`rsync (remote sync)` 🔄 📁

- Directories with many files.
- Synchronizing local and remote folders.
- Resuming interrupted large transfers.
- Filtering what gets copied (e.g., exclude logs).

```
rsync [options] [source] [destination]
```

`scp (secure copy)` 🔑🔒

- Single files, small in size (<~1GB).
- Quick, one-off transfers.
- When simplicity is the priority:

```
scp [source_path] [user@host]:[destination_path]
```

Examples 💡

1. Pushing data from your local machine to a remote server:

```
rsync -avh --progress /path/to/local/source/ username@remote.host:/path/to/remote/destination/
```

2. Pulling data from a remote server to your local machine

```
rsync -avh --progress username@remote.host:/path/to/remote/source/ /path/to/local/destination/
```

3. Securely copy a file from Scholar to your local machine

```
scp username@scolar.rcac.purdue.edu:~/results.txt .
```

Note 📝

scp always copies the entire file and does not resume if interrupted. For larger files, directories, or unreliable connections, use rsync or Globus.

Key options (rsync) ⚙

- `-a` (archive): A master flag that preserves permissions, ownership, timestamps, and symbolic links. It's equivalent to `-rlptgoD`. This is the most important option.
- `-v` (verbose): Shows a list of the files being transferred.
- `-h` (human-readable): Displays file sizes in an easy-to-read format (e.g., 1.2G, 5.5M).
- `--progress`: Shows a progress bar for each file, which is useful for monitoring large files.
- `--partial`: Keeps partially transferred files on the destination. If the transfer is interrupted, you can run the exact same command to resume it from where it left off.
- `-n` or `--dry-run`: performs a trial run without transferring any data, showing you exactly what files would be copied or deleted.
- `--exclude`: Skips files or directories that match a specific pattern.
- `--delete`: Use with extreme caution. This flag deletes files on the destination if they do not exist on the source, making the destination an exact mirror. It can cause permanent data loss if used incorrectly

File transfer clients

- 💻 WinSCP (Windows): GUI for file transfers
- 🌐 Supports FTP, FTPS, SCP, SFTP, WebDAV, and Amazon S3.
- 🔗 winscp.net

- 🍏 Cyberduck (macOS & Windows): GUI client for file transfers and cloud storage.
- 🌐 Supports FTP, SFTP, WebDAV, Amazon S3, Google Cloud Storage, Backblaze B2, Dropbox, and more.
- 🔗 cyberduck.io

Globus for file transfers

Bioinformatics projects involve massive datasets that must move between sequencing centers, HPC clusters, and collaborators. Standard tools are often slow, insecure, or unreliable. Globus is built for researchers because it is:

- ⚡ **Fast & Reliable**: High-speed, secure transfers with auto-resume.
- 🔁 **"Fire-and-Forget"**: Start a transfer, close your laptop, get an email when done.
- 🔒 **Secure**: Authentication + data integrity checks on all transfers.

Globus Endpoints

A. Purdue Endpoints

- Go to transfer.rcac.purdue.edu → login with Purdue.
- In **File Manager**, search for cluster/storage (e.g., Data Depot, Fortress).
- Authenticate → access your files.

transfer

secure endpoint

A

2 Globus moves the data for you

secure endpoint

B

1 you submit a transfer request

3 Globus notifies you once the transfer is complete
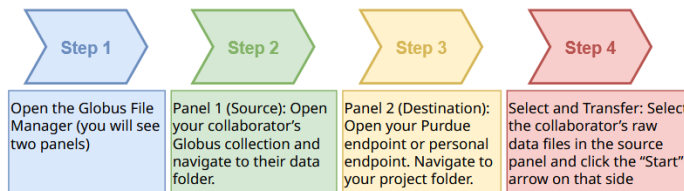
B. Your Collaborator

- Your collaborator must have a Globus endpoint (institutional or personal).
- Ask them to share their collection or provide a share link.
- Once shared, search for their collection in File Manager and authenticate.
- You can now transfer files directly between their endpoint and yours.
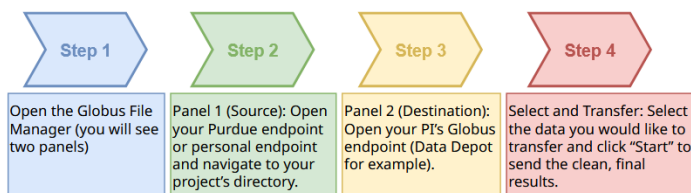
C. Your Personal Computer

- First-time? Log in to transfer.rcac.purdue.edu to activate account.
- Install **Globus Connect Personal** (Windows/Mac/Linux).
- Associate with your Purdue email.
- Name your endpoint (e.g., `JaneDoe-Work-Laptop`).
- Choose folders for transfers.

How to Transfer Data?

Use Case 1: Uploading Raw Data from a Collaborator

Step 1	Step 2	Step 3	Step 4
Open the Globus File Manager (you will see two panels)	Panel 1 (Source): Open your collaborator's Globus collection and navigate to their data folder.	Panel 2 (Destination): Open your Purdue endpoint or personal endpoint. Navigate to your project folder.	Select and Transfer: Select the collaborator's raw data files in the source panel and click the "Start" arrow on that side

Use Case 2: Sharing Final Results with a PI or Collaborator

Step 1	Step 2	Step 3	Step 4
Open the Globus File Manager (you will see two panels)	Panel 1 (Source): Open your Purdue endpoint or personal endpoint and navigate to your project's directory.	Panel 2 (Destination): Open your PI's Globus endpoint (Data Depot for example).	Select and Transfer: Select the data you would like to transfer and click "Start" to send the clean, final results.

✅ **Best Practices for Big Data Workflows**

- 📁 **Raw = read-only** → keep separate "processed" folder.
- 🔄 **Use sync** → only new/changed files transfer.
- 🏷️ **Label transfers** → track activity easily.
- ✓ **Trust checksums** → automatic file integrity verification.
- 🗄️ **Archive small files** → compress before transfer for speed.