# Data Transfers

# Getting Started with RC


- Daniel Trahan
- *Email: Daniel.Trahan@Colorado.edu*

- *RC Homepage: https://www.colorado.edu/rc*

- Slides available for download at:
- https://github.com/ResearchComputing/Data_Transfers_Fall_2021

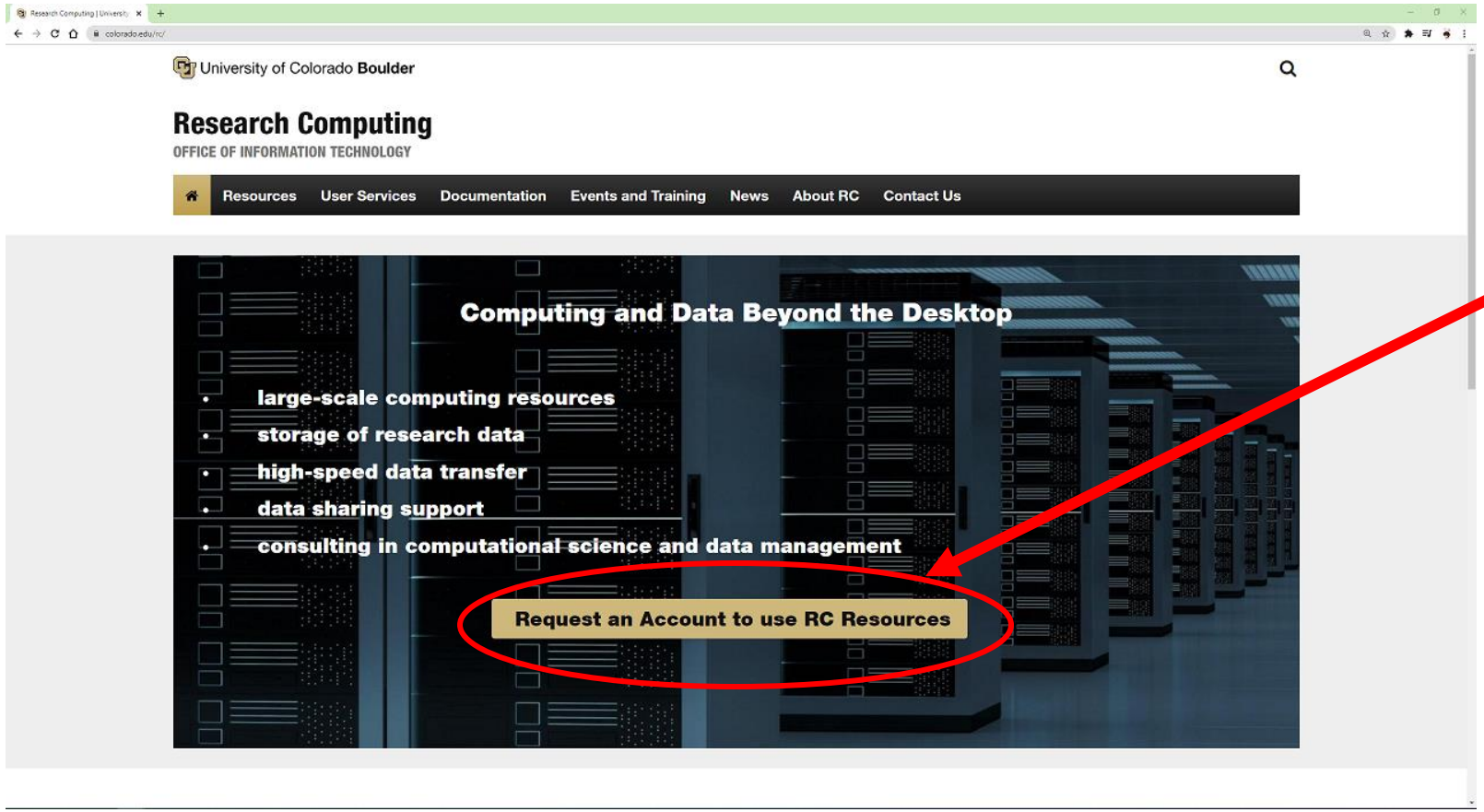
# Outline

- Getting an RC Account
- Navigating the RC Ecosystem
- Linux Tools
  - SCP
  - SFTP
  - RSYNC
  - SSHFS
  - SMB
- Globus Demo
- Sharing Data
- Getting A Petalibrary Allocation

# RC Accounts

- Before accessing your resources, Research Computing requires users to obtain specialized RC accounts.
- Why?
  - A lot of CU users do not need access to HPC or Enterprise grade storage
    - Highly valuable resource!
  - RC requires the setup of 2-factor authentication
  - Other institutions utilize RMACC Summit
- Accounts can quickly and easily be obtained through our website at [www.Colorado.edu/rc](www.Colorado.edu/rc)

# Getting an RC Account




Click here to get started!

