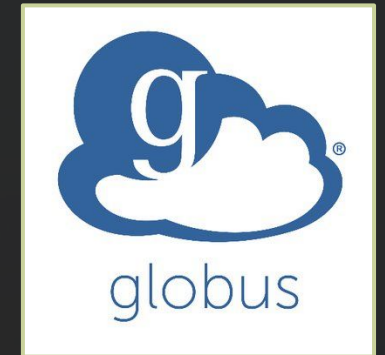


Using Globus at NCAR



Brian Vanderwende
CISL Consulting Services

February 20, 2020

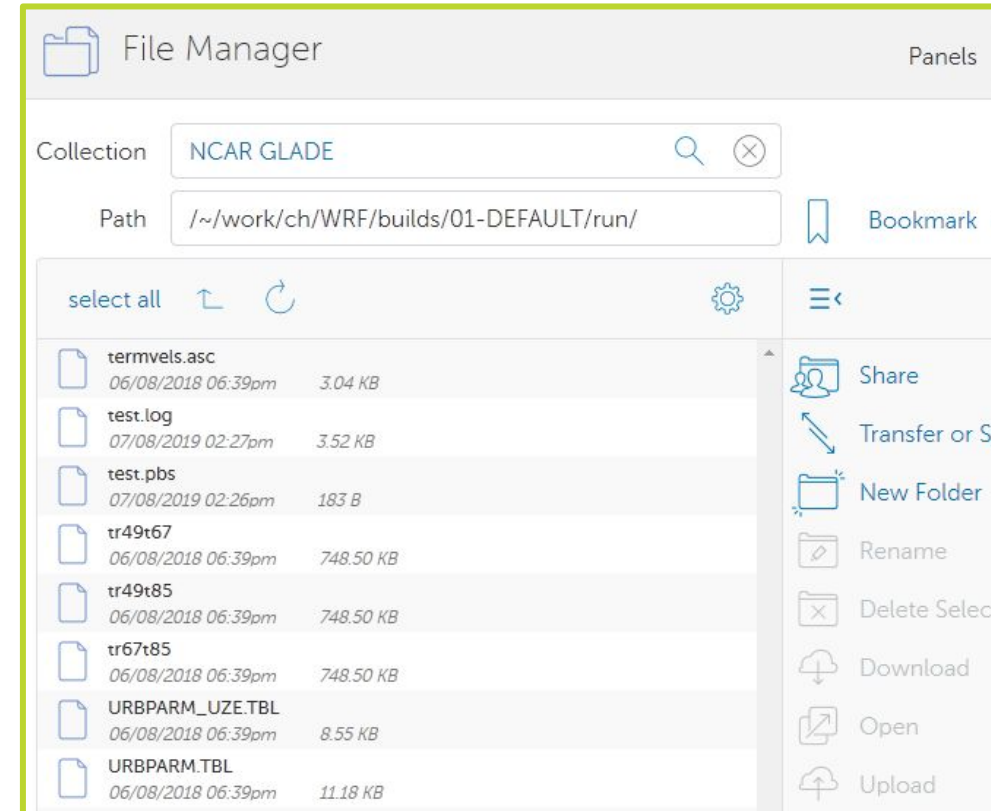


Globus is a tool for fast and reliable data transfers between internal and external storage platforms

- Globus transfers use the GridFTP protocol, which can send multiple chunks of data in parallel (unlike rsync) to achieve high transfer rates
- GridFTP allows for fault-tolerant transfers, so you can resume a transfer if network connectivity is poor
 - Globus automatically resumes halted transfers
- Transfers can be made either at the command line or via the Globus web interface
 - The web interface provides a GUI and makes cross-platform transfers easy
- The Globus service attempts to manage multiple accounts with many varied means of authentication

Globus is not a data management tool (though it has some capabilities)

- You can list directory contents, move, copy, and delete files and folders
- However, these operations do not have the versatility of equivalent command-line tools like cp, mv, and rm
- **Globus transfers do not preserve permissions** - the default permissions on the destination are assigned
 - Side-effect: binaries lose execution permissions after a transfer
- **Symbolic links are skipped in a transfer**

