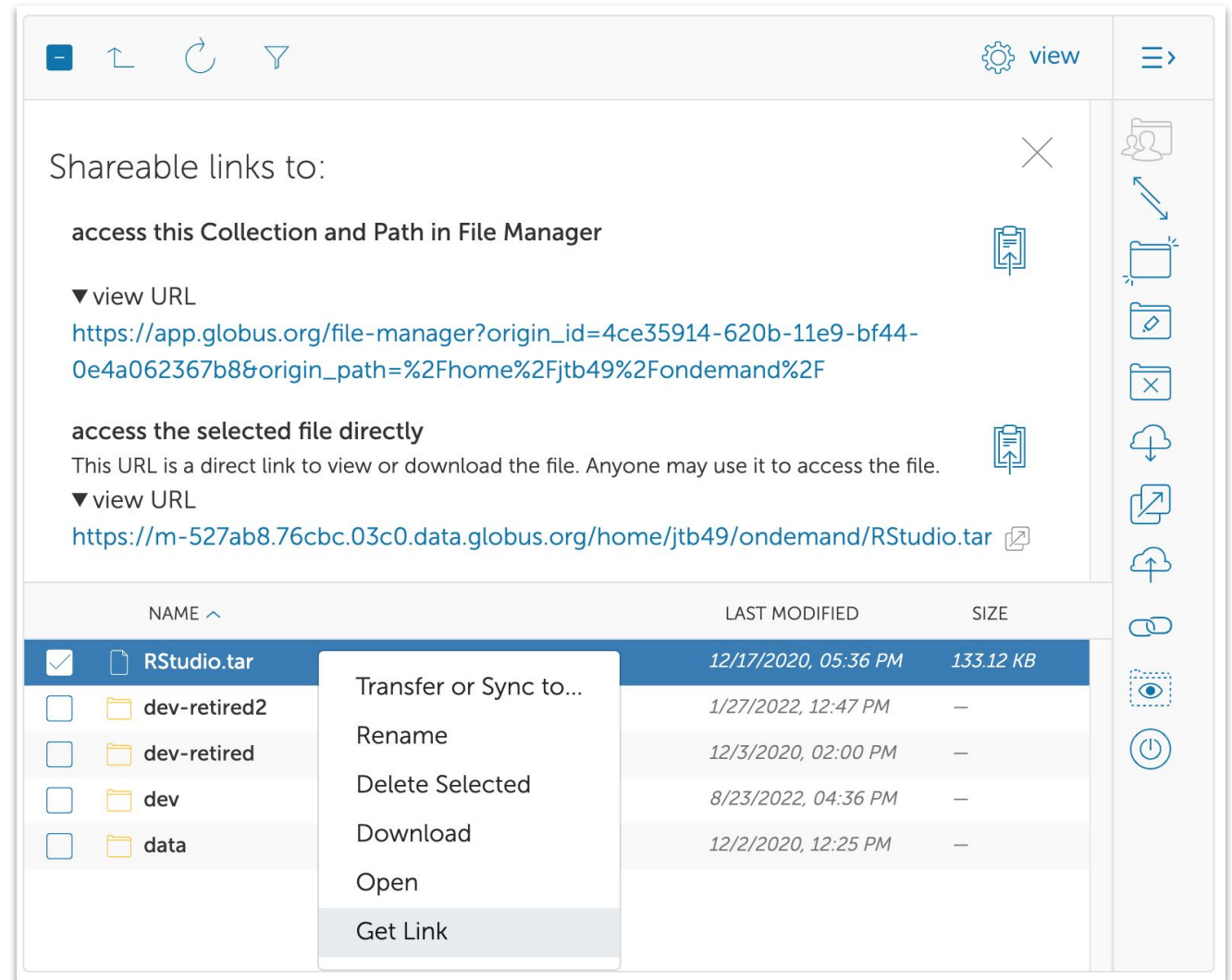
Default Directory [Browse](#) [?](#)

Force encryption on transfers to and from this collection

[Submit](#) [Cancel](#)

Sharing data (via link)

- Two kinds of globus links
 - 1st: “view” link
(Goes to a specific folder)
 - 2nd: “open file”
(Not available for dirs)
- Share individual files
 - Specific Globus users
 - All Globus-auth users
 - Anonymous/web



The screenshot displays the Globus File Manager interface. At the top, there are navigation icons (back, forward, refresh, search) and a 'view' button. Below the navigation bar, there is a section for 'Shareable links to:' with a close button (X). This section contains two options:

- access this Collection and Path in File Manager**: Includes a 'view URL' and a long URL: https://app.globus.org/file-manager?origin_id=4ce35914-620b-11e9-bf44-0e4a062367b8&origin_path=%2Fhome%2Fjtb49%2Fondemand%2F
- access the selected file directly**: Includes a note: 'This URL is a direct link to view or download the file. Anyone may use it to access the file.' and a 'view URL' pointing to: <https://m-527ab8.76cbc.03c0.data.globus.org/home/jtb49/ondemand/RStudio.tar>

Below the links, there is a table listing files and folders. A context menu is open over the 'RStudio.tar' file, showing options: Transfer or Sync to..., Rename, Delete Selected, Download, Open, and Get Link.

NAME ^	LAST MODIFIED	SIZE
<input checked="" type="checkbox"/> RStudio.tar	12/17/2020, 05:36 PM	133.12 KB
<input type="checkbox"/> dev-retired2	1/27/2022, 12:47 PM	—
<input type="checkbox"/> dev-retired	12/3/2020, 02:00 PM	—
<input type="checkbox"/> dev	8/23/2022, 04:36 PM	—
<input type="checkbox"/> data	12/2/2020, 12:25 PM	—

Sharing with non-NAU, new-to-Globus users

Jason Buechler (jtb49@nau.edu) shared folder "/bulk_data") on "Beekman Lab" with jason@po.cwru.edu Inbox x



Globus Notification <no-reply@globus.org>

10:57 AM (34 minutes ago)



to jtb3 ▾

Globus user Jason Buechler (jtb49@nau.edu) shared the folder "/bulk_data" on the endpoint "Beekman Lab" (endpoint id: d8628595-1653-4486-a08a-532a2af9534f) with jtb3@po.cwru.edu.

Additional Message:

hi, me

Use this URL to access the share:

https://app.globus.org/file-manager?&origin_id=d8628595-1653-4486-a08a-532a2af9534f&origin_path=/&add_identity=39276e6a-6be3-4cb8-a4af-b5994777e4eb

