Petalibrary

Research Computing
UNIVERSITY OF COLORADO **BOULDER**

Be Boulder.

Petalibrary

- Daniel Trahan
- *Email: Daniel.Trahan@Colorado.edu*
- *RC Homepage: https://www.colorado.edu/rc*

Sign in! http://tinyurl.com/curc-names

- Slides available for download at: https://github.com/ResearchComputing/Petalibrary_Spring_2021

Outline

- What is Petalibrary
- Data Transfers
- Sharing Data
- Computing with Petalibrary
- Organizing Data
- Getting A Petalibrary Allocation

What is Petalibrary?

- Research Computing offers a **paid**, long-term storage solution closely coupled with RC resources.
- Petalibrary
  - Large scale subsidized storage solution
  - Enterprise Grade
  - RC Staff supported with assistance on transfer strategies
  - Available in several flavors:
    - Active – Disk
    - Archival – Tape
    - Active Storage with Archive copy

Hardware Specifications

- Active Storage
  - Spinning disk platters for frequent reads and writes
  - BeeGFS filesystem
  - Parallel file I/O capable
  - RAID-6 file protection
  - Allocations located at: `/pl/active/`
- Archive Storage
  - Tape storage for infrequent reads and writes
  - iRods backed with StrongBox
  - Redundant copies of Data on separate tapes
  - Allocations located at: `/pl/archive/`

Accessing your files

- Petalibrary Allocations are all located in the following paths:
- Active:

```
[userXXXX@login12 ~]$ cd /pl/active/<your-allocation>
```

- Archive:

```
[userXXXX@login12 ~]$ cd /pl/archive/<your-allocation>
```

Why Petalibrary?

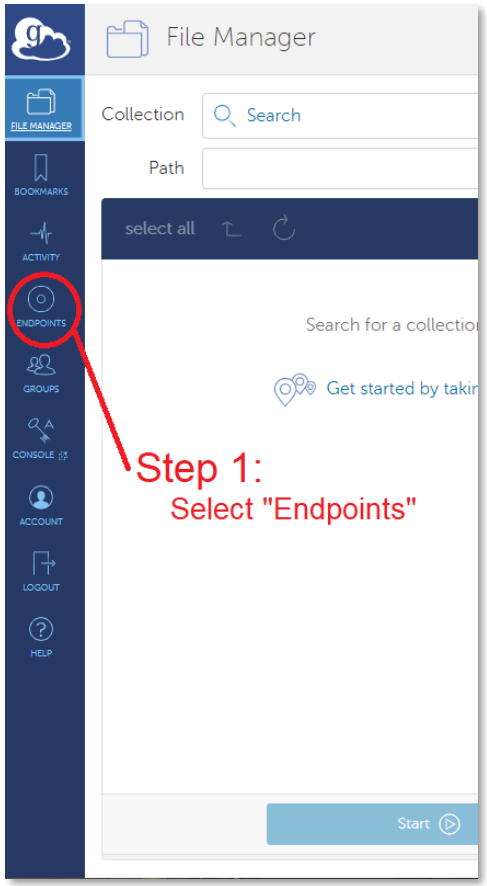
- Most users do not need a substantial sum of storage for their computational needs.
- Consider if…
  - You exceed your 10TB scratch space and need long term storage
  - You need long term managed archival services for your data
  - You own a Blanca node and need a place to compute against
  - You wish to have a group shared storage space that isn't located on a user's project space.
  - You need cost effective enterprise grade storage at a subsidized price