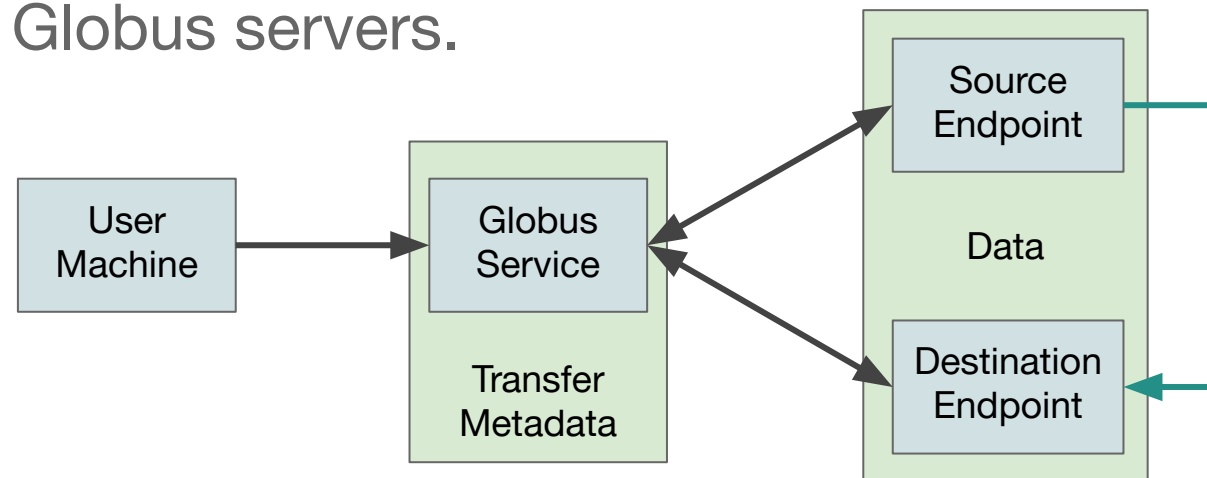


So when should you use Globus?

1. You are transferring any significant amount of data (e.g., ≥ 100 MB) to or from an external site that supports Globus
2. You are moving large amounts of data (e.g., ≥ 1 GB) between NCAR endpoints and want the transfer to be managed in the background
3. You want to share data with external collaborators who do not have access to NCAR systems via the NCAR Data Sharing Service

The basics of a Globus transfer


Selected data are scheduled to be transferred between two “collections” via GridFTP “endpoints”. This transfer is then managed in the background and the result logged on Globus servers.



A Globus transfer may require up to 3 logins - to the Globus service, and to the source and destination endpoints

**There are multiple ways to
use the Globus service at
NCAR...**

Method 1: the web graphical interface at www.globus.org



Log in to use Globus Web App


Use your existing organizational login
e.g., university, national lab, facility, project

Look-up your organization...

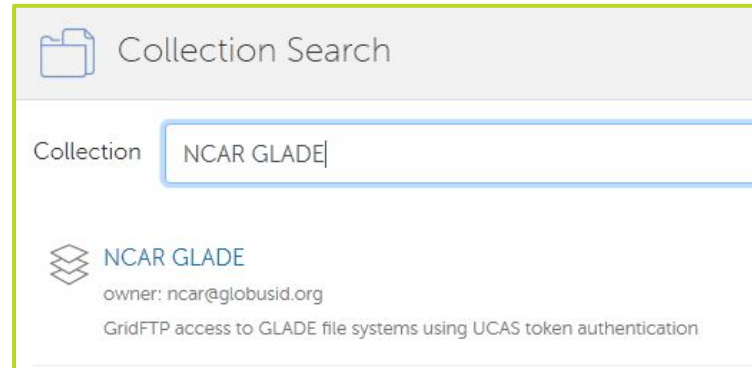
Didn't find your organization? Then use [Globus ID to sign in.](#) (What's this?)

