NAU
NORTHERN
ARIZONA
UNIVERSITY

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 the local 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.

How your bytes become a file

- **Local**
  - *SOFTWARE* ➡️ disk
  - disk ➡️ OS ➡️ disk
  - disk ➡️ *SOFTWARE* ➡️ disk
- **Remote**
  - *remote* disk

➡️**SOFTWARE**

➡️network

➡️**SOFTWARE**

➡️*local* disk

How your bytes become a file (II)

- **Local**
  - *WORD PROCESSOR* ➡️ disk
  - *WEB BROWSER* ➡️ disk
  - *GLOBUS PERSONAL* ➡️ disk
- **Remote**
  - *remote* HPC disk (e.g. @ ASU)
    - ➡️**GLOBUS SW** (running on cluster)
      - ➡️network
        - ➡️**GLOBUS SW** (on Monsoon)
          - ➡️*local* HPC disk (e.g. /home, /scratch, /projects, etc)

Architecture of a Globus transfer

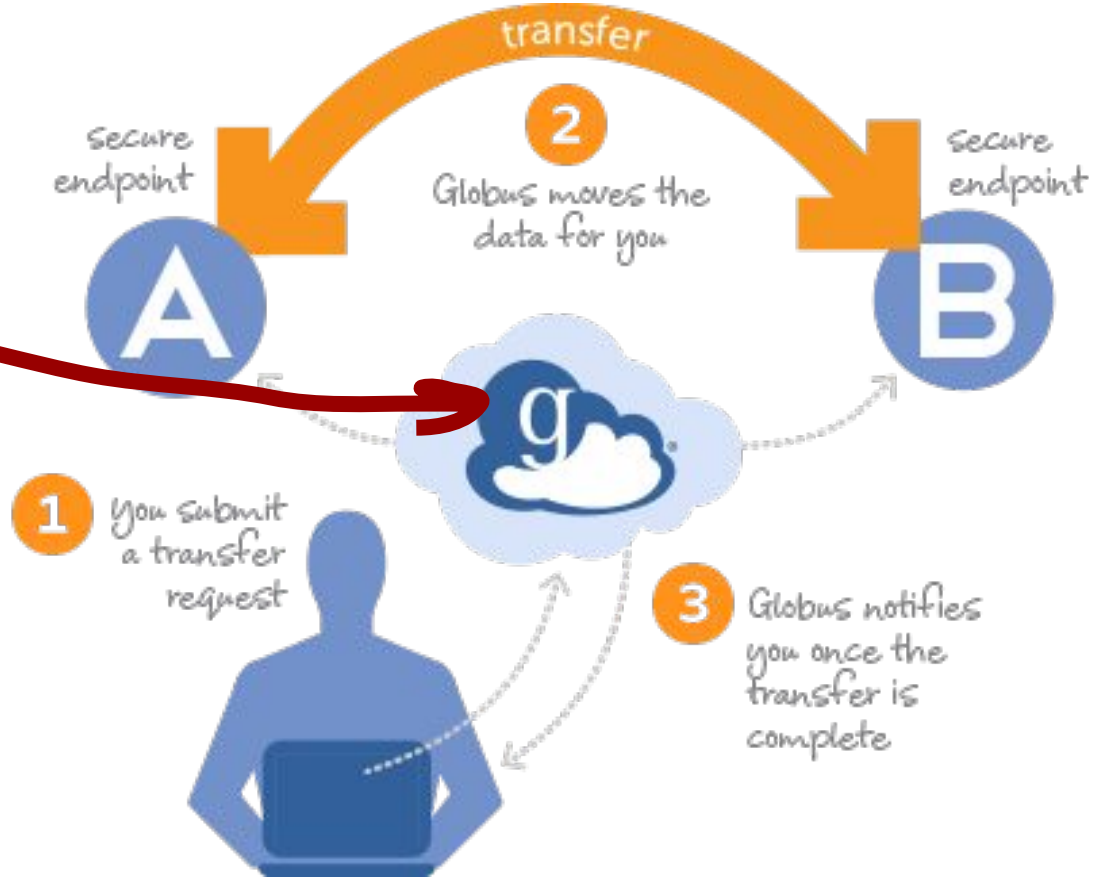
remote HPC disk (e.g. @ ASU)

➡ GLOBUS SW (running on cluster, 'A')

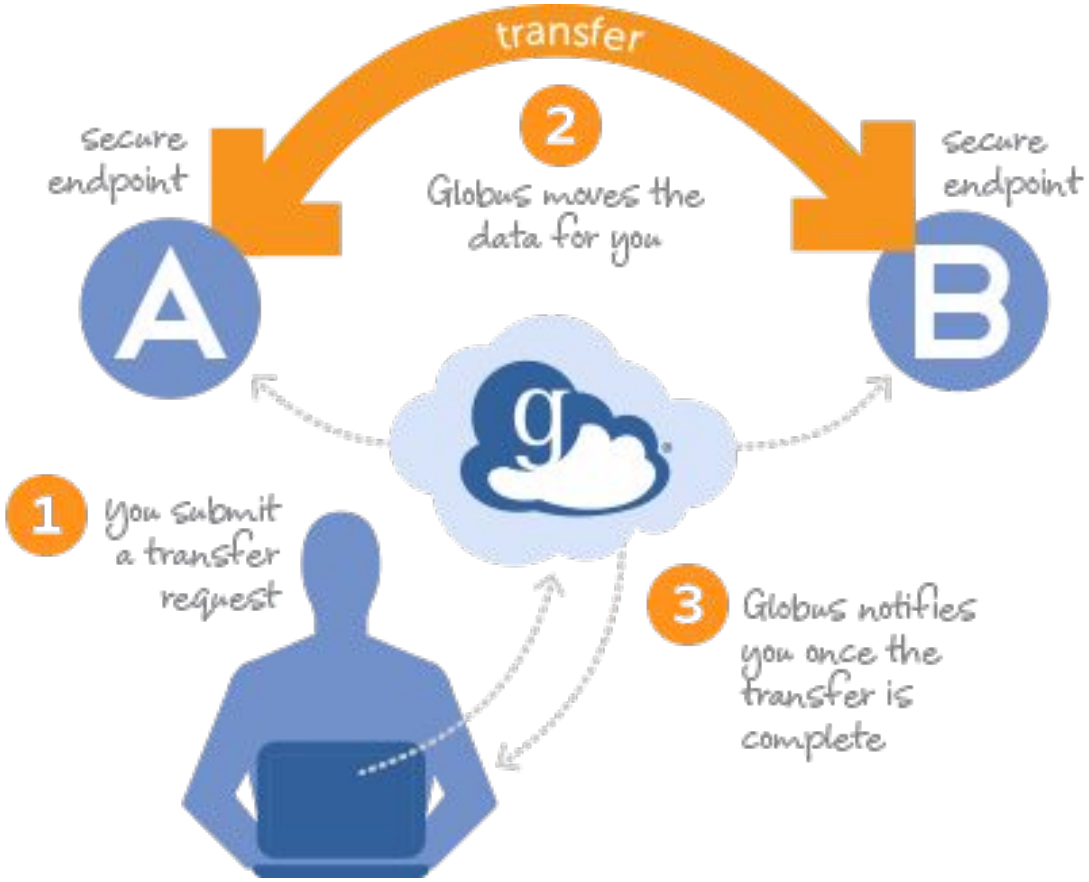
➡ network

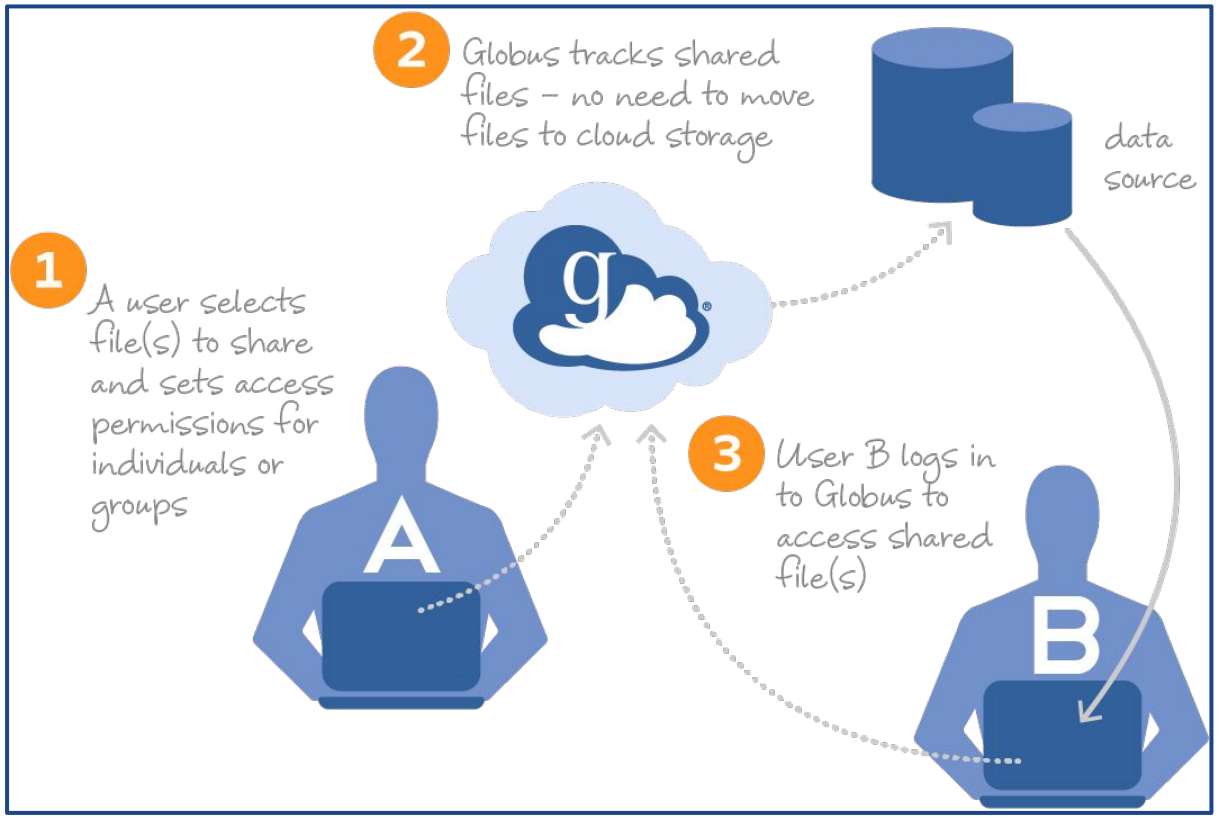
➡ GLOBUS SW (on Monsoon, 'B')

➡ *local* HPC disk (e.g. /home, /scratch, /projects, etc)



Globus transfers

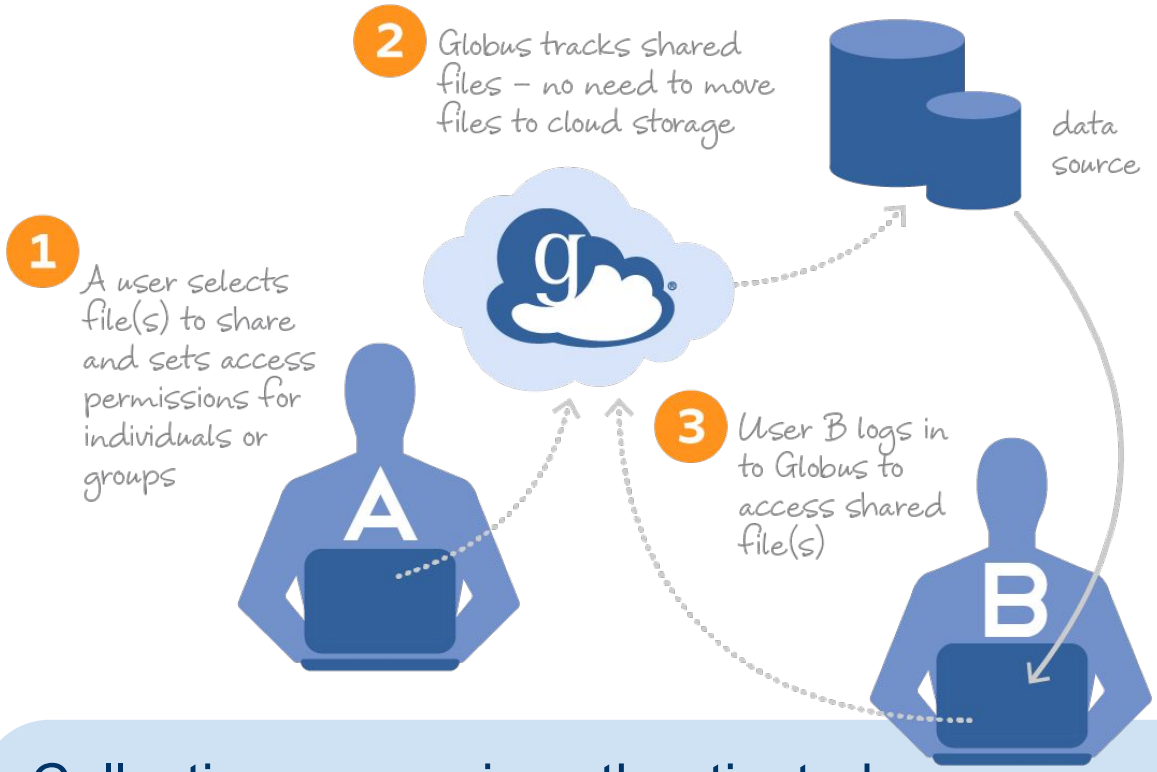




What is a Globus ‘collection’?

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

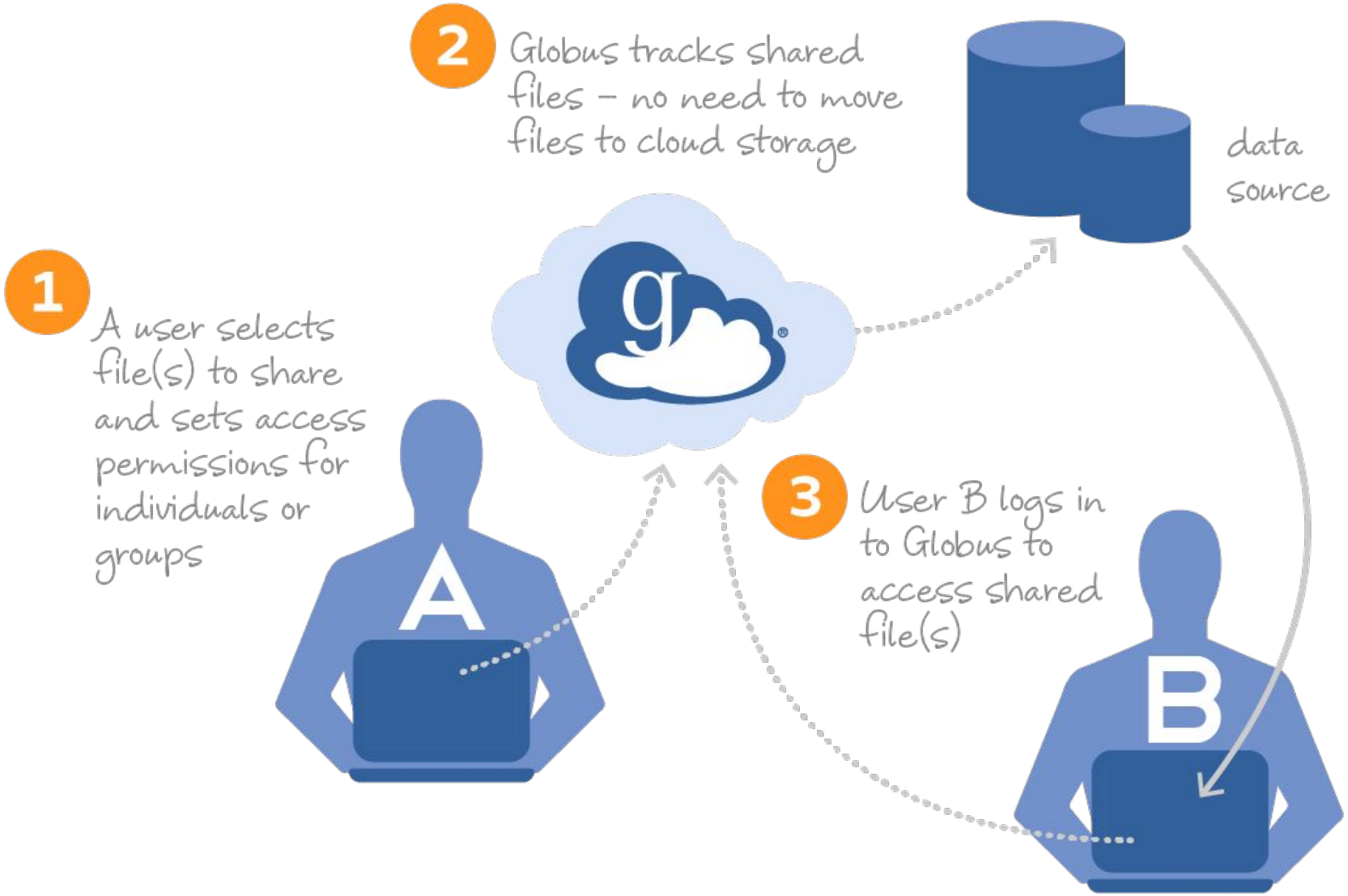
- **Mapped collection**: Access requires a user to have a local account. Their Globus identity *is mapped* to their local account and its local filesystem.
- **Guest collection**: Access is based on permissions explicitly granted by an authenticated collection owner.