# Duo Accounts

- After requesting your account you will soon be sent an invitation to RC's Duo 2-factor authentication utility.
  - You will be prompted to set up a smart device with the Duo application.
  - Follow the steps prompted by the email invitation to activate your device.
- Don't have a smart Device? No Problem!
  - RC also offers 2 factor authentication tokens to access your account.
  - Contact rc-help@colorado.edu for more information

# Accessing RC Services

- There are numerous ways to access RC Resources from your local machine
  - Globus (if just managing files!)
  - Command line (Most common for Jobs)
  - Jupyterlab
  - Visualization Cluster
- Globus is probably the easiest solution when accessing files your files on RC resources

# Access through the Command Line

- Other users may find the use of Globus to be overly clunky when wanting to check on their files.
- Users can log into RC servers on their local machines by opening a terminal and running the command:

```
ssh <your-username>@login.rc.colorado.edu
```

- For users given temporary accounts please login with the following command:

```
ssh <your-username>@tlogin1.rc.colorado.edu
```

# Basic Navigation Commands

- Change directories

```
cd <relative-or-full-path>
```

- List contents of a directory

```
ls <optional-path>
```

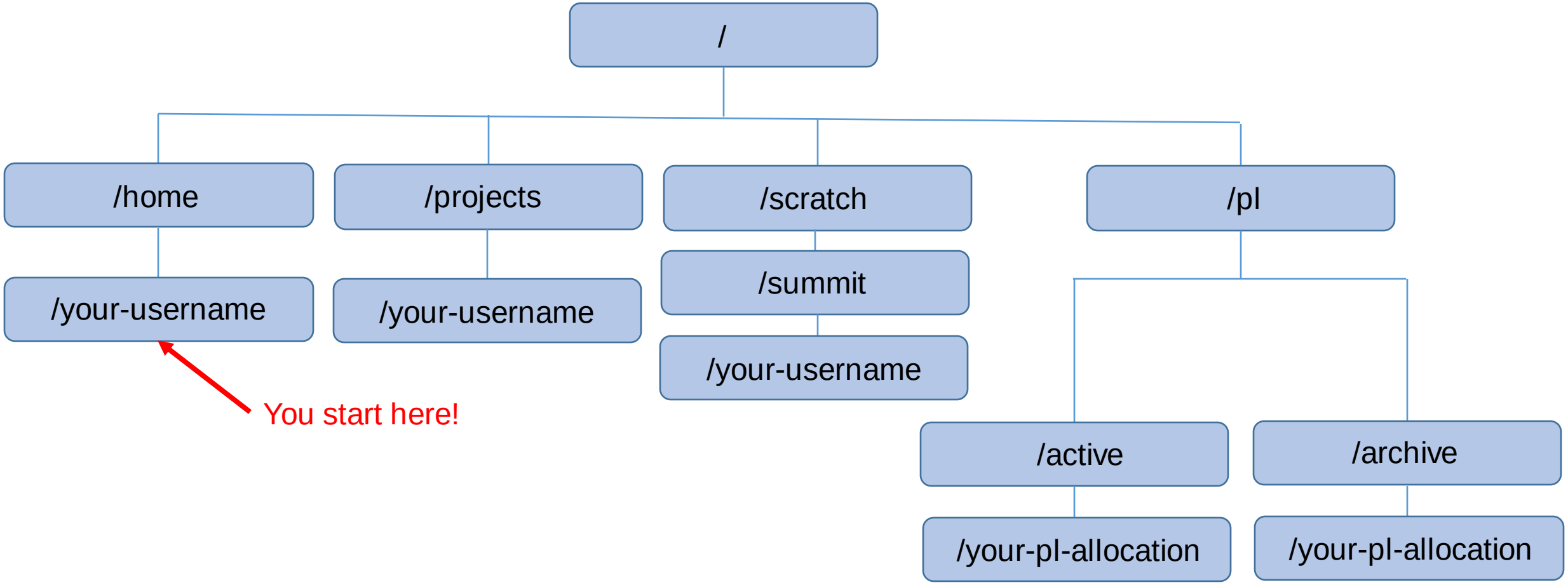
- Print current working directory

```
pwd
```