Continue

Or


 Sign in with Google  Sign in with ORCID ID

Log into the Globus service - no dedicated method for NCAR... use either Google account or create a Globus ID



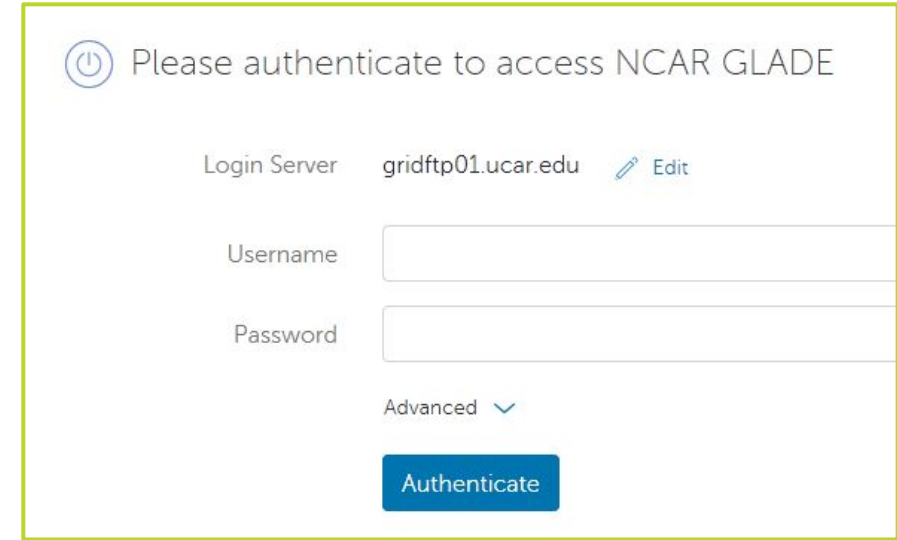
Collection Search

Collection NCAR GLADE


 NCAR GLADE
owner: ncar@globusid.org
GridFTP access to GLADE file systems using UCAS token authentication

Search for collections (endpoints) ... NCAR offers:

- **NCAR GLADE**
- **NCAR Campaign Storage**




Please authenticate to access NCAR GLADE

Login Server gridftp01.ucar.edu  Edit

Username

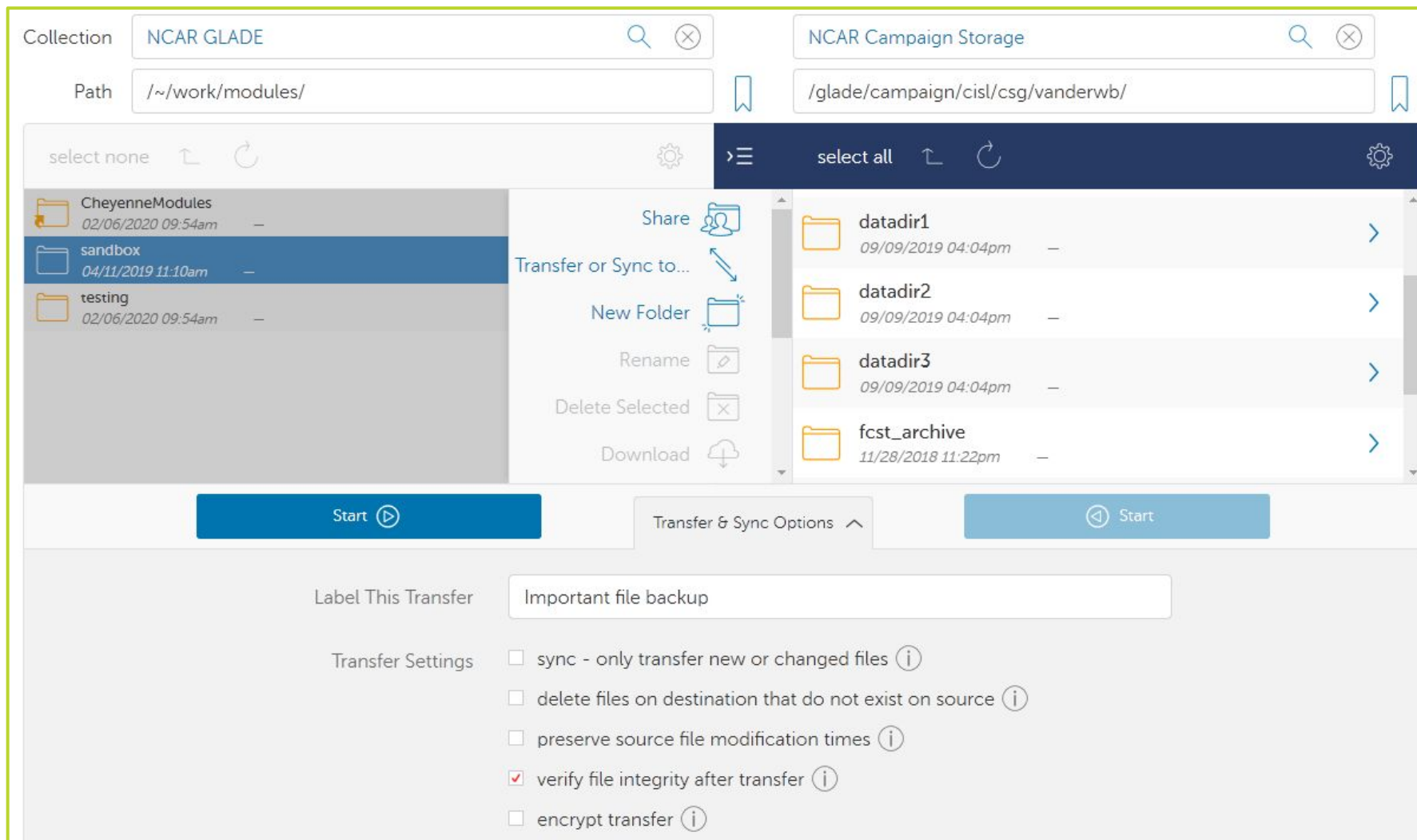
Password

Advanced 

Authenticate

Authenticate to the endpoint. For NCAR, use two-factor auth method (currently Duo Mobile or Yubikey)

Default lifetime: 24 hours




Select files and/or folders, and optionally name and configure your transfer. The selected data will be scheduled to be copied to the active directory on the destination endpoint

Activity List ✓ NCAR Campaign Storage to NCAR GLADE
transfer completed

Overview Event Log

Task Label	NCAR Campaign Storage to NCAR GLADE	<table border="0" style="width: 100%;"> <tr><td style="text-align: right;">9</td><td>Files</td></tr> <tr><td style="text-align: right;">2</td><td>Directories</td></tr> <tr><td style="text-align: right;">83.92 MB</td><td>Bytes Transferred</td></tr> <tr><td style="text-align: right;">13.82 MB/s</td><td>Effective Speed</td></tr> <tr><td style="text-align: right;">0</td><td>Skipped</td></tr> </table>	9	Files	2	Directories	83.92 MB	Bytes Transferred	13.82 MB/s	Effective Speed	0	Skipped
9	Files											
2	Directories											
83.92 MB	Bytes Transferred											
13.82 MB/s	Effective Speed											
0	Skipped											
Source	NCAR Campaign Storage ⓘ											
Destination	NCAR GLADE ⓘ											
Task ID	cd4b6fd6-513b-11ea-971b-021304b0cca7											
Owner	Brian Vanderwende (vanderwb@globusid.org)											
Condition	SUCCEEDED											
Requested	2020-02-16 09:13 pm	Hide debug data										
Completed	2020-02-16 09:13 pm											
Transfer Settings	<ul style="list-style-type: none"> • verify file integrity after transfer • transfer is not encrypted • overwriting all files on destination 											

Debug Data 

DATA_TYPE	task
bytes_checksummed	0

- Metadata, logs, and debug data can be viewed for recent transfers in the activity tab
- Use this info to track and confirm success, along with transfer rates between endpoints
- Any transfer faults (recoverable or otherwise) can be viewed in the Event Log
- **This metadata does not persist indefinitely on the web app**

Method 2: the Globus command-line interface

The Globus CLI is a Python package that enables command-line interaction with the Globus service to schedule and inspect transfers

- Available in default user environment on our data-access nodes:
ssh -l username data-access.ucar.edu
- Also available on Cheyenne* and Casper via NCAR Package Library:
module load python; ncar_pylib