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)

What is a Globus ‘collection’?

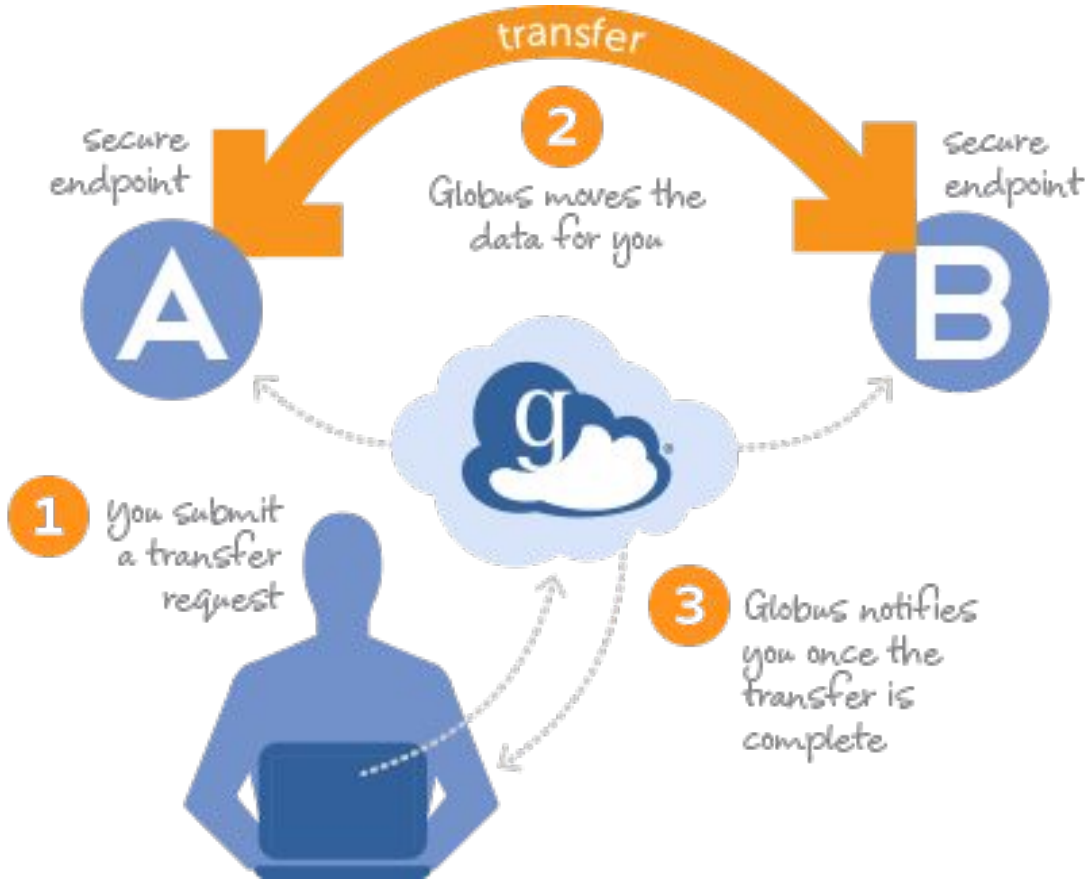


Globus ‘collections’

A Globus “collection” is simply a pre-defined set of files accessible via Globus’ filesharing software.

For example, Monsoon hosts the **NAU HPC Filesystems** collection through Globus’ server-software (which runs on the dtn1 login-node).

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



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 or 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 a fantastic solution for transferring data between/among different collaborators, *especially* when the size of the data or number of files gets relatively large.

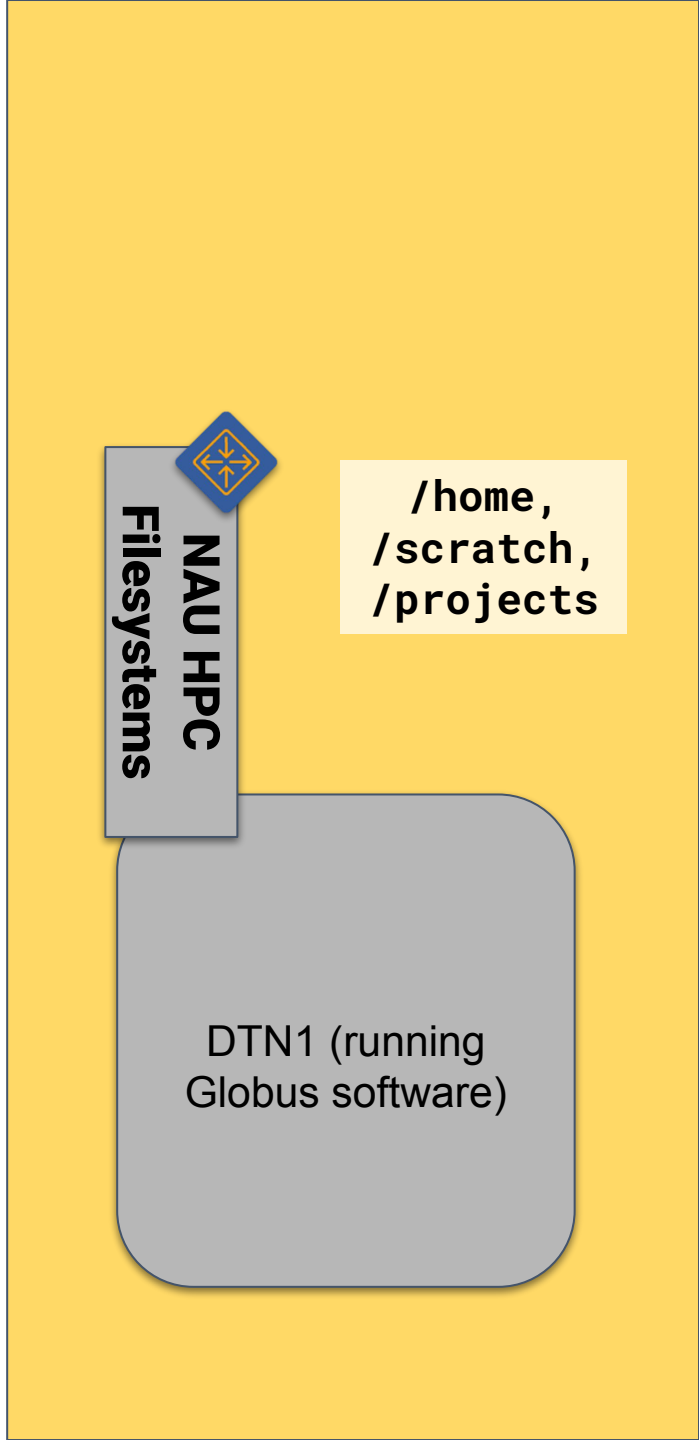
- Monsoon ↔ personal PCs*
- Monsoon ↔ other HPCs/Edus**
- My Monsoon ↔ *your* Monsoon
- PC_A ↔ PC_B*
- Remote collaborator(s) ↔ Monsoon
  - NAU-affiliated, or not
  - Different slurm or enterprise group
- Time consuming & complex xfers
- Advanced
  - CLI tools (via conda) and APIs

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

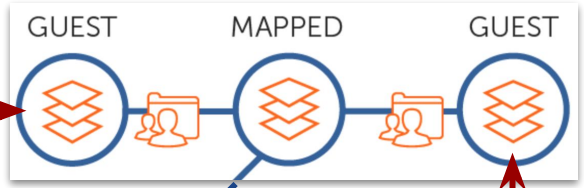
Globus ‘collections’

- A Globus “collection” is simply a pre-defined set of files accessible via Globus’ filesharing software.
- For example, Monsoon hosts the **NAU HPC Filesystems** collection through Globus’ server-software (which runs on the dtn1 login-node).
- Individual users may also install/run Globus Connect Personal on their own desktop/laptop to enable transfers between PC ↔ any other endpoint.



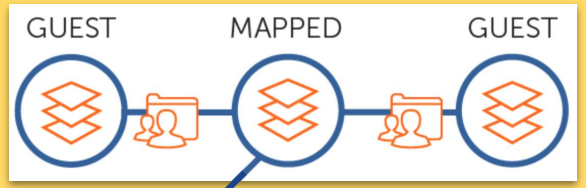
Globus ‘collections’

A “guest collection” is created by YOU!



YOU control sharing-access to guest collections!