# RC Filesystem

- As with most Linux distributions, after logging into RC Resources you will be placed on your RC home directory.
- RC's file system is broken up into 4 major components
  - Home – 2GB
  - Projects – 250GB
  - Scratch – 10TB
  - PL (Petalibrary) - Contact your PI about purchasing PL Space!
- Backups occur regularly on Home and Projects.
  - PL Active capability is coming soon!
- Scratch will delete files older than 90 days.

# RC File system map



# RC Data transfer nodes

- In addition to the usual RC login points, we also provide Data transfer nodes for faster transfers.
- The upcoming commands can utilized these nodes IF you are on CU's VPN.
- Address to transfer to data transfer nodes:

```
scp file1 <username>@login.rc.colorado.edu:<remote-path>
```

# Command line options (1)

- SCP / *Secure Copy*
  - `scp` is a very simple file transfer command that allows users to transfer one file to another remote system.

```
scp file1 <username>@login.rc.colorado.edu:<remote-path>
```

  - Useful with small or quick file transfers
- SFTP / *Secure File Transfer Protocol*
  - `sftp` is a similarly simple protocol that loads a user into a sftp prompt with both local and remote file systems accessable from a single prompt.

```
sftp<username>@login.rc.colorado.edu
```

  - Useful for multiple/repetative small file transfers.

# Command line options (2)

- rsync
  - `rsync` is a popular linux utility for updating changed files to a remote filesystem.

```
rsync -v file1 <username>@login.rc.colorado.edu:<remote-path>
```

  - Useful when working on a file on both remote and local machines with modifications that need to be updated
  - Flags:

```
-v    # verbose mode
-r    # recursive (directory)
-t    # sync based off timestamp
-c    # sync changed files based on content
-a    # archive mode
```

# Command line options (3)

- sshfs
  - Mount a remote directory to a local Unix operating system!
  - Mac and Linux Exclusive:

```
sshfs <username>@login.rc.colorado.edu:<path> <local-mountpoint>
```

  - But can windows do something similar?
- smb mounting
  - Universal mounting protocol that is built into every operating system
  - Contact RC to get this set up!

# Globus

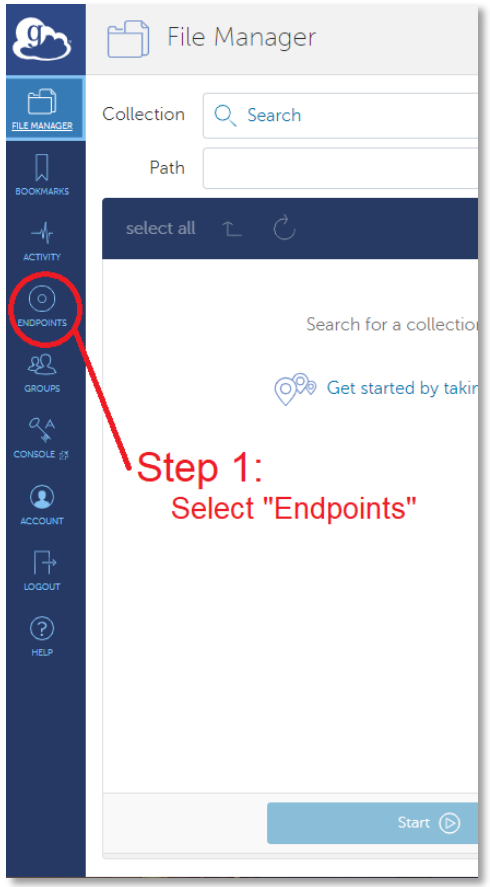
- Globus
  - By far the most stable and recommended way for data transfers
  - Fast transfers
  - Transfers continue if a user disconnects
  - Web GUI option or Globus Connect Personal
- Demo:

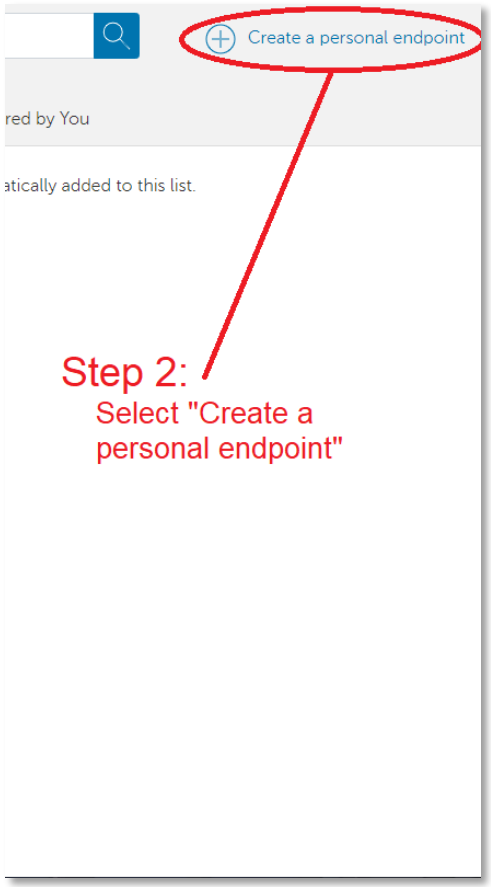
# Globus Demo (1)

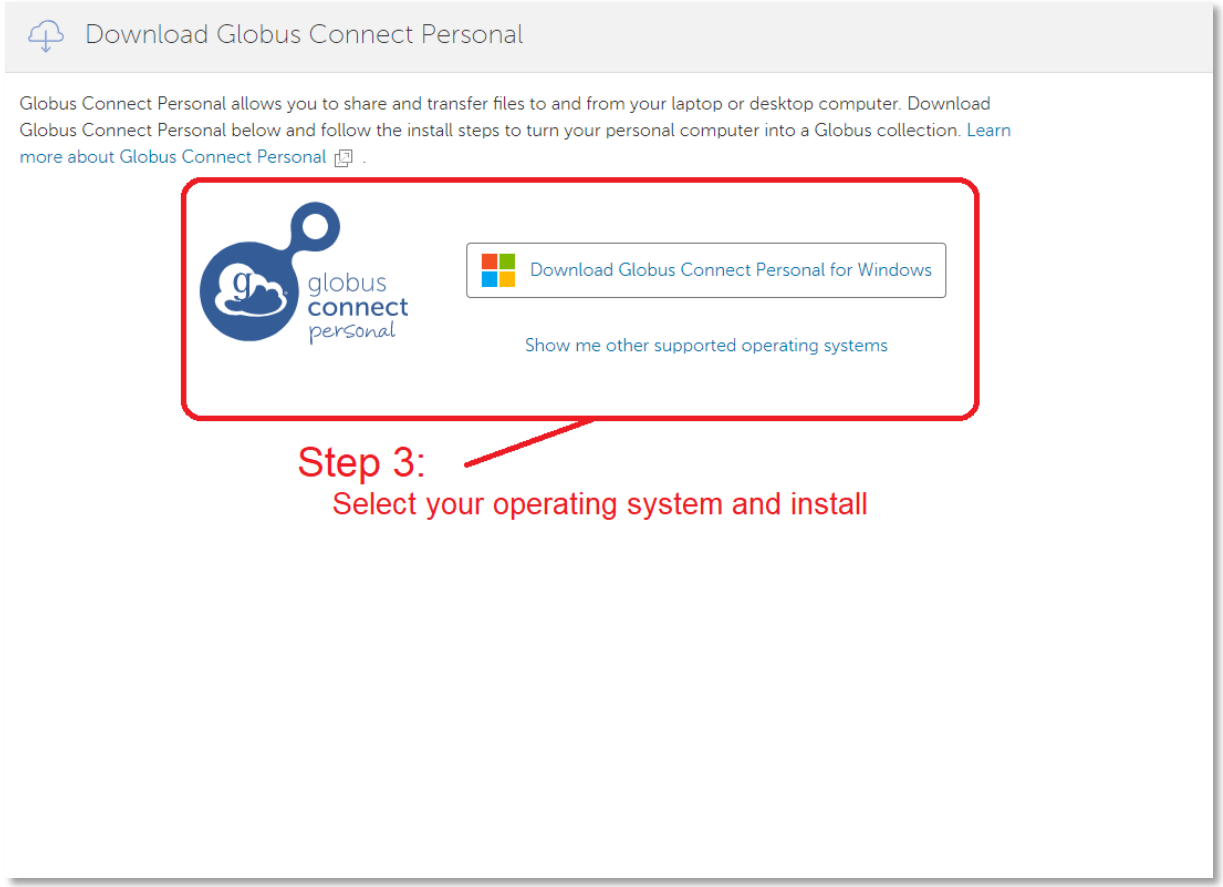
- Globus login is simple and quick: https://app.globus.org
  1. Select University of Colorado at Boulder under the dropdown menu
  2. Login with your CU credentials
  3. Continue with onscreen prompts until you are brought to the Globus WebGUI
- Installing a Globus Endpoint on your local machine
  1. 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

