



Data Transfer with Globus

Instructors: Brandon Reyes

- Website: www.rc.colorado.edu
- Documentation: <https://curc.readthedocs.io>
- Helpdesk: rc-help@colorado.edu



- Slides: https://github.com/ResearchComputing/data_transfer_with_globus_quick_byte
- Survey: <http://tinyurl.com/curc-survey18>

Outline

- Ways to access your data
- Filesystem structure of CURC resources
- Data transfer using Globus
 - Web-based interface
 - Local Globus endpoint
 - Connecting local and CURC endpoints

Accessing Data on CURC Resources

- When you use CURC resources the data is not on your local machine
- Ways to access the data from your local machine
 - Command line (a variety of tools)
 - Open OnDemand (straightforward GUI)
 - **Globus (GUI with some set up required)**

General Filesystem Structure

/home (2GB)

- Small important data
- Backed up frequently
- Not for sharing files or job output

/projects (250GB)

- Medium sized important data
- Software
- Can be shared with others
- Backed up, but less frequently
- Not for job output

/scratch/alpine (10TB)

- Large data
- Can be shared with others
- Fast Data transfer to compute nodes
- Not backed up!
- Purged after 90 days!

Filesystem documentation: <https://curc.readthedocs.io/en/latest/compute/filesystems.html>

GUI option - Globus

Globus is a service that allows for users to reliably move, share, and discover data

- Command line version is also available
- Our recommended way to transfer data
 - Stable and fast data transfers
 - Transfers continue if a user disconnects
 - Web GUI or Globus Connect Personal GUI
- Supported on all major operating systems
 - Works well with cloud storage providers



Globus Demo

- Logging into the web-based interface for Globus
 - Provides a nice GUI for managing files on CURC resources
- Installing a Globus Endpoint on your local machine
 - Allows you to interact with files on your local machine via Globus

Globus Web-based interface

Globus login is simple and quick

1. Navigate to <https://app.globus.org>
 - CU Boulder users
 - Select “University of Colorado at Boulder” in the dropdown menu
 - CSU users
 - Select “Colorado State University”
 - AMC and RMACC users
 - Select “ACCESS CI (formerly XSEDE)”
2. Login with your credentials
3. Continue with onscreen prompts until you are brought to the Globus Web GUI

Globus Web GUI

Let's take a look!

Local Globus Endpoint

- Required if you want to transfer data between your machine and CURC resources
- Navigate to <https://www.globus.org/globus-connect-personal>

Install Globus Connect Personal

Create a Globus collection on your laptop. Globus Connect Personal is available for all major operating systems.



Globus Connect Personal for Mac

Mac OS X 10.9 or higher

INSTALL NOW >



Globus Connect Personal for Windows

currently supported Windows versions

INSTALL NOW >



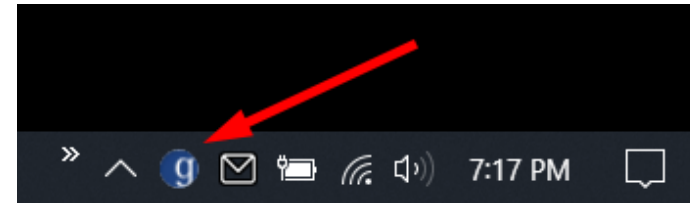
Globus Connect Personal for Linux

for common x86 distributions

INSTALL NOW >

Local Globus Endpoint-Installed

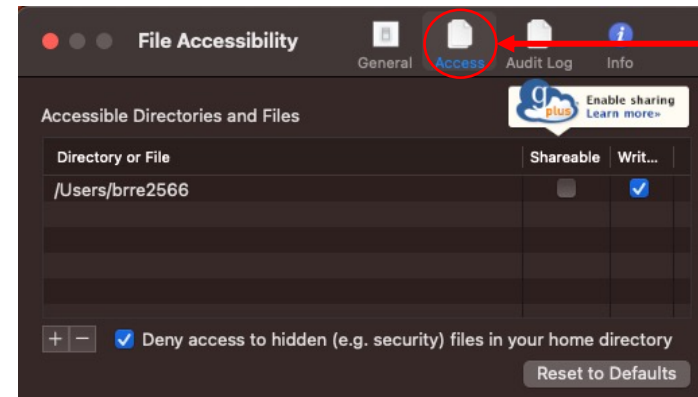
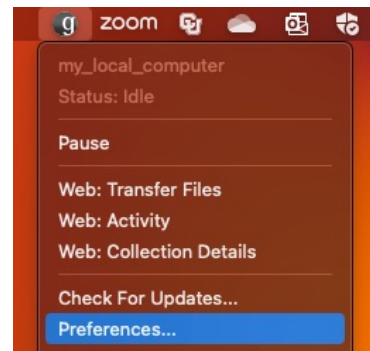
- For “Collection Name” you can use anything name you would like
 - You will use this as a reference name later, so please remember it!
- For Mac and Windows, it is easy to see that Globus has been installed:



- For Linux, at the end of the installation, the setup will exit with a message stating it was successfully set up and the main Globus Connect Personal application will launch