NAU HPC Filesystems

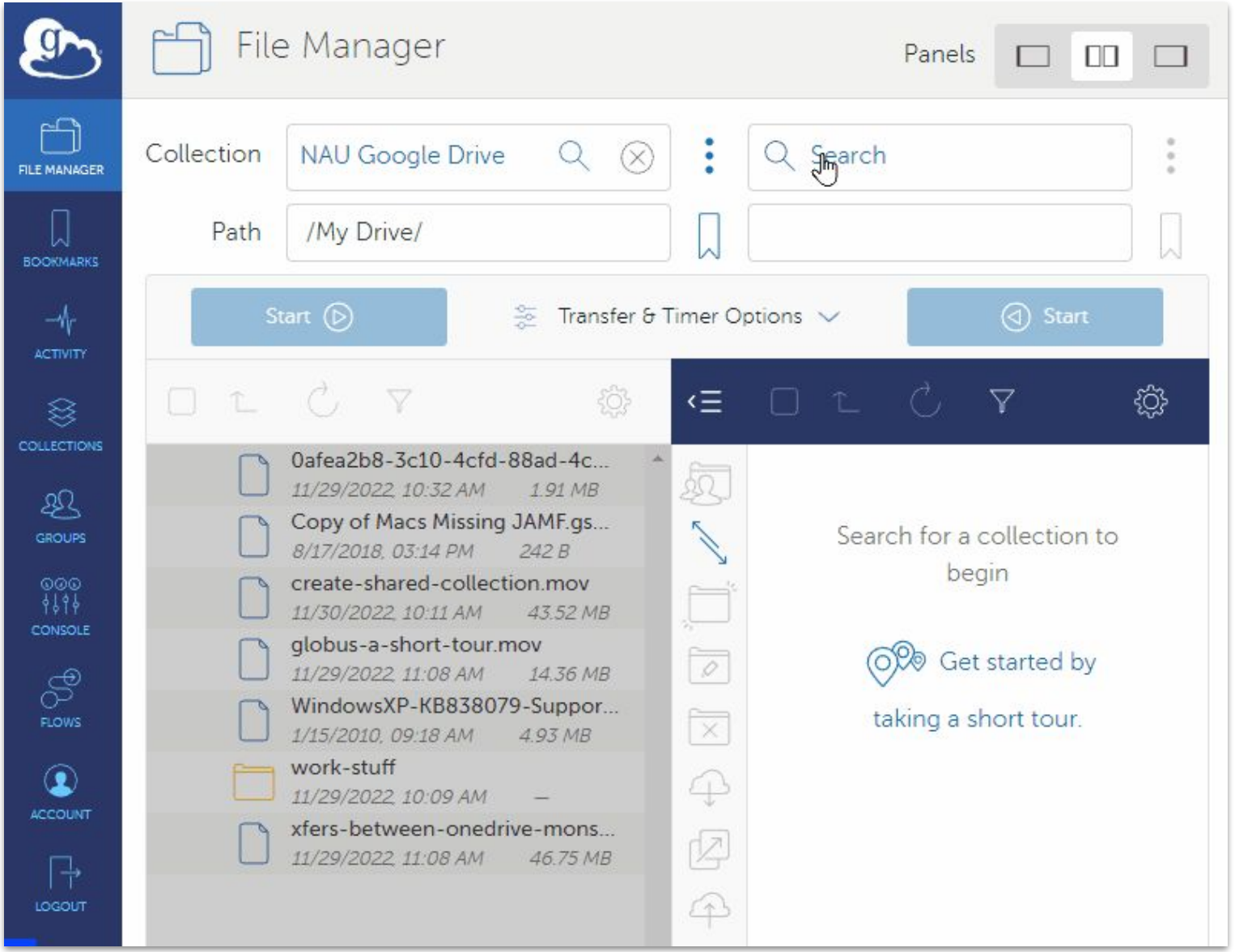


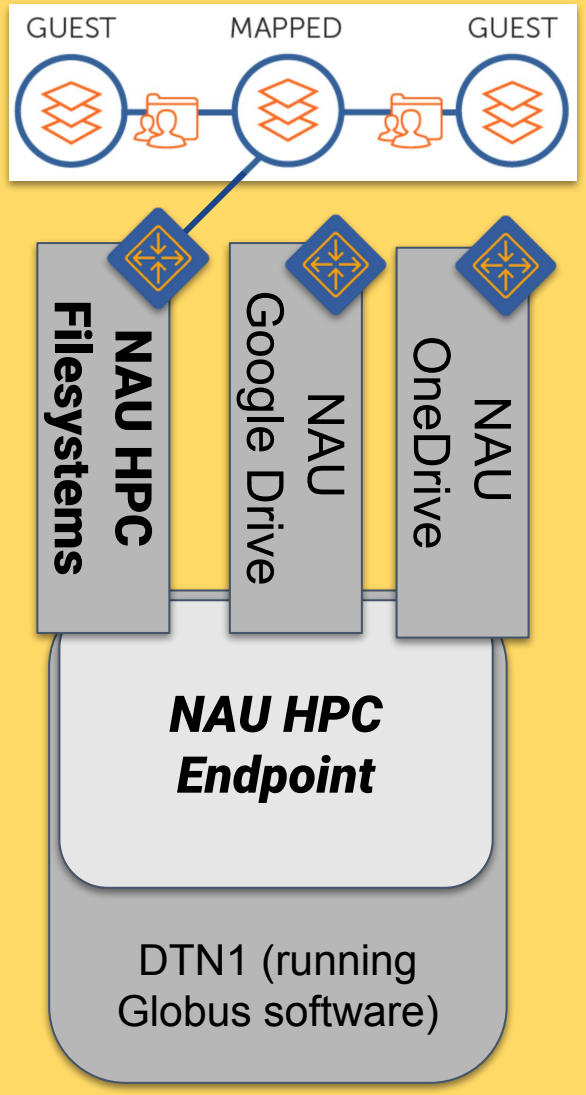
NAU HPC Filesystems

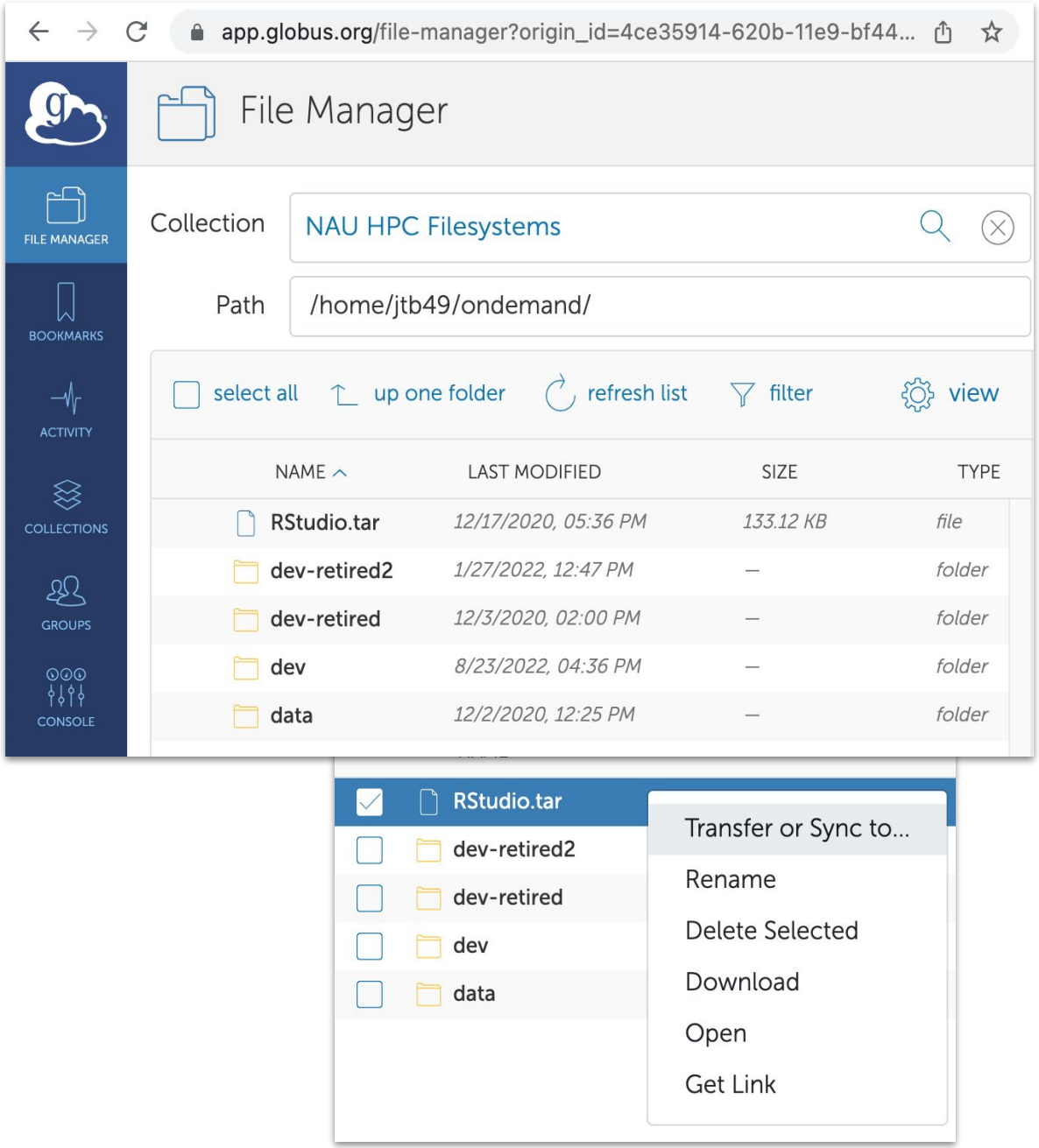
NAU HPC Endpoint

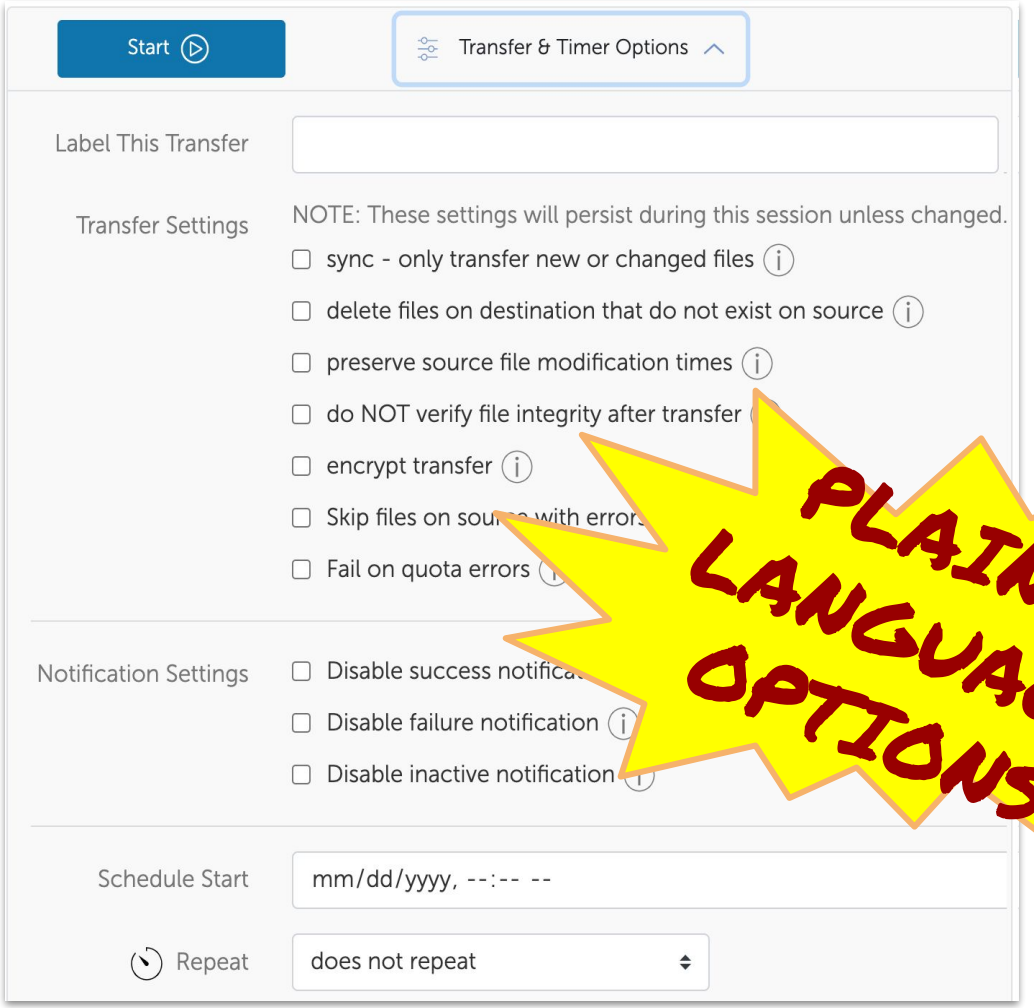
DTN1 (running Globus software)

Interface: Example transfer



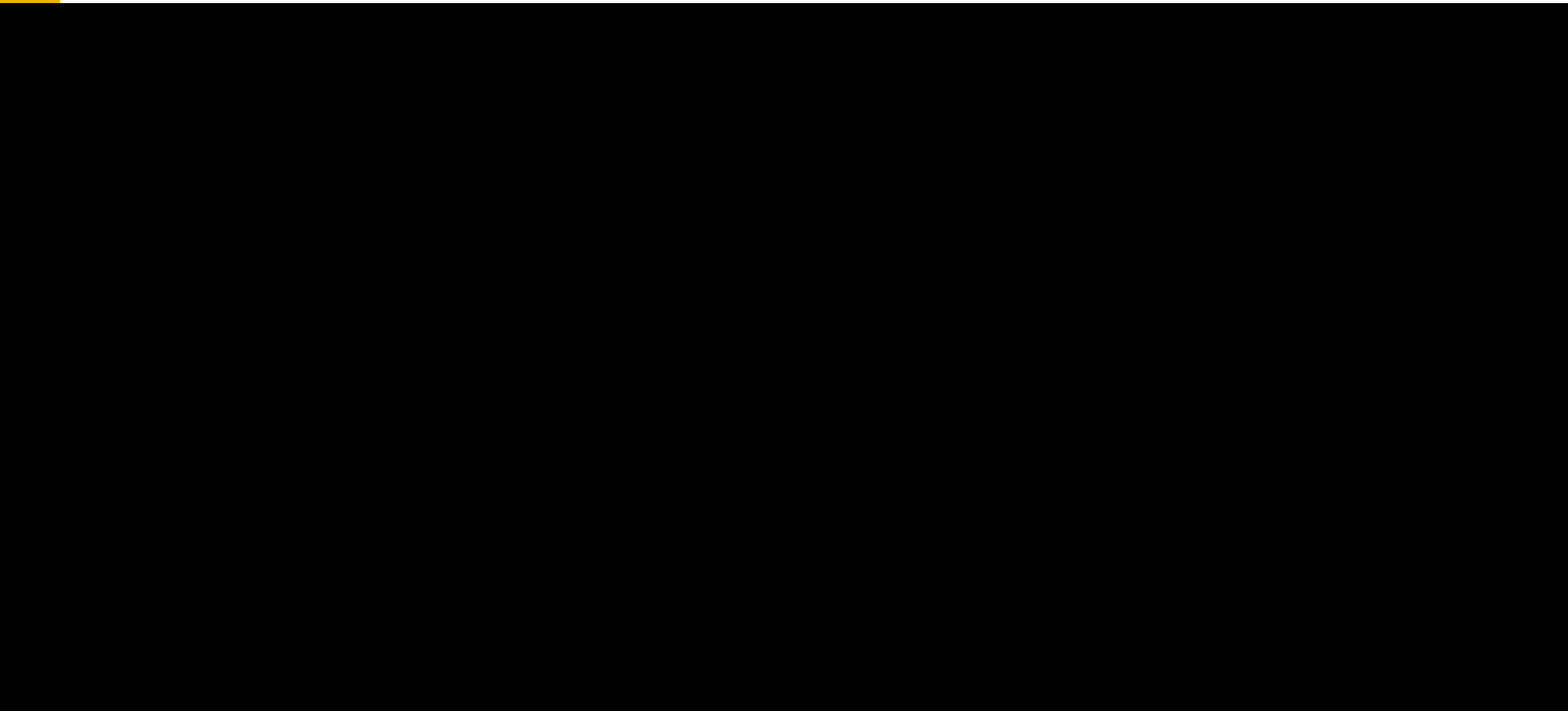




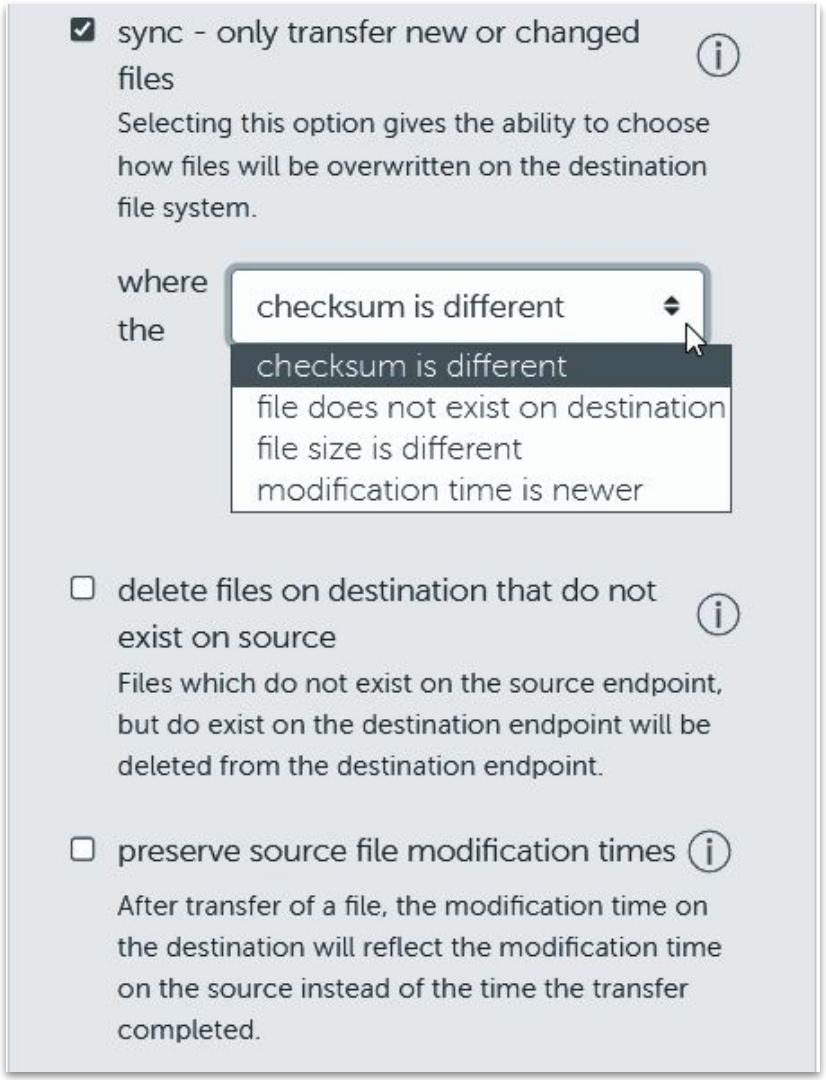


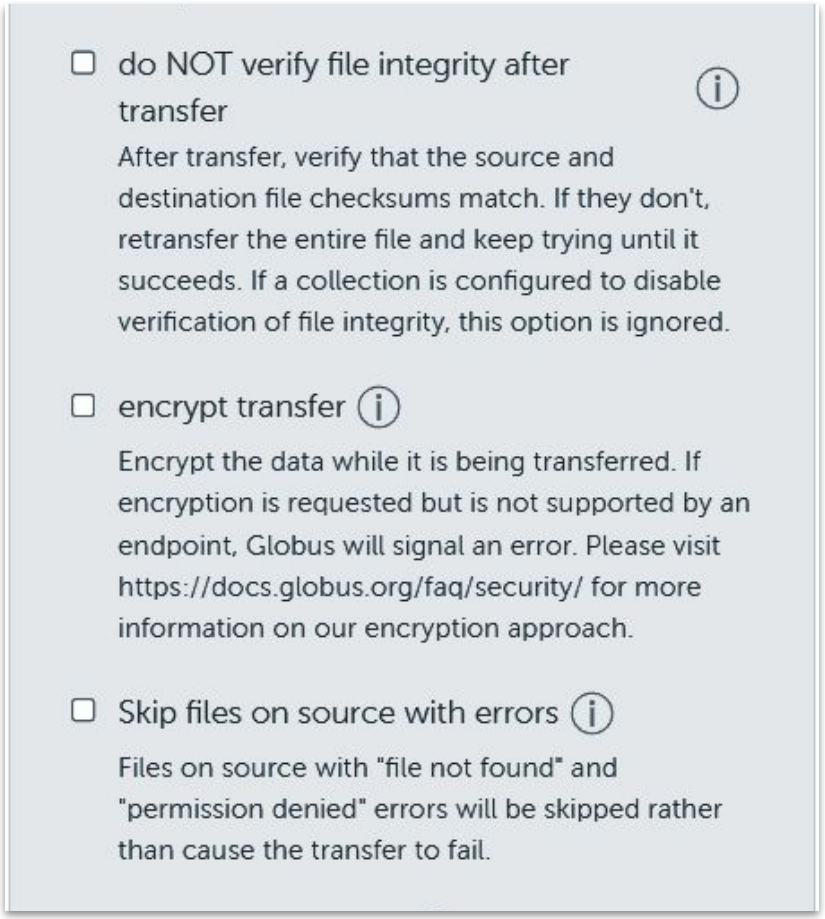
PLAIN-LANGUAGE OPTIONS!