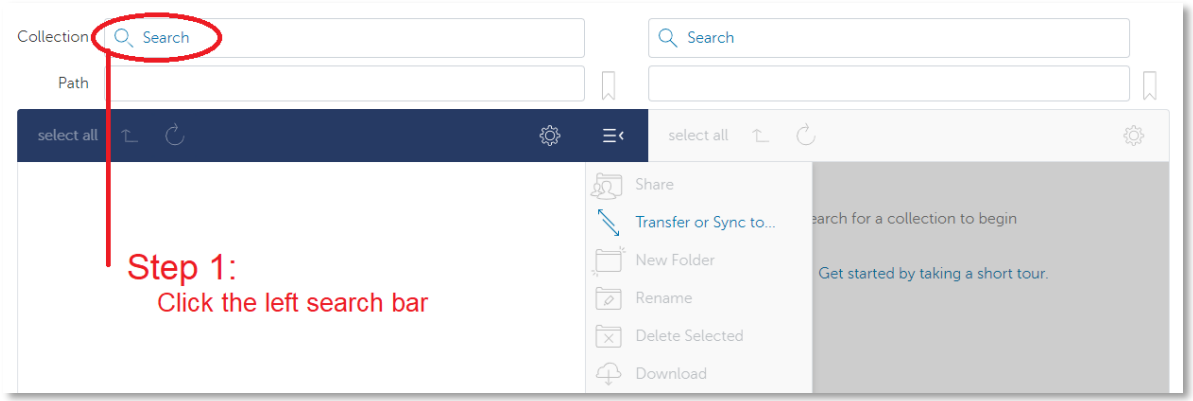
# Globus Demo (2)

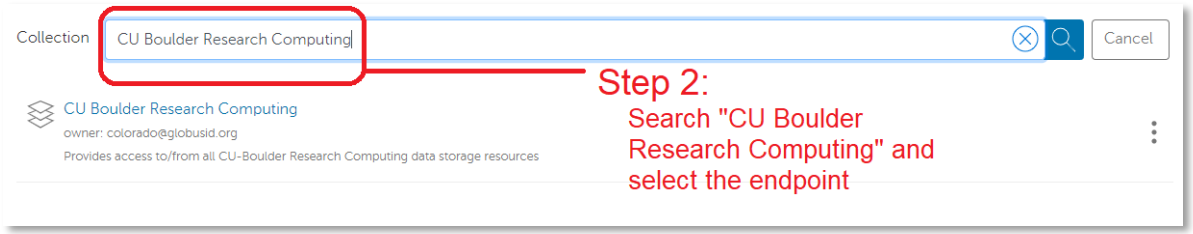
- Transferring Files can be done through the GUI
- From the File Manager tab:
  1. Click the “Two Panel” view button at the top right.
  2. Click the top left Search bar.
  3. Search “CU Boulder Research Computing” and select the end point.
  4. Sign into Research Computing’s Endpoint
  5. Click the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