* Globus CLI commands do not work on Cheyenne batch nodes because these nodes lack internet connectivity and thus cannot reach the Globus service

Authentication with the Globus CLI

```
cheyenne$ module load python
cheyenne$ ncar_pylib
cheyenne$ globus login
Please authenticate with Globus here:
-----
https://auth.globus.org/v2/oauth2...
-----

Enter the resulting Authorization Code here: XXX

You have successfully logged in to the Globus CLI!

cheyenne$ globus endpoint search 'NCAR GLADE' --filter-owner-id
ncar@globusid.org --format UNIX --jq 'DATA[0].id'
d33b3614-6d04-11e5-ba46-22000b92c6ec
cheyenne$ EPGLADE=d33b3614-6d04-11e5-ba46-22000b92c6ec
cheyenne$ globus endpoint activate --force --no-autoactivate
--myproxy --myproxy-lifetime 24 $EPGLADE
Myproxy username: vanderwb
Myproxy password:
Endpoint activated successfully using a credential fetched from
a MyProxy server.
cheyenne$ globus endpoint is-activated --format UNIX --jq
"expire_time" $EPGLADE
2020-02-19 20:53:32+00:00
```

Options Defined

- format UNIX** - provide output in parsable format
- jq 'FIELD'** - restrict output of command to a specific field
- force** - activate endpoint even if already activated (ensures lifetime is correct)
- no-autoactivate** - use this authentication method, even if another is already active
- myproxy** - activation method that uses Duo/Yubikey for NCAR endpoints
- myproxy-lifetime N** - specify the proxy lifetime in hours (default is 12; max is 720)

File transfers with the Globus CLI (bash)

```
cheyenne$ EPSTORE=$(globus endpoint search 'NCAR Campaign'
--filter-owner-id ncar@globusid.org --format UNIX --jq
'DATA[0].id')
cheyenne$ globus transfer --recursive --sync-level mtime
--label "Model Project - Data Storage"
$EPGLADE:/glade/scratch/$USER/output/run04
$EPSTORE:/glade/campaign/LAB/GROUP/$USER/model_proj/run04
Message: The transfer has been accepted and a task has been
created and queued for execution
Task ID: 9be00ecc-529c-11ea-971b-021304b0cca7
cheyenne$ TID=9be00ecc-529c-11ea-971b-021304b0cca7
cheyenne$ globus task wait --timeout 3600 $TID
cheyenne$ globus task show $TID
Label:                Model Project - Data Storage
Task ID:              9be00ecc-529c-11ea-971b-021304b0cca7
Is Paused:           False
Type:                TRANSFER
Directories:         1
Files:               3
Status:              SUCCEEDED
...
cheyenne$ globus task show --successful-transfers $TID >
files.$TID
cheyenne$ globus task event-list $TID > eventlog.$TID
```

Options Defined

- recursive** - transfer a specified directory and all of its contents
- sync-level LEVEL** - determine which files will be clobbered at destination (here, we specify files with newer modification time on source)
- label TEXT** - provide a name for the transfer
- timeout SECONDS** - maximum time to wait on an active transfer
- successful-transfers** - causes show command to list all files that were copied from source to destination

Method 3: long-lived authentication with gcert and gci

- Authentications with the Globus service should persist until the user logs out of the service
- Endpoint authentications, however, have expirations - this design is intended to protect your data!
- Since myproxy authentication requires keypress and browser interaction, it does not permit robust unattended usage

To facilitate unattended workflows, we provide InCommon certificates which can be used to authenticate NCAR endpoints without user interaction. **This method is intended for those who want to use Globus in scheduled cron jobs or batch scripts.**

Steps to configure and use InCommon certificate

1. Submit a ticket at <http://support.ucar.edu> and request a free certificate
2. Copy the certificate to your home directory on Cheyenne
3. From Cheyenne, run the **gcert** command to prepare and activate the certificate

Once activated, the certificate can be used to activate endpoints via the CLI:

```
globus endpoint activate --force --no-auto-activate --delegate-proxy  
~/.${USER}-globus.cert --proxy-lifetime 720 $EPGLADE
```