Demo ‘A’: Quick Globus interface tour

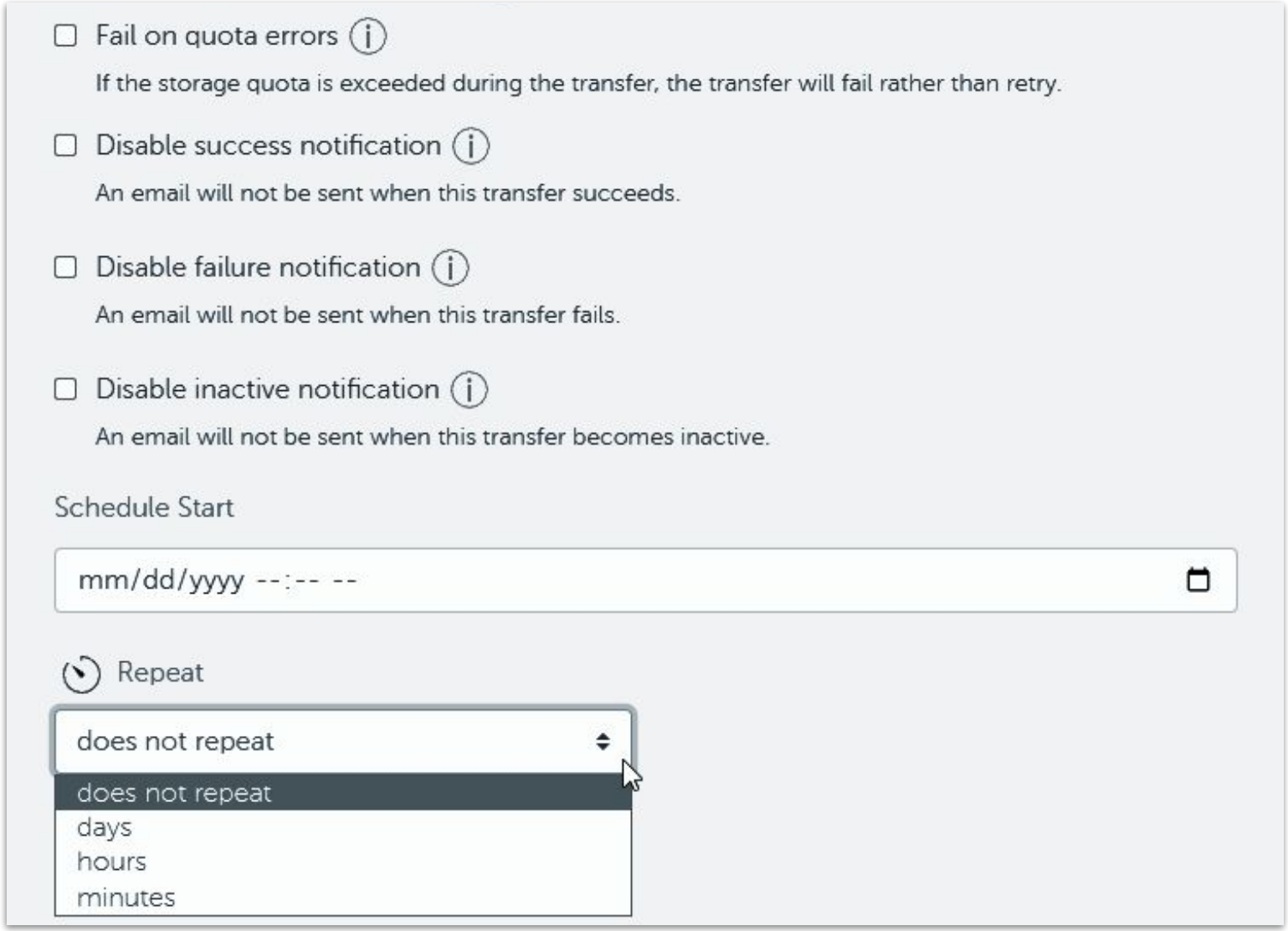


Transfer options (1/2)

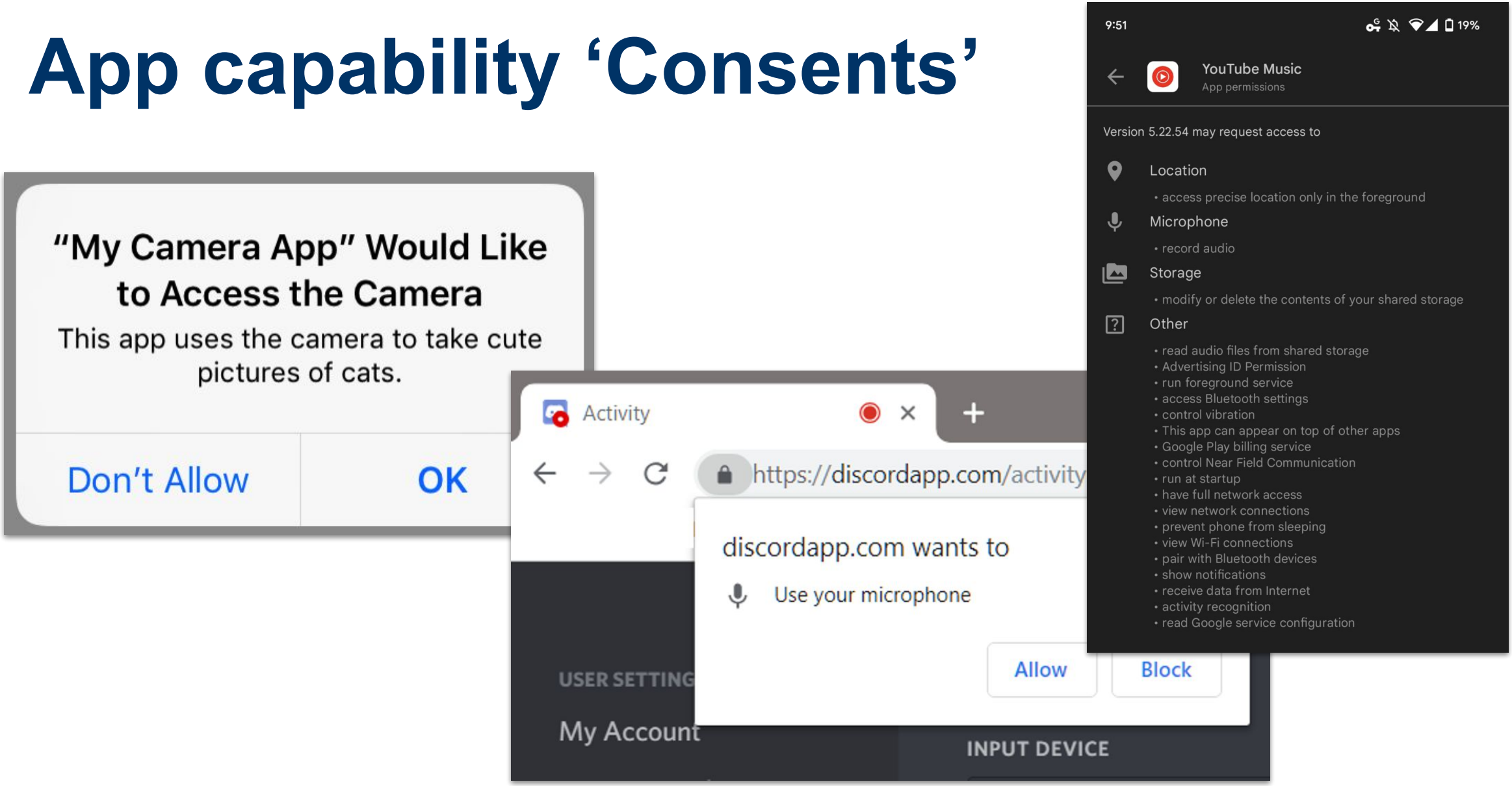




Transfer options (2/2)

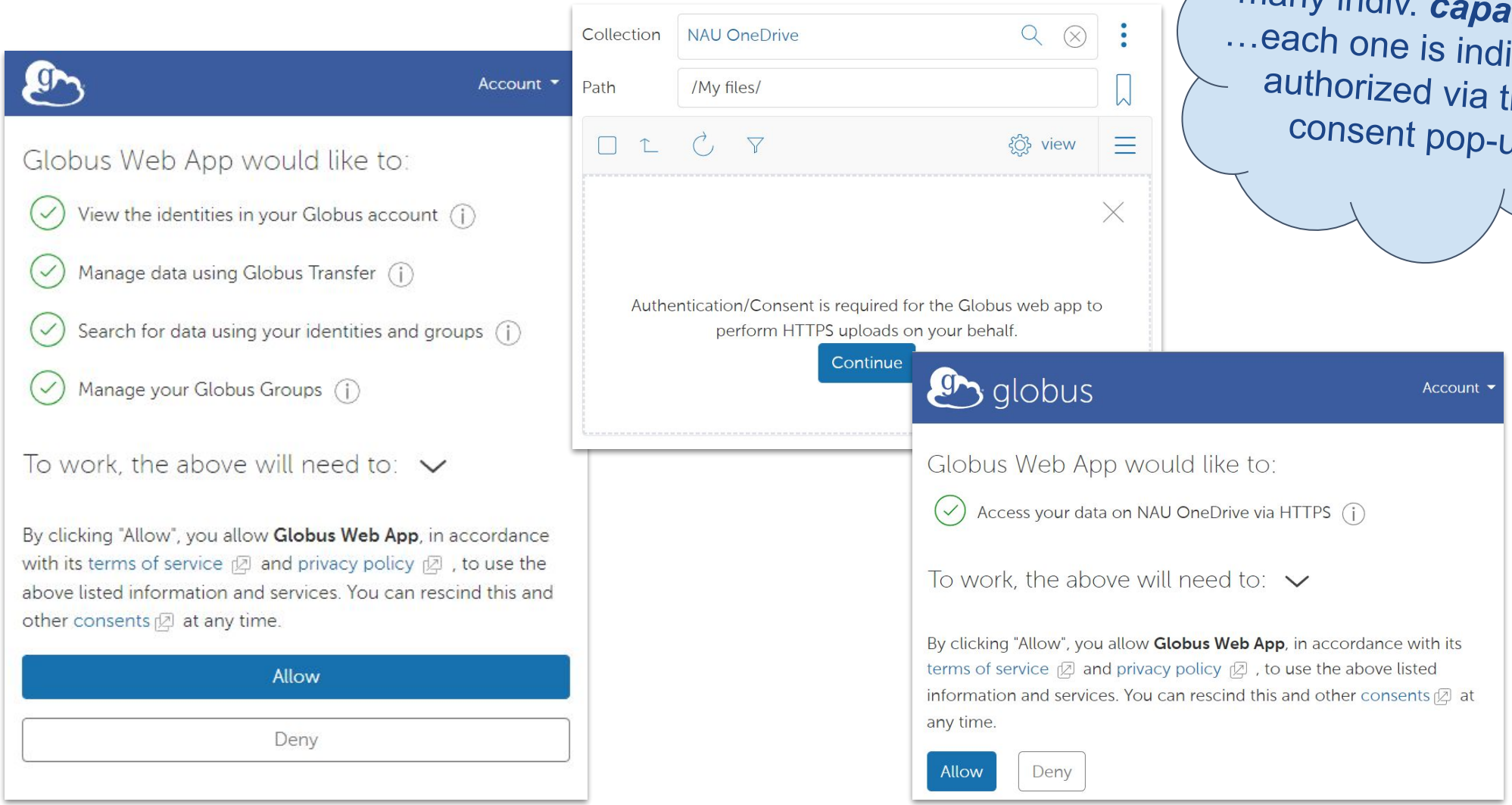


App capability ‘Consents’

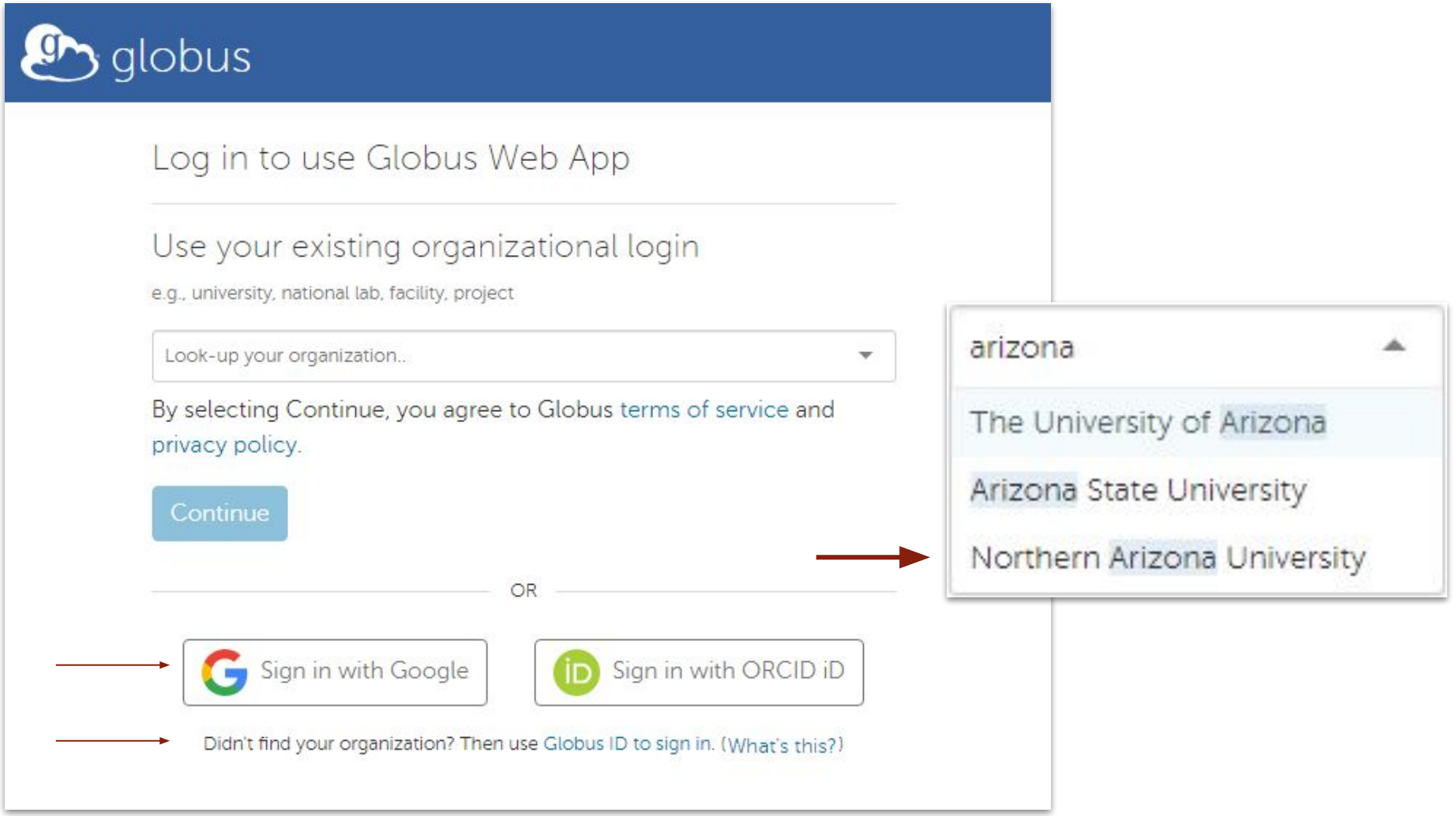


About audit ‘Consents’