The Globus Team

support@globus.org

A collaborator uploads to Monsoon via web browser

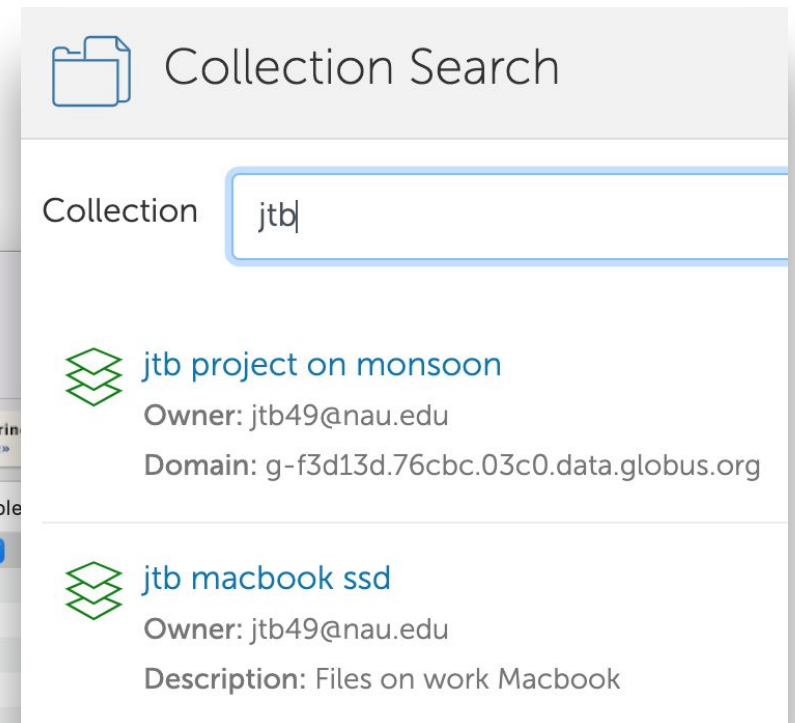
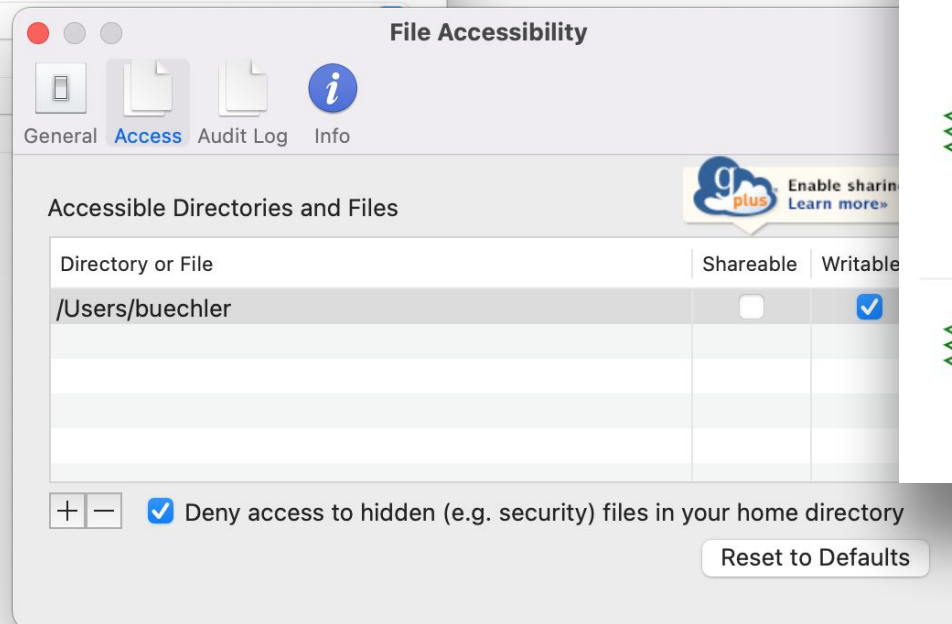
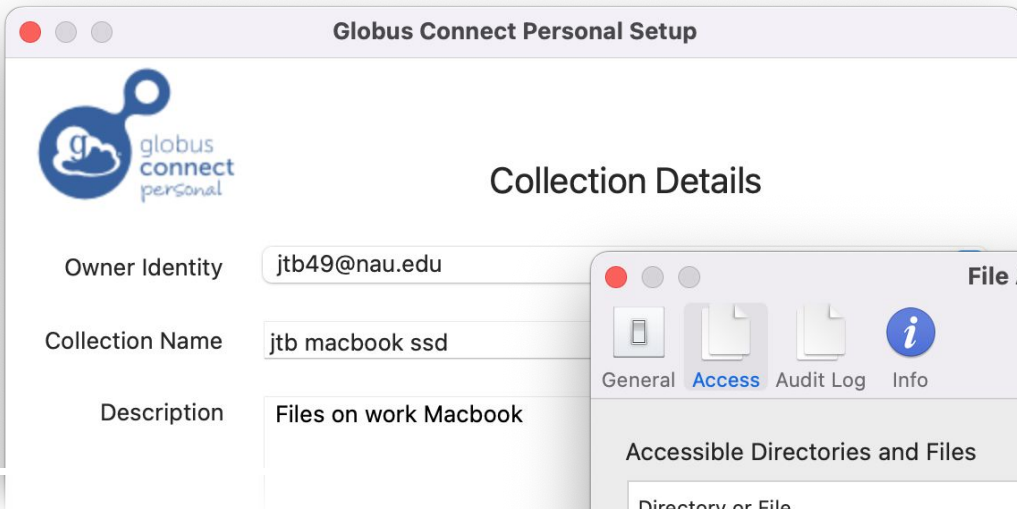
The screenshot displays a web-based File Manager interface. The main window shows a collection named "Beekman proj collection" at the path "/bulk_data/". A table lists files with columns for NAME, LAST MODIFIED, and SIZE. An "Uploads" window is open in the foreground, showing a list of files with green checkmarks indicating successful uploads.

NAME	LAST MODIFIED	SIZE
cluster_workshop_ondemand_jb.pdf	11/17/2022, 05...	1.58 MB
cluster_workshop_ondemand_jb.pptx	11/17/2022, 05...	7.17 MB
cluster_workshop_ondemand_orig.p...	11/17/2022, 05...	6.79 MB
cluster_workshop_ondemand_orig.pptx	11/17/2022, 05...	4.32 MB
ondemand-editor.png	11/17/2022, 05...	212.67 KB
cluster_workshop_ondemand_jb.pptx	11/17/2022, 05...	