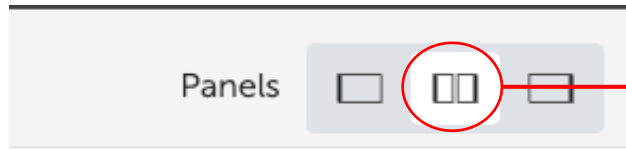
Local Globus Endpoint-Configure

- Once the Globus endpoint has been installed on your computer, you then need to configure it.
- On Mac and Windows follow the “Configuration” section
- On Linux follow the “Running” section
- During Configuration you will need to let Globus know what directories or files it can have access to

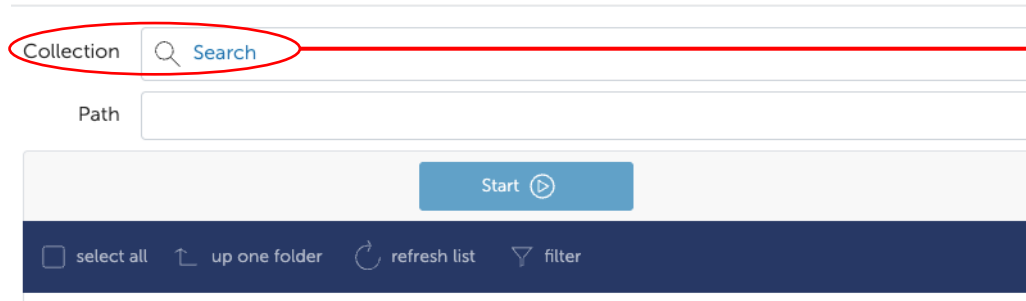


Select the “Access” tab and add folders you want to be accessible by Globus

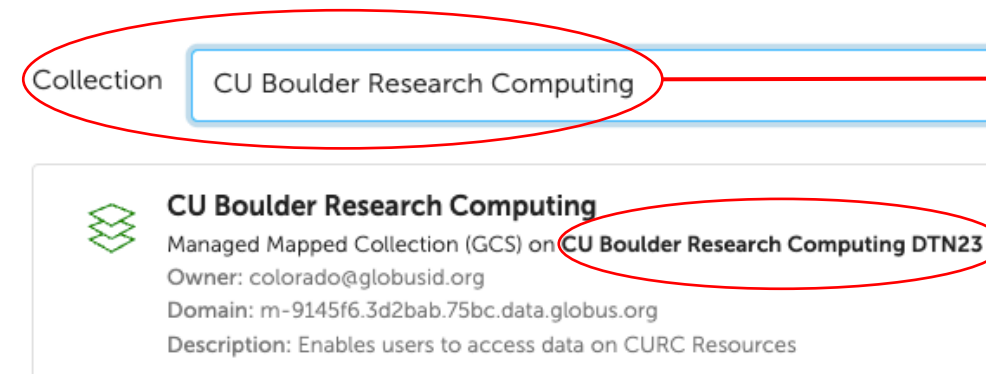
Connecting local endpoint to CURC resources



Step 1: Select “set two pane”
in the upper right corner

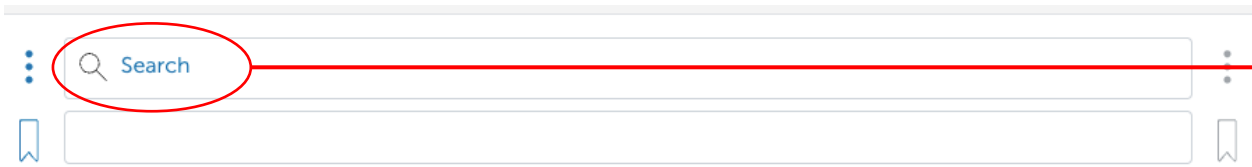


Step 2: In the left pane
click the search bar



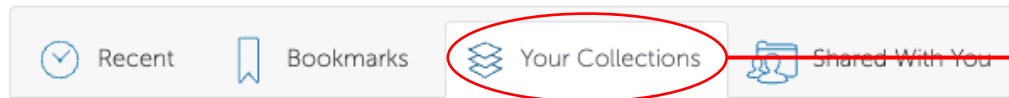
Step 3: Search “CU Boulder
Research Computing”

Make sure you are
using DTN23!

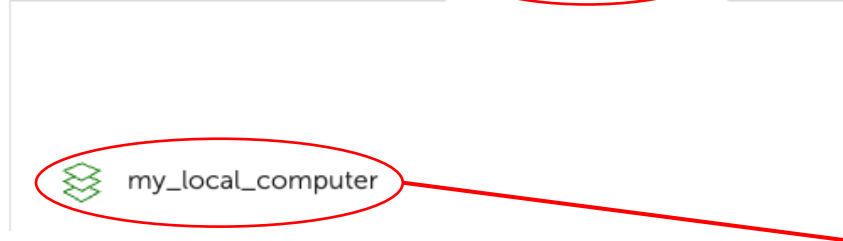


Step 4: In the right pane,
click on “Search”

Collection
Start typing the name of a data collection or select one below



Step 5: Select the “Your
Collections” tab



Step 6: Click the collection
name that **you** chose when
you created your local
Globus endpoint

Let's check it out!

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.
 - See <https://scholar.colorado.edu/concern/datasets/9593tw13k>
 - Can assign various permissions to specific users or all users withing Globus
 - More information on here: <https://docs.globus.org/how-to/share-files/>

Thank you!!

For more help contact rc-help@colorado.edu

Additional documentation: <https://curc.readthedocs.io/en/latest/compute/data-transfer.html>

Survey: <http://tinyurl.com/curc-survey18>