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



Globus sign-in options



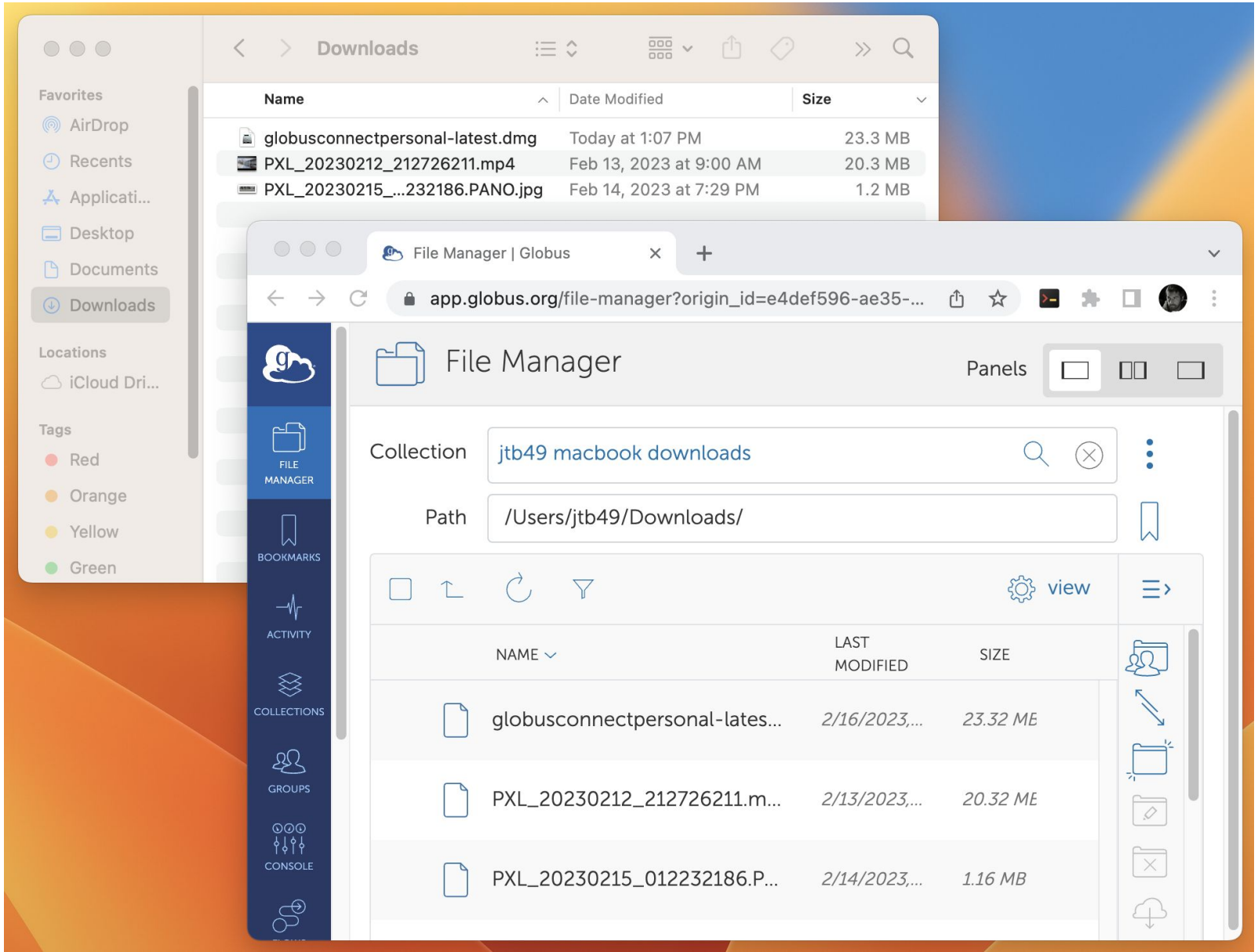
Exercise #1: login & browse your files

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

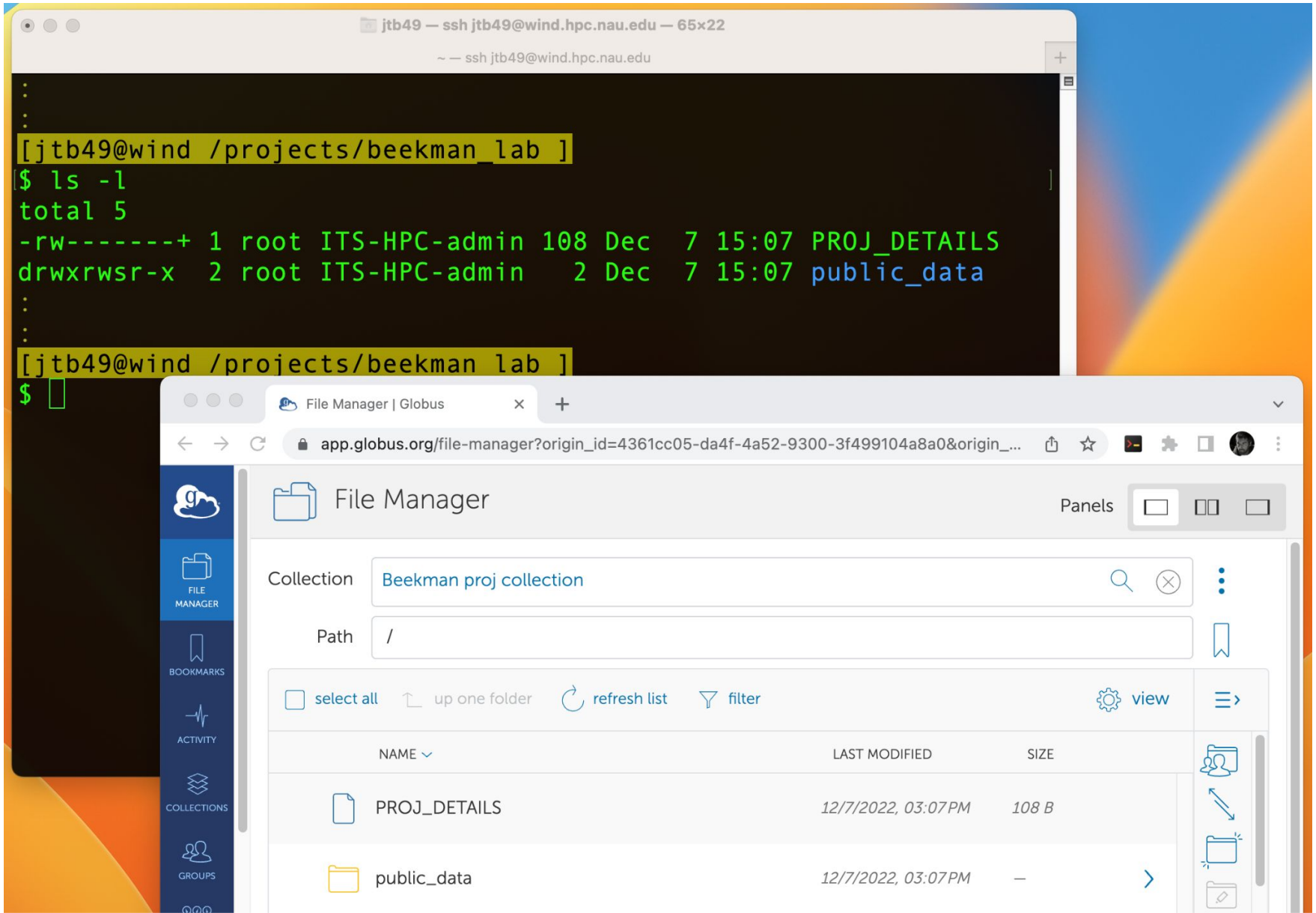
NOTE: Non-HPC users can
- Browse OneDrive & G-Drive,
- but not NAU HPC Filesystems

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

Guest collection via Globus Personal



Guest collection via mapped collection

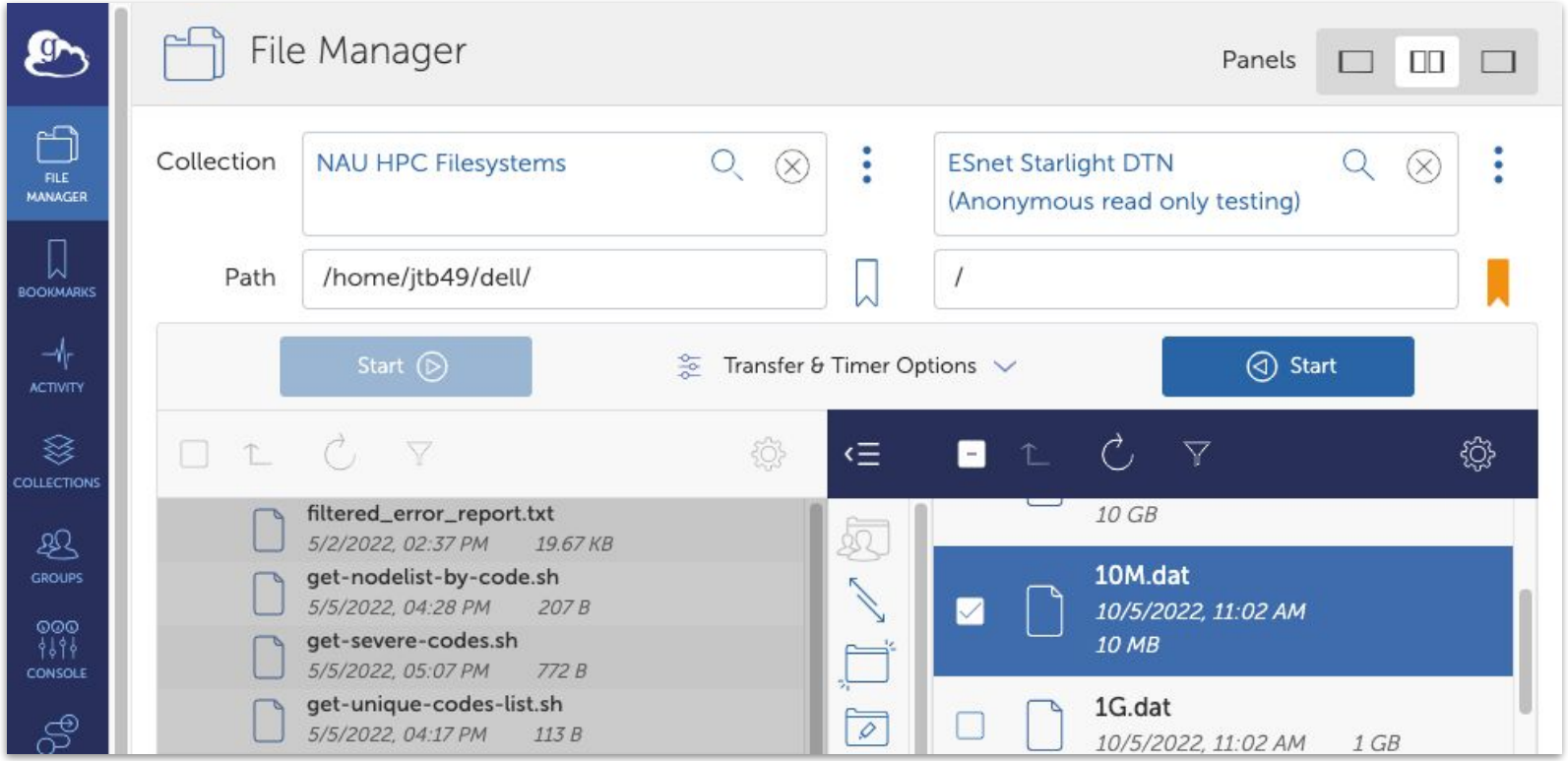


Demo ‘B’: Transfer between collections, via web

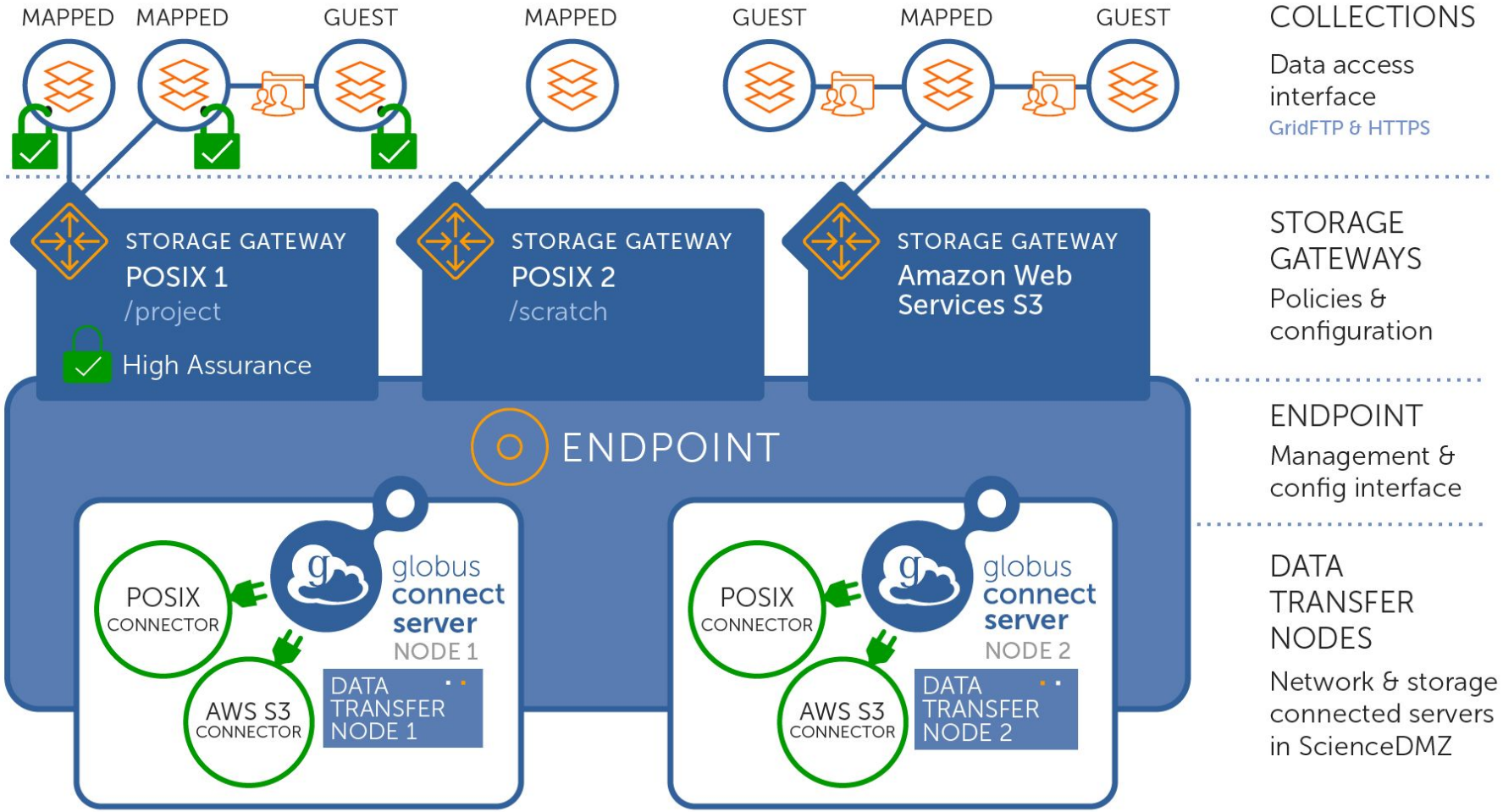
Exercise #2: copy a small file to Monsoon