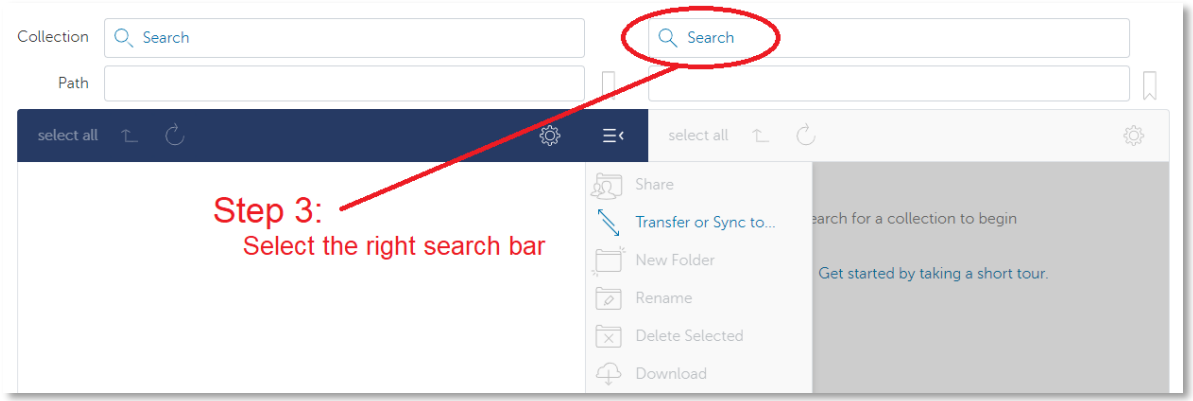

File Manager
FILE MANAGER
BOOKMARKS
ACTIVITY
ENDPOINTS
GROUPS
CONSOLE
ACCOUNT
LOGOUT
HELP
Collection
Search
Path
select all
Search for a collectio
Get started by takir
Start
Step 1:
Select "Endpoints"

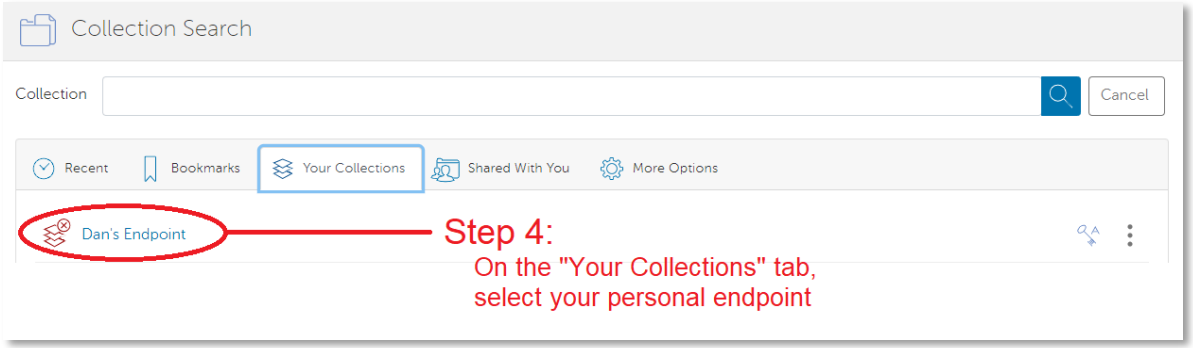

Create a personal endpoint
red by You
atically added to this list.
Step 2:
Select "Create a personal endpoint"


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


Collection
Search
Path
select all
Share
Transfer or Sync to...
New Folder
Rename
Delete Selected
Download
Get started by taking a short tour.
Step 1:
Click the left search bar


Collection
CU Boulder Research Computing
Cancel
CU Boulder Research Computing
owner: colorado@globusid.org
Provides access to/from all CU-Boulder Research Computing data storage resources
Step 2:
Search "CU Boulder Research Computing" and select the endpoint


Collection
Search
Path
select all
Share
Transfer or Sync to...
New Folder
Rename
Delete Selected
Download
Get started by taking a short tour.
Step 3:
Select the right search bar


Collection Search
Collection
Cancel
Recent
Bookmarks
Your Collections
Shared With You
More Options
Dan's Endpoint
Step 4:
On the "Your Collections" tab, select your personal endpoint

# Sharing Data

- Other RC Users
  - To share files 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

# Unix Groups

- Unix Groups
  - 3 Levels of permissions:
    - User
    - Group
    - Other
  - All users have a group associated with their username
  - Permissions can be set for an individual file with the `chmod` command

```
chmod g+rx file.exe
```

# 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.
- Petalibrary Exclusive!
- 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

# Petalibrary Notes

- *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
```

- Confidential Data is unsupported and *should not be stored on Petalibrary*

# Thank you!

- Please fill out the survey: http://tinyurl.com/curc-survey18
- Contact information: rc-help@Colorado.edu
- Slides: https://github.com/ResearchComputing/Data_Transfers_Fall_2021