Uploads

- cluster_workshop_ondemand.pdf ✓
- cluster_workshop_ondemand_orig.pptx ✓
- ondemand-editor.png ✓
- cluster_workshop_ondemand_jb.pptx ✓

Download from: www.globus.org/globus-connect-personal



Globus CLI

```
$ module load anaconda3 && conda activate globus
```

```
(globus): $ globus endpoint search nau | grep NAU | tr -s ' '
```

```
9b7bebd2-94da-4b73-8093-78d3d03889dd | cbc@nau.edu | NAU HPC Endpoint  
aa21c1e8-ba46-4e73-879b-50c9f946a887 | cbc@nau.edu | NAU Google Drive  
4f4bbfcb-005f-40d7-97ee-9cba31b40bfc | cbc@nau.edu | NAU HPC Filesystems  
b81db7c2-e355-420d-b5ce-6e903ad24594 | cbc@nau.edu | NAU OneDrive
```

NOTE 1: in Globus CLI, use ``endpoint search``
to find both collections AND endpoints

NOTE 2: Globus web usually hides “actual”
endpoints

```
$ globus ls b81db7c2-e355-420d-b5ce-6e903ad24594
```

```
The collection you are trying to access data on requires you to grant consent for the Globus CLI to access it.
```

```
message: Missing required data_access consent
```

```
Please run
```

```
  globus session consent 'urn:globus:auth.....-6e903ad24594/data_access]'  
to login with the required scopes
```

```
$ globus session consent 'urn:globus:auth.....-6e903ad24594/data_access]'
```

```
Please authenticate with Globus here:
```

```
-----
```

```
https://auth.globus.org/v2/oauth2/authorize?cli.....ne&prompt=login
```

```
-----
```

```
Enter the resulting Authorization Code here: uPAbsSuBw1QhXjivE0PLo1430BBHeS
```

```
You have successfully updated your CLI session.
```

```
$ globus ls b81db7c2-e355-420d-b5ce-6e903ad24594
```

```
Contacts/
```

```
Desktop/
```

```
Documents/
```

```
Downloads/
```

```
Microsoft Teams Chat Files/
```

A Globus CLI transfer

```
$ globus transfer -r --encrypt b81db7c2-e355-420d-b5ce-6e903ad24594:"Microsoft Teams Chat Files" 4f4bbfcb-005f-40d7-97ee-9cba31b40bfc:/home/jtb49/demo/
```

The collection you are trying to access data on requires you to grant consent for the Globus CLI to access it.

message: Missing required data_access consent

Please run

```
globus session consent 'urn:globus:auth.....-6e903ad24594/data_access]'
to login with the required scopes
```

```
$ globus session consent 'urn:globus:auth.....-6e903ad24594/data_access]'
```

Please authenticate with Globus here:

<https://auth.globus.org/v2/oauth2/authorize?cli.....ne&prompt=login>

Enter the resulting Authorization Code here: YSiSa2UXnW0euFzMxt0l8TPha9z7s9

You have successfully updated your CLI session.

```
$ globus transfer -r --encrypt b81db7c2-e355-420d-b5ce-6e903ad24594:"Microsoft Teams Chat Files" 4f4bbfcb-005f-40d7-97ee-9cba31b40bfc:/home/jtb49/demo/
```

Message: The transfer has been accepted and a task has been created and queued for execution

Task ID: 6c971d9a-703a-11ed-8423-8d2923bcb3ec

```
$ globus task show 6c971d9a-703a-11ed-8423-8d2923bcb3ec | grep -v Subtasks
```

```
Label:           None
Task ID:         6c971d9a-703a-11ed-8423-8d2923bcb3ec
Is Paused:       False
Type:            TRANSFER
Directories:     1
Files:           131
Status:          SUCCEEDED
Request Time:    2022-11-29T23:06:20+00:00
Faults:          0
Completion Time: 2022-11-29T23:06:58+00:00
Source Endpoint: NAU OneDrive
Source Endpoint ID: b81db7c2-e355-420d-b5ce-6e903ad24594
Destination Endpoint: NAU HPC Filesystems
Destination Endpoint ID: 4f4bbfcb-005f-40d7-97ee-9cba31b40bfc
Bytes Transferred: 296728329
Bytes Per Second: 7613995
```

```
$ globus task event-list 6c971d9a-703a-11ed-8423-8d2923bcb3ec
```

Time	Code	Is Error	Details
-----	-----	-----	-----
2022-11-29T23:06:58+00:00	SUCCEEDED	False	{"files_succeeded":131}...
2022-11-29T23:06:58+00:00	PROGRESS	False	{"bytes_transferred":29...
2022-11-29T23:06:23+00:00	STARTED	False	{"concurrency":2,"para1...

Thank You