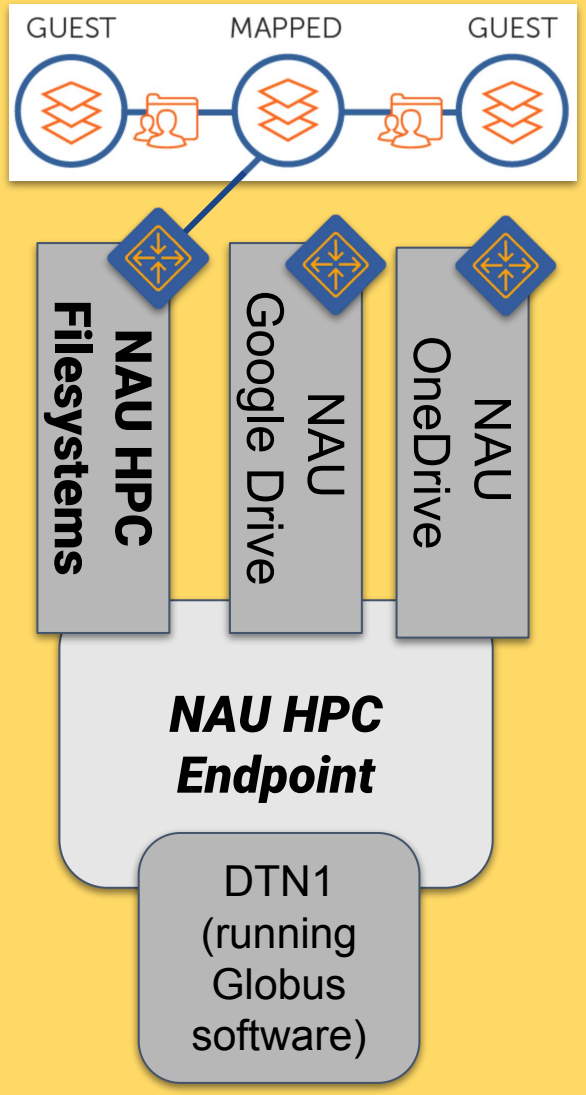
- Set two-panel view
- In the new panel:
  - Find either your G-/OneDrive **or** this **ESnet** collection →
  - Select a file
  - Click

- Check status via *sidebar* > **ACTIVITY**
- Bonus: upload a file from your desktop



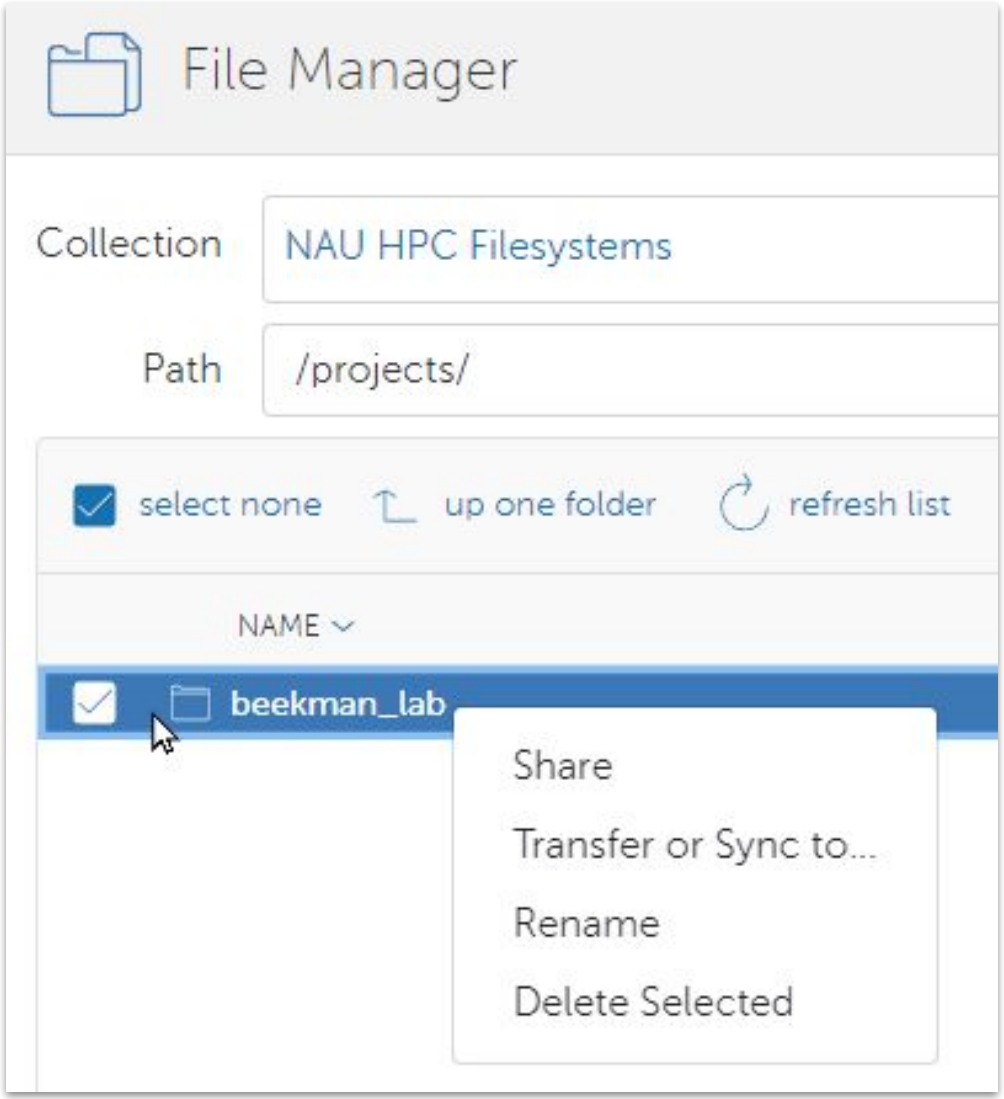
Architecture & Interface

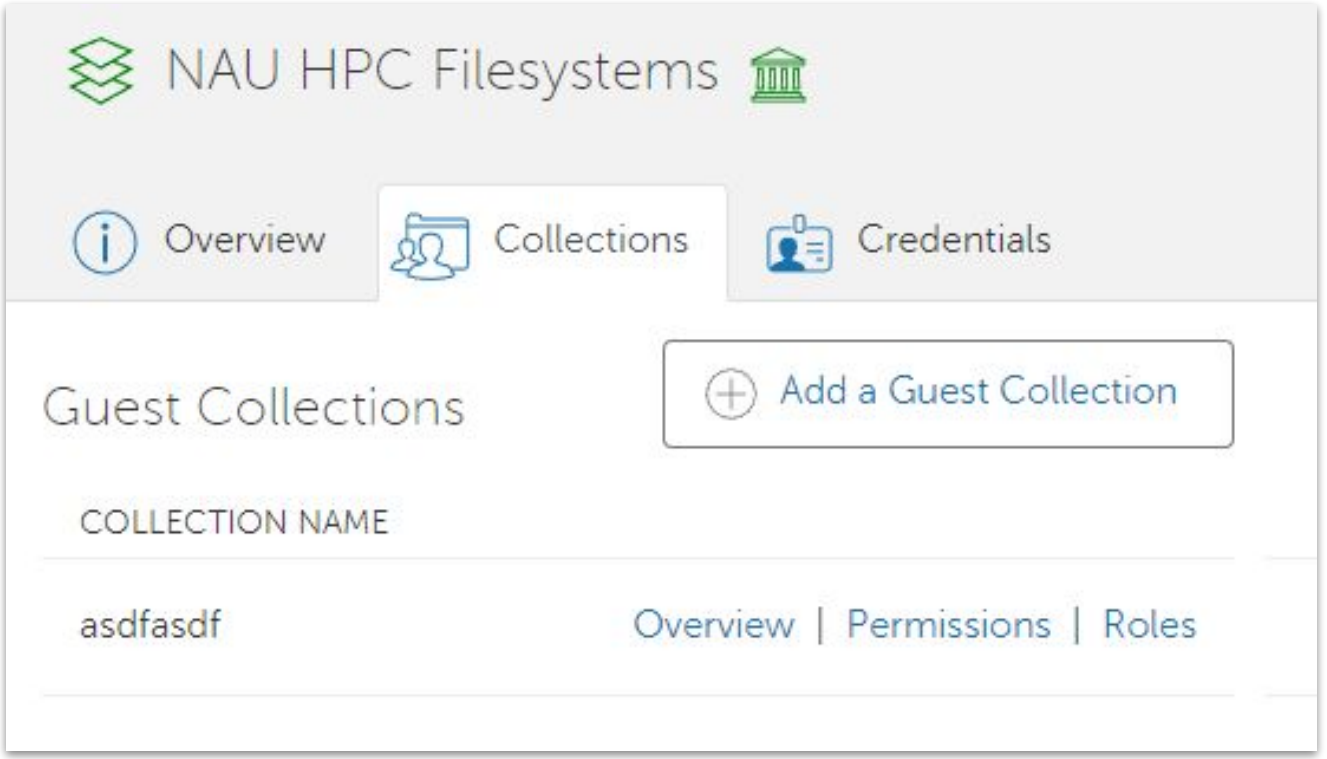




https://docs.globus.org/globus-connect-server/v5/#globus_connect_version_5_terminology