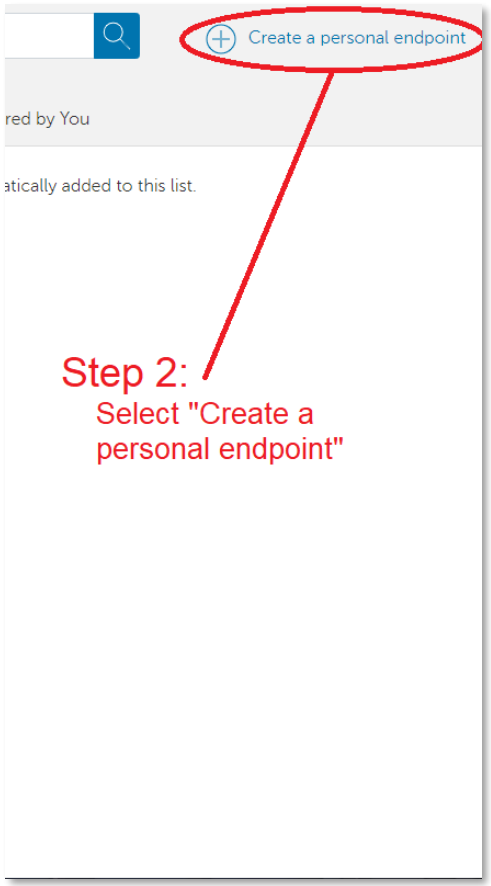
Data Transfers to Petalibrary

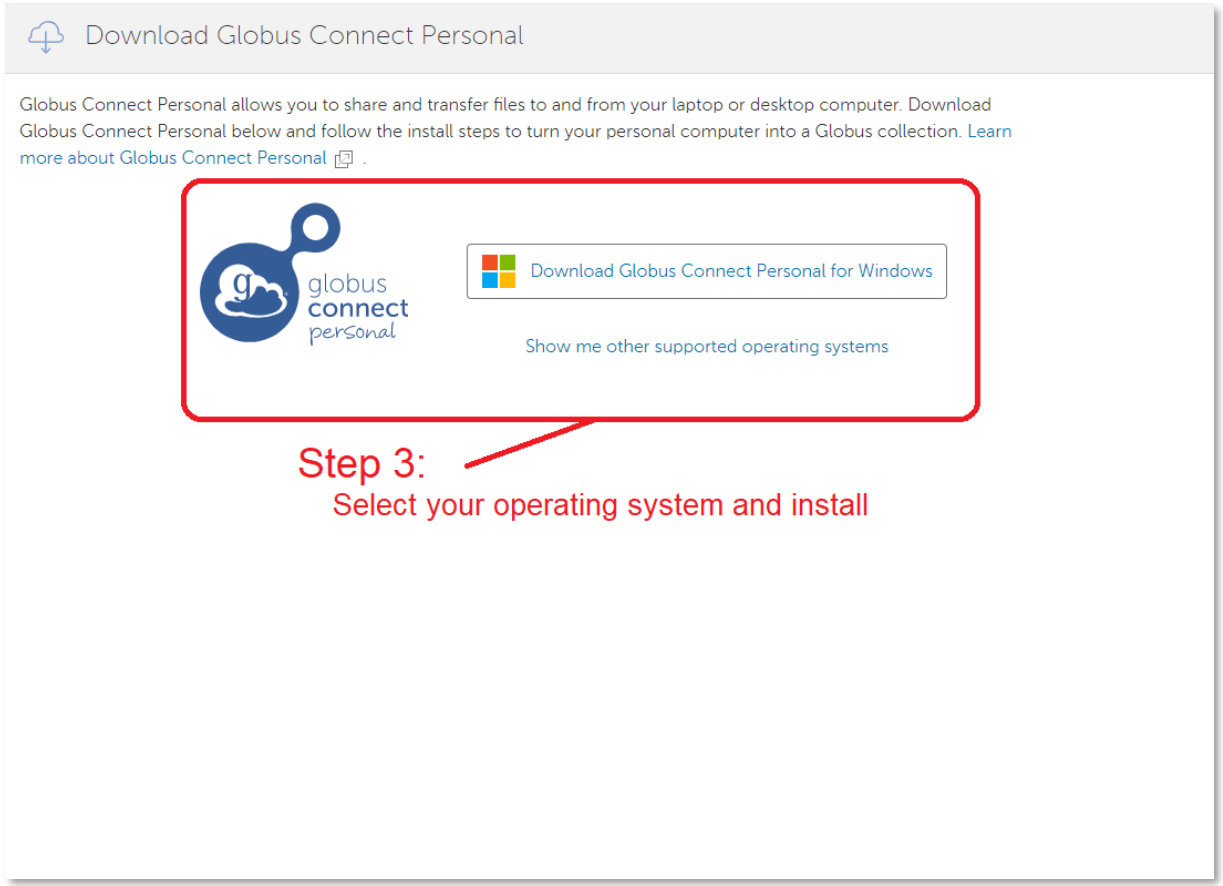
- Data transfers to Petalibrary are done like any other directory on RC resources.
- Globus
  - By far the most stable and recommended way for data transfers
  - Fast transfers
  - Transfers continue if a user disconnects
  - Web GUI option or Globus Connect Personal
- SCP/SFTP
  - Secure Copy and Secure File Transfer Protocol
  - Straightforward method of transferring data
  - Generally, recommend only to move small files less than a Gigabyte.