You can also rerun the **gcert** command any time to activate endpoints

gci - a simplified interface to the Globus CLI with certificate integration

- Simple commands to put files on Campaign Storage or get files from Campaign Storage
- Respects relative and absolute paths, unlike CLI commands
- Automatically authenticates GLADE and CS endpoints using certificate

Transfer data file from working directory on GLADE to CS space

gci put data1.dat:lab/group/\$USER/data1.dat

Conditionally transfer data directory from CS to working directory on GLADE

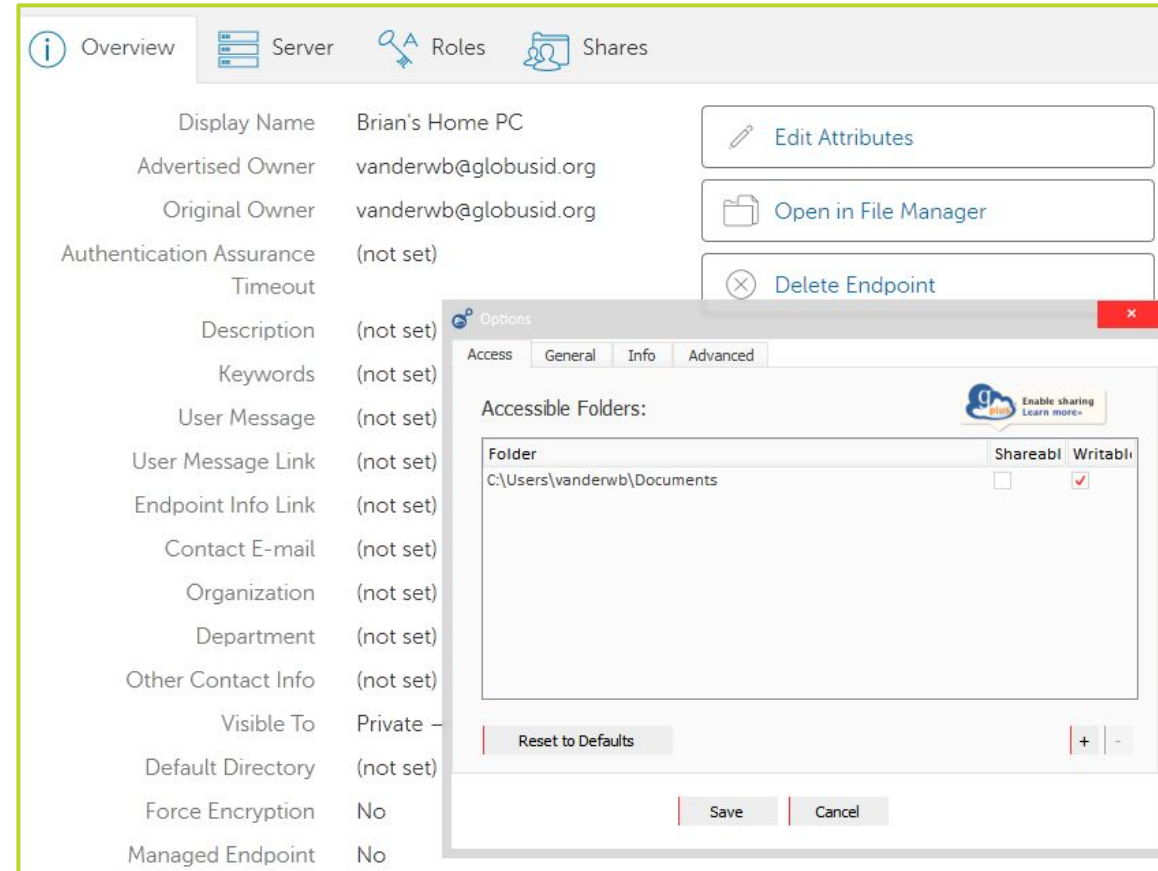
gci cget -r /glade/campaign/lab/group/\$USER/datadir2:`pwd`