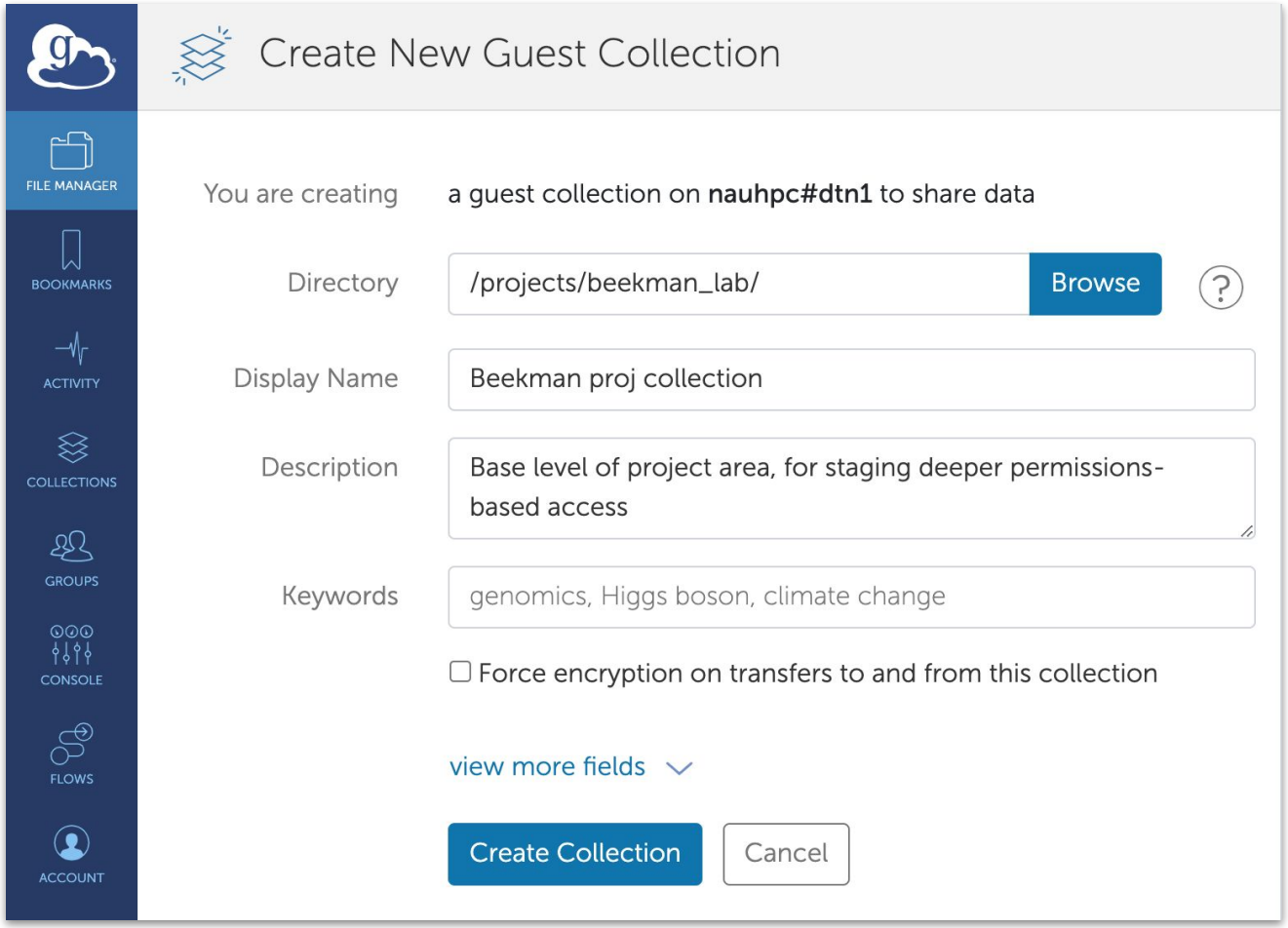
Collab/Sharing via Guest Collections



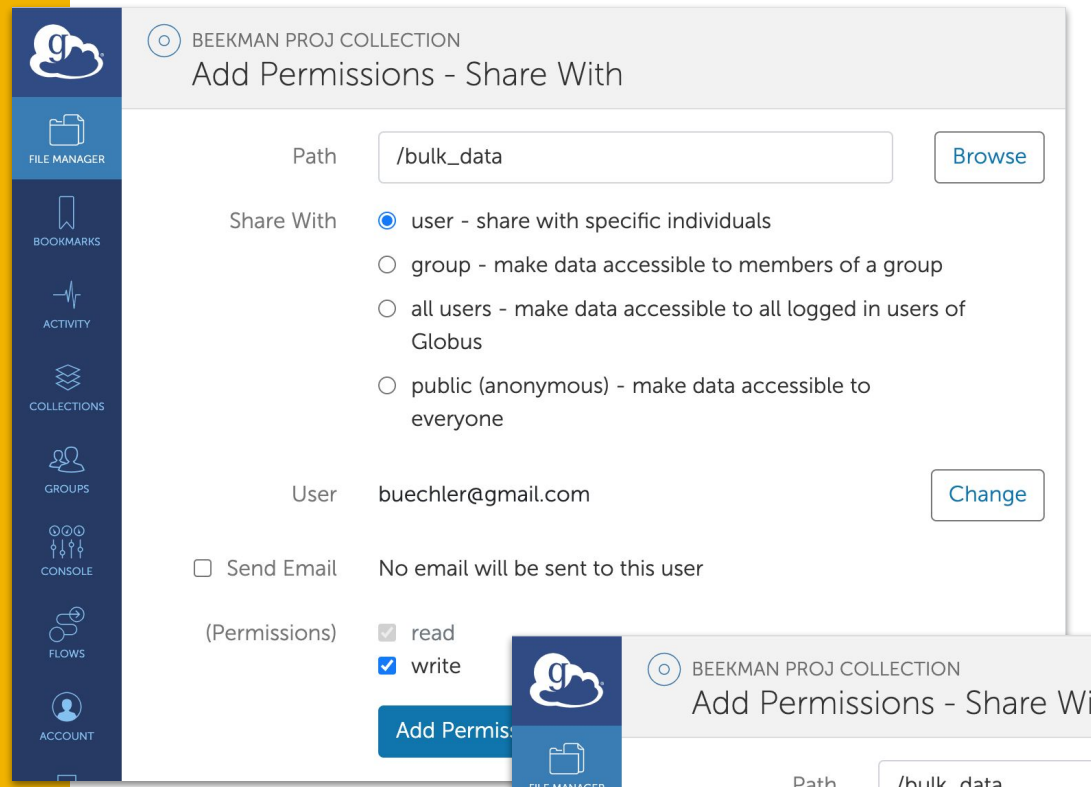


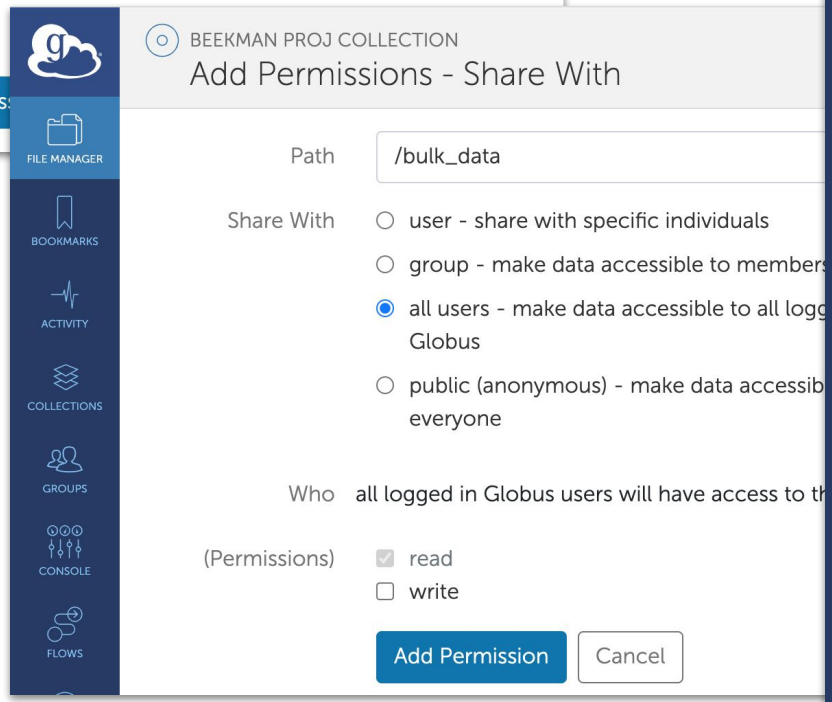
Adding a Guest Collection

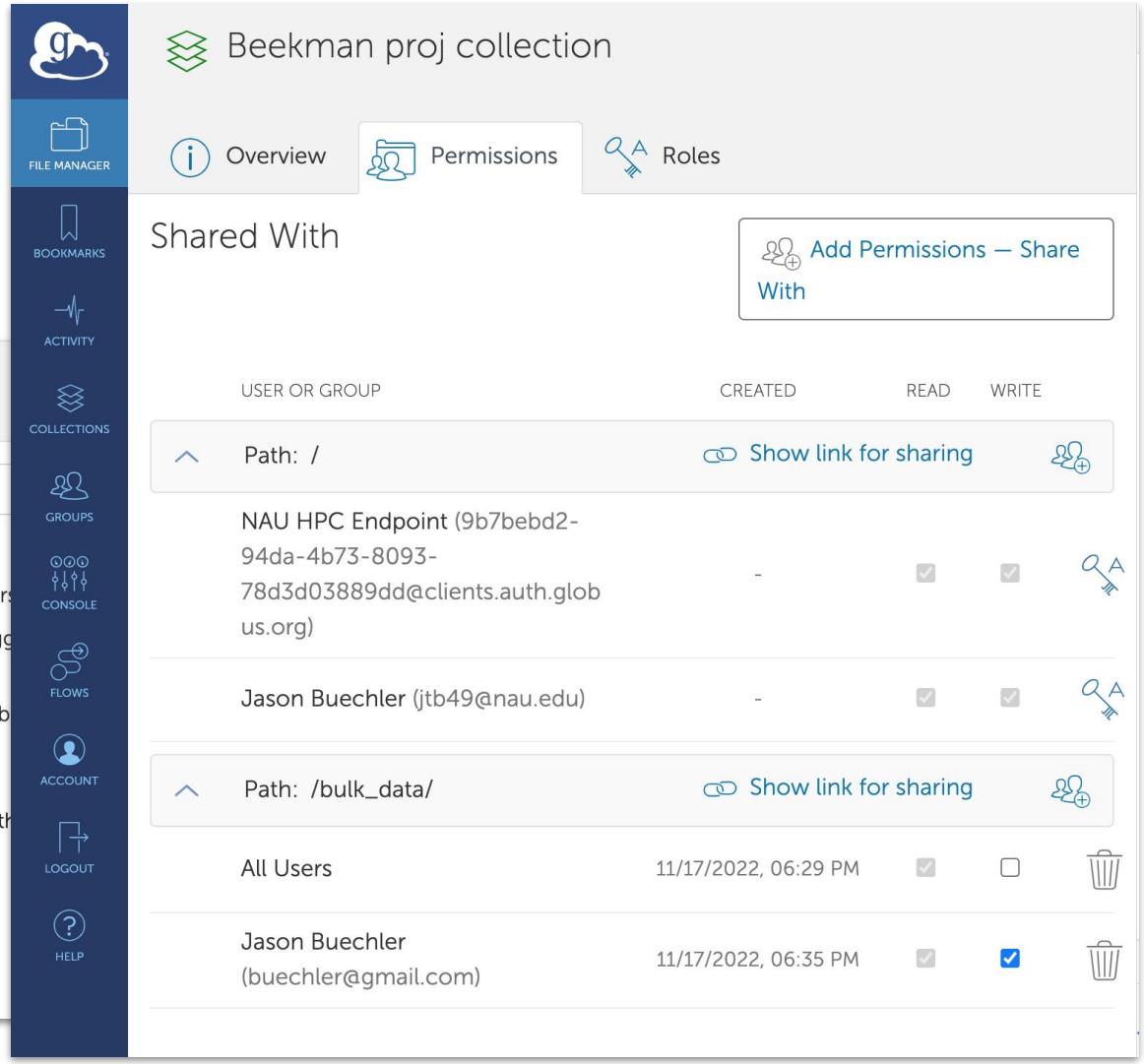
- Think of a collection like its own pre-defined filesystem
  - "/" in the file-browser equates to the chosen directory →
- **On top of** your Guest Collection, you will define Access Control Lists for files & directories
- Share with
  - Specific Globus users
  - All Globus-auth users
  - Anonymous/web



Friendliest UI in the game!





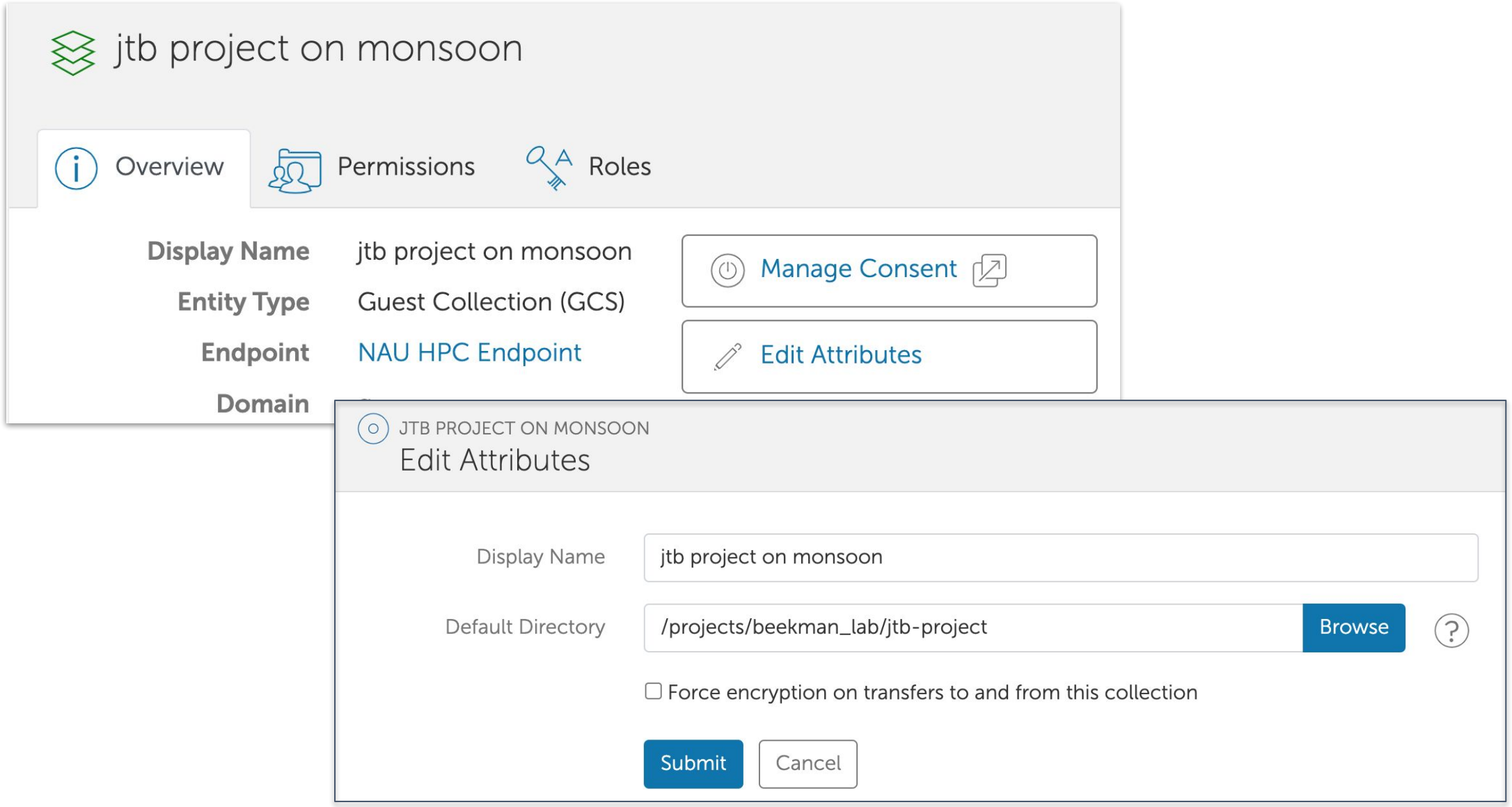


Demo ‘C’: setup a shared area

Exercise #3: create/share a Guest Collection

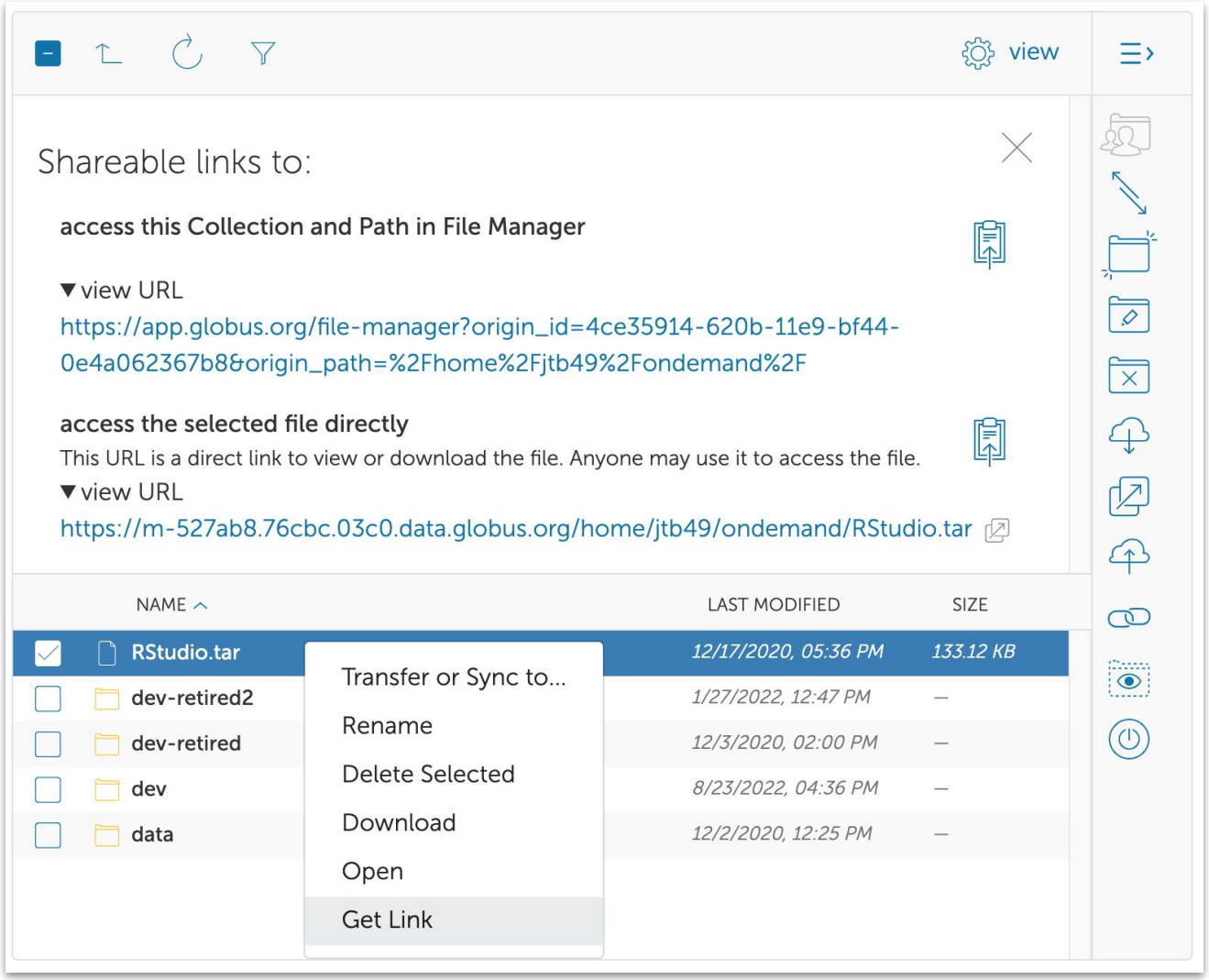
- Create a new directory in your home (e.g. ~/jb_collection)
- “Share” it, via side-toolbar or right-click
- Click “Add a guest collection” and fill-out required fields
- Click “Create collection”
  - Note you (and NAU’s “endpoint”) have default access
- Click “Add Permissions – Share With”
  - Give 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

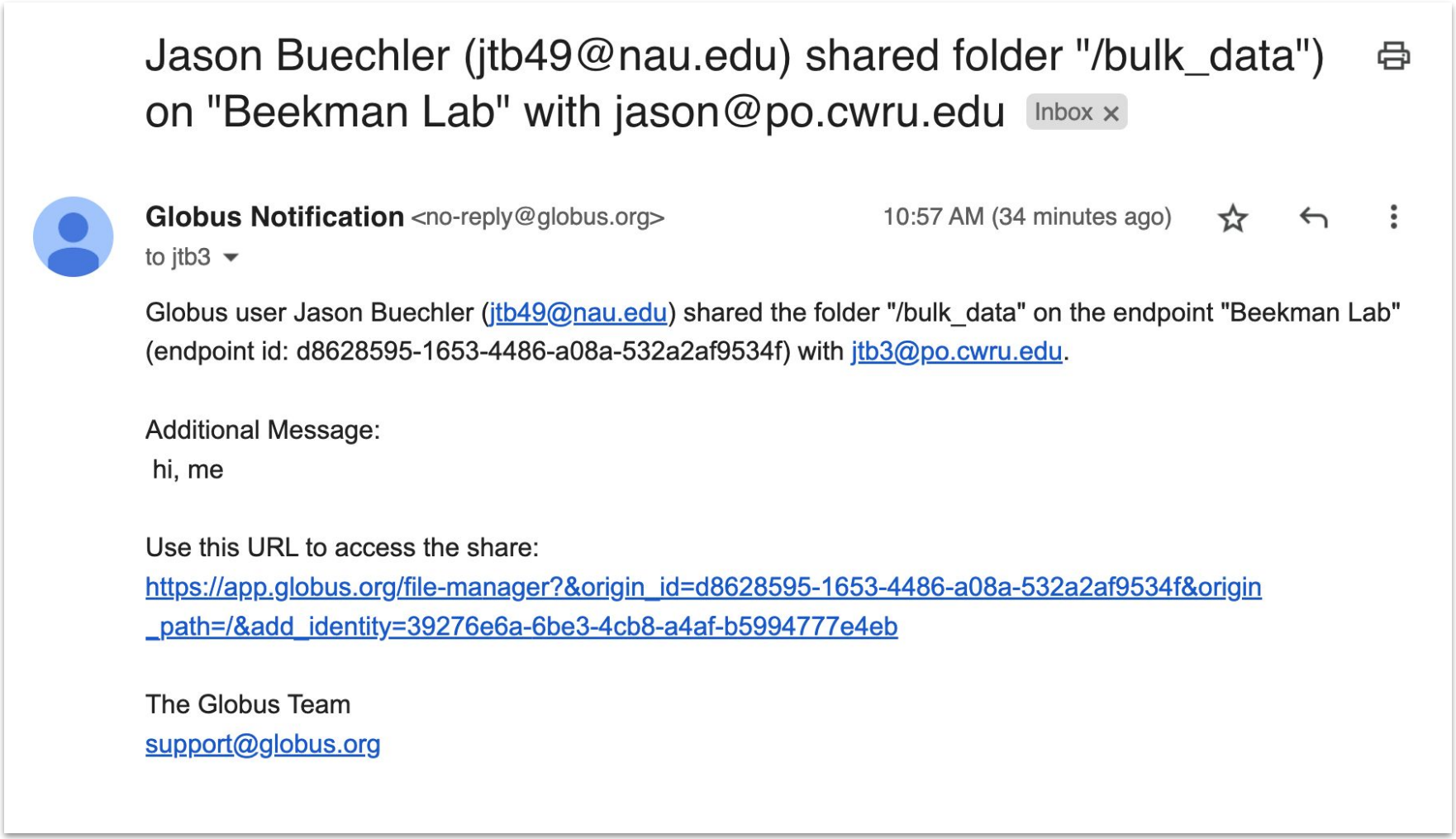


Sharing data (via link)