Globus Demo (1)

- Globus login is simple and quick: https://app.globus.org
  1. Select University of Colorado at Boulder under the dropdown menu
  2. Login with your CU credentials
  3. Continue with onscreen prompts until you are brought to the Globus WebGUI
- Installing a Globus Endpoint on your local machine
  1. Navigate down to Endpoints on the sidebar
  2. Click create an endpoint on the top right of the page
  3. Select your operating system and download the installer
  4. Follow the prompts on the installer and complete the installation

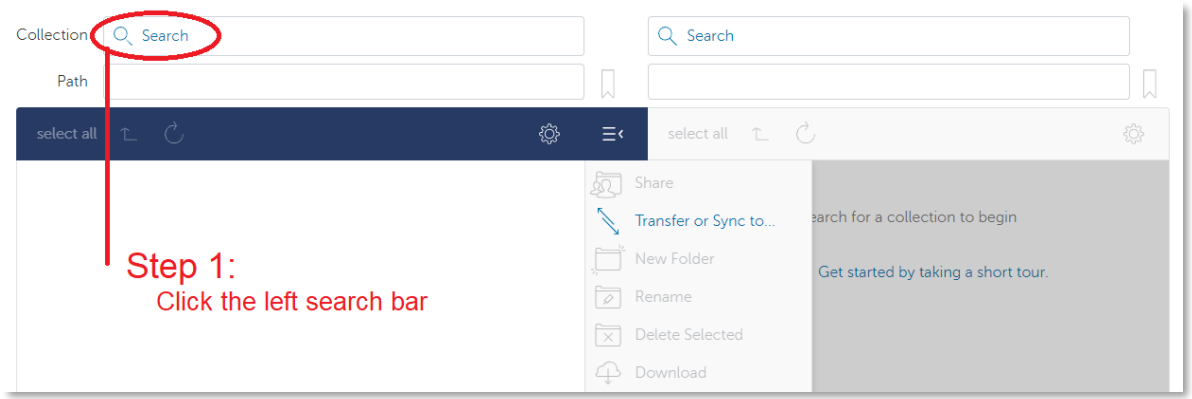

File Manager
FILE MANAGER
BOOKMARKS
ACTIVITY
ENDPOINTS
GROUPS
CONSOLE
ACCOUNT
LOGOUT
HELP
Collection
Search
Path
select all
Search for a collectio
Get started by takir
Start
Step 1:
Select "Endpoints"


Create a personal endpoint
red by You
atically added to this list.
Step 2:
Select "Create a
personal endpoint"

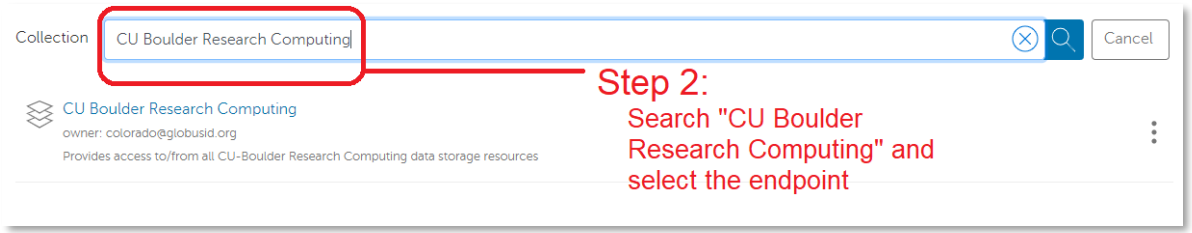

Download Globus Connect Personal
Globus Connect Personal allows you to share and transfer files to and from your laptop or desktop computer. Download Globus Connect Personal below and follow the install steps to turn your personal computer into a Globus collection. Learn more about Globus Connect Personal .
globus
connect
personal
Download Globus Connect Personal for Windows
Show me other supported operating systems
Step 3:
Select your operating system and install

Globus Demo (2)

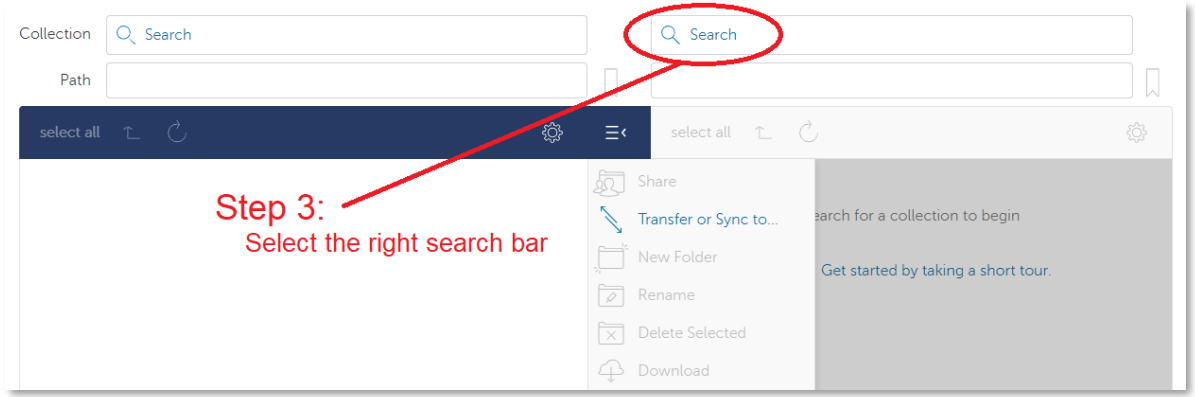
- Transferring Files can be done through the GUI
- From the File Manager tab:
  1. Click the “Two Panel” view button at the top right.
  2. Click the top left Search bar.
  3. Search “CU Boulder Research Computing” and select the end point.
  4. Sign into Research Computing’s Endpoint
  5. Click the right search bar
  6. On the ‘Your Collections’ tab, choose the endpoint you created
  7. Transfer your files!