Using your workstation as an endpoint with Globus Connect Personal

- You can use Globus to perform transfers to and from your workstation
- Your workstation then becomes an endpoint you can use via either the web service or the CLI
- The endpoint is active whenever your machine is connected to the internet and the utility is loaded

Download the utility from

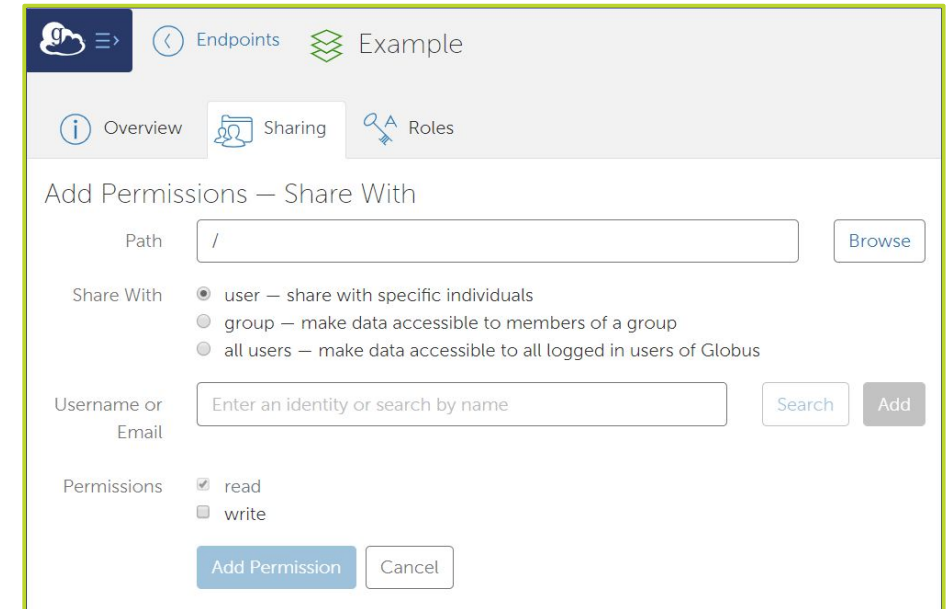
<https://www.globus.org/globus-connect-personal>



Sharing data with external collaborators via the NCAR Data Sharing Service

Users with Cheyenne accounts can request a Data Sharing space, from which you can serve data to individuals without accounts via a Globus endpoint

- The default shared space quota is 50 TB
- The space can only be accessed via Globus or the data access nodes
- Files are not backed up and are deleted after 45 days of inactivity
- Permissions are managed via Globus



The screenshot shows the Globus Data Sharing interface. At the top, there is a navigation bar with a menu icon, a back arrow, and the text 'Endpoints' and 'Example'. Below this is a secondary navigation bar with three tabs: 'Overview' (selected), 'Sharing', and 'Roles'. The main content area is titled 'Add Permissions — Share With'. It contains a 'Path' field with a dropdown menu showing '/' and a 'Browse' button. Below this is the 'Share With' section, which has three radio button options: 'user — share with specific individuals' (selected), 'group — make data accessible to members of a group', and 'all users — make data accessible to all logged in users of Globus'. There is a 'Username or Email' field with a placeholder 'Enter an identity or search by name', a 'Search' button, and an 'Add' button. At the bottom, there are two checkboxes for 'Permissions': 'read' (checked) and 'write' (unchecked). Finally, there are 'Add Permission' and 'Cancel' buttons.

<https://www2.cisl.ucar.edu/resources/storage-and-file-systems/using-the-ncar-data-sharing-service>

Some tips for using Globus at NCAR

- Authentication of the GLADE and Campaign Storage endpoints is linked
 - If you authenticate one, the other becomes usable too
- If you authenticate an endpoint on the website, it will be active at the command line and vice versa
- The Globus API can be used in Python scripts to interact with the Globus service programmatically and integrate Globus into your applications

Getting assistance from the CISL Help Desk

<https://www2.cisl.ucar.edu/user-support/getting-help>

- Walk-in: ML 1B Suite 55
- Web: <http://support.ucar.edu>
- Phone: 303-497-2400

Specific questions from today and/or feedback:

- Email: vanderwb@ucar.edu