- 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



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 ×

Globus Notification <no-reply@globus.org> 10:57 AM (34 minutes ago)

to jtb3

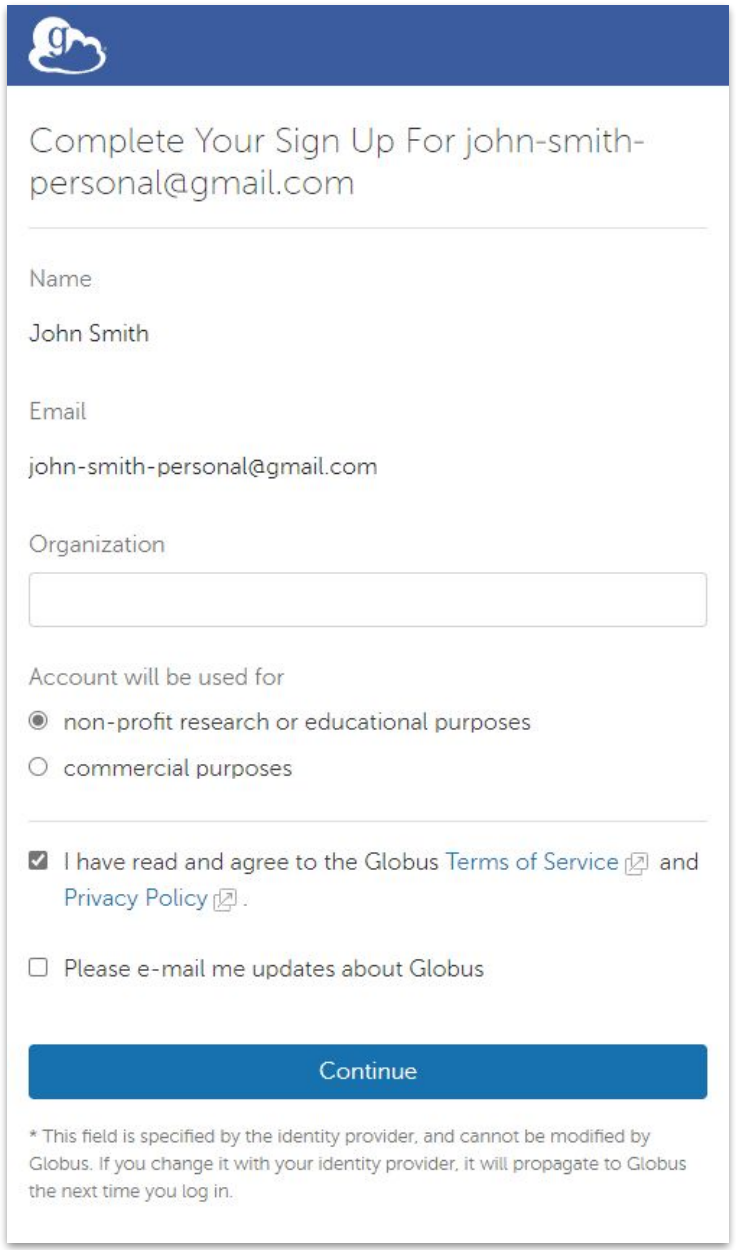
Globus user Jason Buechler (jtb49@nau.edu) shared the folder "/bulk_data" on the endpoint "Beekman Lab" (endpoint id: d8628595-1653-4486-a08a-532a2af9534f) with jtb3@po.cwru.edu.

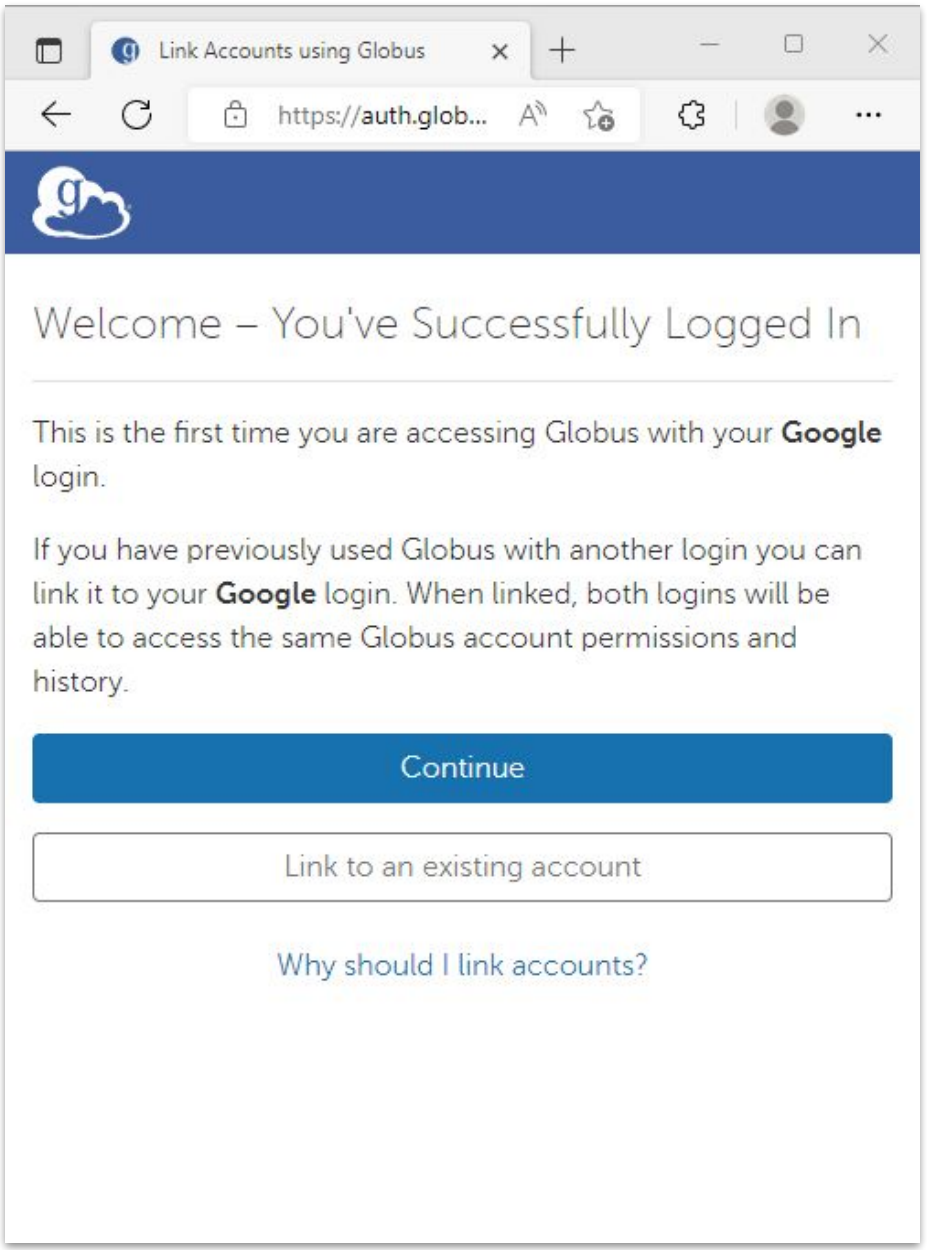
Additional Message:
hi, me

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

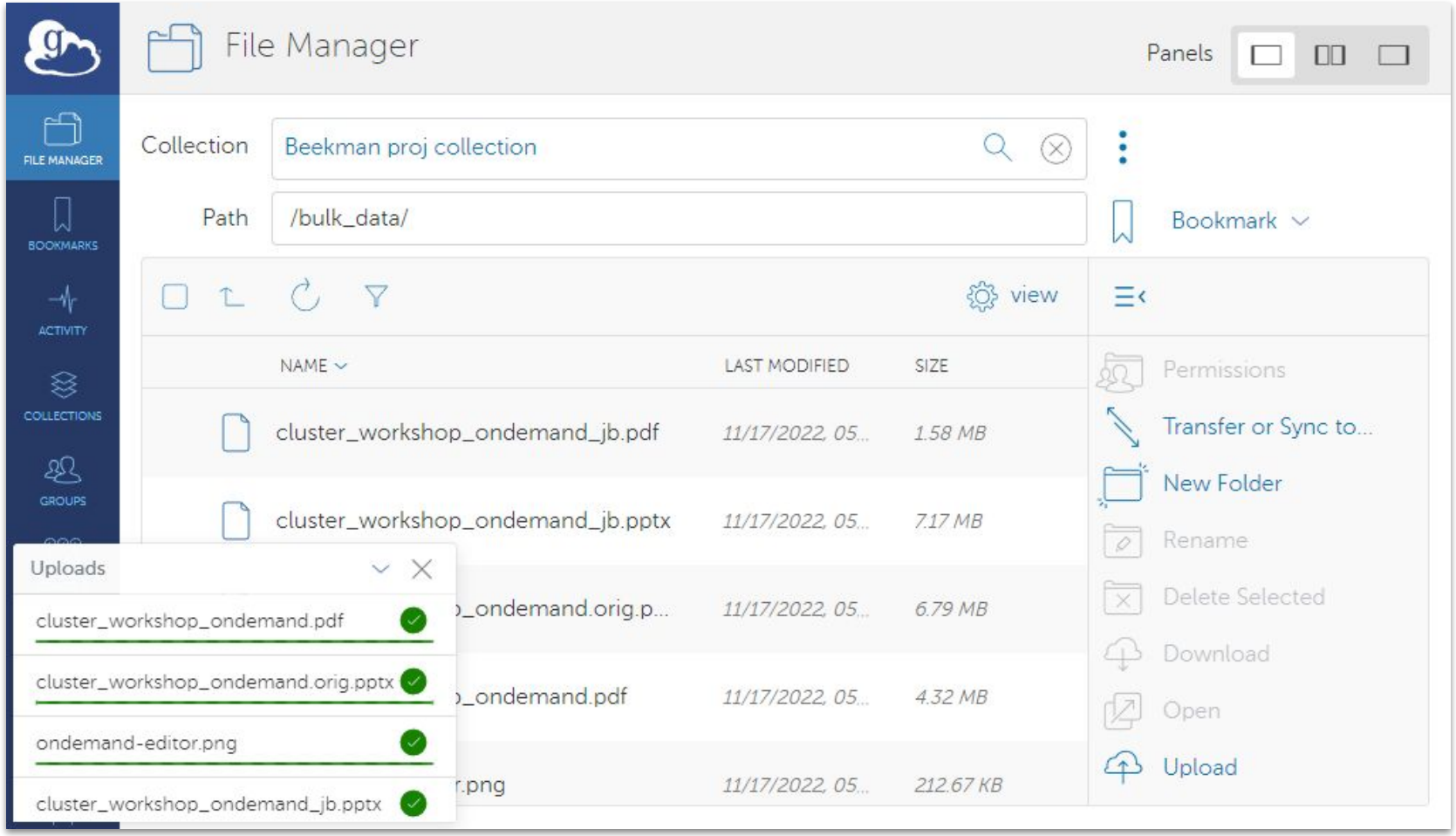
The Globus Team
support@globus.org

Non-NAU login with Globus

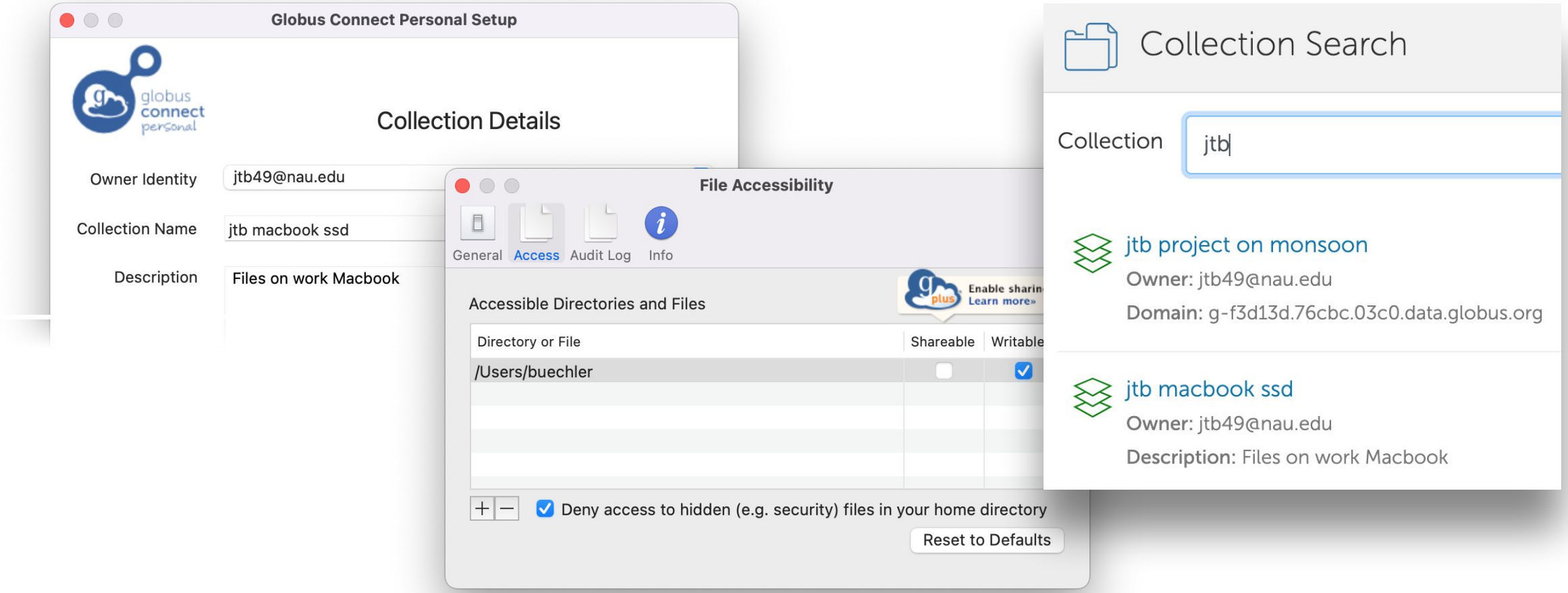




POV: A collaborator uploads to Monsoon via (just) web browser



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: uPAbsSuBw1QhXjivEOPLo143OBBHeS

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: YSiSa2UXnWOeuFzMxt0l8TPha9z7s9
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,"paral…
```

Thank You