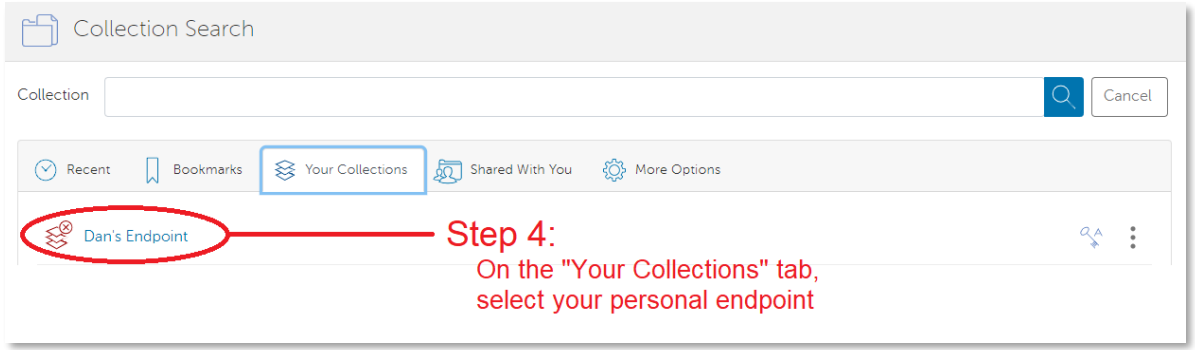
Step 1:
Click the left search bar



Step 2:
Search "CU Boulder Research Computing" and select the endpoint



Step 3:
Select the right search bar



Step 4:
On the "Your Collections" tab, select your personal endpoint

More on Data transfers

- sshfs
  - Mounting the RC filesystem to your drive remotely!
  - Single sign in for multiple data transfers
  - Great when needing to repeatedly access files on RC Resources
- rsync/rclone
  - Another utility to transfer files
  - Particularly useful in repeated file transfers and synchronization of file sets
  - Snapshot like backups
  - https://github.com/ResearchComputing/Documentation/blob/dev/docs/compute/rclone.md

Sharing Data

- Other RC Users
  - To share a Petalibrary space with other RC users. Simply contact RC with a list of users you would wish to allow access.
  - RC will place the chosen users in the owner's group
  - The owner can then set up permissions in the space
  - On premise collaborators can also access Petalibrary files with Globus Shared Endpoints
- Off-premise collaborators
  - Off premise collaborators can only access Petalibrary files through Globus Shared Endpoints

Globus Shared Endpoints

- Globus offers ‘shared endpoints’ which don’t require a user to have an account with RC.
- RC provides this capability for easy access of Data.
- Generates a shared collection that can be accessed with a link.
  - Can assign various permissions to specific users or all users withing Globus
  - More information on here: https://docs.globus.org/how-to/share-files/

Data Publishing with Petalibrary

- Using Globus shared endpoints can be a great way to publish your data while maintaining the convenience of having it Petalibrary.
- Example: https://scholar.colorado.edu/concern/datasets/9593tw13k

Computing with Data on Petalibrary

- Petalibrary Active allocations are available for active compute
- Blanca
  - Lacks supported performant scratch space for compute
  - Petalibrary Active is parallel I/O capable
  - Petalibrary can be used as a location to have performant I/O
- Summit
  - Has a dedicated performant scratch space for compute
  - Petalibrary has comparable performance with Scratch
  - Permanent location

Checking your Petalibrary Allocation

- *curc-quota* – Research Computing tool to monitor disk usage.
  - Provides detailed summary of your core storage
  - Provides detailed summary of scratch space on compile and compute nodes
  - Also lists current capacity of all Petalibrary allocations you have access to

```
[userXXXX@login12 ~]$ curc-quota
```

Getting a Petalibrary Allocation

- Any user affiliated with Research Computing and CU Boulder has access to Purchase a Petalibrary allocation
- Contact us at: rc-help@colorado.edu to get started!
- Petalibrary MOU
- Core Facilities receive their first x.TB free for their first year

Storage Solution	FY20 Service Fees
Active Storage	$45/TB/year
Active w/ Archive Copy	$55/TB/year
Archive Storage	$20/TB/year

Thank you!

- Please fill out the survey: http://tinyurl.com/curc-survey18
- Sign in! http://tinyurl.com/curc-names
- Contact information: rc-help@Colorado.edu
- Slides: https://github.com/ResearchComputing/Petalibrary_Spring_2021