

Remote access to eBIC at Diamond

Remote access is available from Linux, Mac OSX and Windows computers using the No machine software (NX). Full instructions including a video can be found on the Diamond web pages:

<http://www.diamond.ac.uk/Beamlines/Mx/Common/Common-Manual/Remote-Access/Connection-to-Diamond.html>

Download and log in then navigate to your directory using NX:

- Download and install the latest NoMachine client from www.nomachine.com
- Run the installed NoMachine client
- First time users should:
 - Create a new connection using the Protocol: NX
 - Set hostname to `nx.diamond.ac.uk` (ensure port is set to 4000)
 - Use password authentication and don't use a proxy
 - Give a name to the new connection, e.g. "Connection to nx.diamond.ac.uk"
- Open this new connection - use this next time you want to access Diamond
- Enter your FedID and Password
- Choose Create a new virtual desktop (You will need Linux desktop – The one with the penguin).
- Choose the display type from the options available then click ok (There are a number of other options but you can just click ok to all)
- Open the applications menu (top left), then go to the System Tools menu and select Terminal
- In the terminal navigate to your directory

```
$ cd /dls/{beamline}/data/{year}/{visit number}
```

e.g.

```
$ cd /dls/m02/data/2020/em15425-1/
```

Available beamlines at eBIC are:

- M01 – Titan Krios (TEM) for CL3 work based at Oxford (temporarily not available)
- M02 – Titan Krios 1 (TEM) located in Zone 2 of Synchrotron Experimental Hall
- M03 – Titan Krios 2 (TEM) located in eBIC Krios hall
- M04 – Talos Arctica (TEM) located in eBIC lab G33
- M05 – Scios (Dual Beam cryo-FIB SEM) located in eBIC lab G28
- M06 – Titan Krios 3 (TEM) located in eBIC Krios hall
- M07 – Titan Krios 4 (TEM) located in eBIC Krios hall
- M11 - Aquilos (Dual Beam cryo-FIB SEM) located in eBIC lab

Checking that data is still being written to your visit directory by EPU

- The Atlas, EPU and raw directories are located in your visit directory
 - The Atlas directory contains the images that make up the grid overview
 - The EPU directory contains all of the images collected during your session except for the movie frames
 - The raw directory contains the movie frames stored in the same directory structure as in the EPU directory
- All these directories are protected from accidental writing/ deleting the data. The only folder one can use to write files and process data is 'processing'.
- In order to check that data is still being collected, type in your session folder

```
cd processing
```

```
$ ls -ltr ../raw/Gr*/Data/* > raw-list.txt  
$ tail raw-list.txt
```

most recent movies will be at the bottom of the list.

- To view most recent non-aligned movie frame averages (as recorded in EPU):

```
$ ls -ltr ../{EPU session}/Gr*/Data/* > epu-list.txt  
$ tail epu-list.txt
```

```
$ module load imod
```

(you need to do it only once per session in your terminal), after that simply type

```
$ 3dmod {file path}/{filename}
```

(you can copy and paste the path and filename of last file displayed after the 'tail' command).

A 3dmod control window and micrograph will pop up. Please, remember to close them down after you have finished checking your micrograph. Please, bear in mind that this micrograph will be NOT motion corrected (unless such option was requested in EPU).

- Alternatively, for viewing the files in this directory you can navigate to respective directory with file browser and check corresponding jpg files:
in processing folder type

```
$ cd ../{EPU session}/{GridSquare number}/Data  
$ nautilus ./
```

- To see number of movies acquired so far, in 'processing' directory type:
\$ wc -l raw-list.txt

- For jpg

\$ display *.jpg (Sometimes the jpegs look poor due to compression)

- Another useful command on linux is eog as this can be used to open up all of the *.jpg currently written to your data directories. This will also give you the total number of images collected so far.

```
$ eog /dls/{beamline}/data/2020/{visit number}/{EPU
$ directory}/Grid*/Data/*.jpg
```

- Please, remember that files in EPU folder are NOT motion corrected sums!
- Regularly refresh file list and check for new files repeating commands (executed from 'processing' folder)

```
$ ls -ltr ../raw/Gr*/Data/*.mrc > raw-list.txt
```

```
$ tail epu-list.txt
```

- Verify that new files appear at the end of the list, and that they have recent time stamp.



Please, note that EPU will stop when LN2 is being refilled (for all sessions), or if you use a K3 camera, when GIF ZLP is being adjusted. However, that should not last longer than 15-20 minutes. Additional time might be required for EPU to move to a new target and acquire an image. Therefore, if after 30 minutes no new files appear at the end of refreshed epu-list.txt, please, contact your Local Contact (LC) directly during normal working hours (Monday-Friday 9am to 5pm) or Experimental Hall Coordinators (EHC) outside normal working hours (EHC contact phone number 01235 77 8787).

Checking that data is still being written to your visit directory for TOMO

- Follow the same setup procedure as detailed for EPU until you reach your visit directory.
- Then navigate to your Tomo4 data folder

```
$ cd {folder you set in tomo4 for storing your tomograms}
```
- Then you need to list the directory by the time that the files were written to disk

```
$ ls -lrt
```
- Check the time stamp on the latest tomogram file (****.mrc) as it should be updated every 2 minutes.
- If you do not see the latest tomogram file being updated for longer than 20 minutes then you should email your local contact (LC) during normal working hours or call the EHC's on 01235 778787

*If you are saving movie frames for your tilt images they will be saved in the raw directory and will be automatically processed and put in a new directory called processed. Within the processed folder there will be a directory for each tomogram that should contain your

tilt series made from movie frame aligned tilt images. This processing pipeline is under construction, feel free to discuss the latest updates with your LC.

Relion on-the-fly processing

Automated data processing at eBIC currently is done using Relion_it with optional use of crYOLO particle picker for true single particles (filament picking and processing is not integrated yet). Your LC should start data automated data analysis in your processing folder. This would require knowledge of pixel size (in super-resolution if this mode is used) and electron dose per square angstrom per movie frame. Also, the smallest and the largest particle dimensions have to be specified for particle picking and analysis. Currently, the main purpose of the Automated data processing at eBIC is ensuring optimal setup and performance of the instrument. Thus, 2D and 3D analysis is done on heavily coarsened data. In case of single particles we expect the Automated pipeline to provide information on drift (output of MotionCorr2), CTF determination (GCTF or Ctffind4), per micrograph particle density (crYOLO particle picker), particle quality, orientation bias and data consistency (2D classification, ab-initio 3D model building and 3D classification). Only C1 symmetry will be assumed during analysis. For filaments, data processing should be capped after CTF estimation until proper tools will be integrated.

How to analyse Relion_it output:

Change to your data processing folder. Type:

```
$ module load EM/relion
```

```
$ relion
```

You can then use Relion GUI or Chimera in the usual way to check data processing progress. Please, note that output of the job which is running at any given time will not be available until its completion. However, previous files can be found in the corresponding job folder denoted as *.old.

Other EM processing software available at diamond:

Diamond has a number of EM processing softwares available. In order to enable you to use them the path to them needs to be set and is done using the module load command. eg:

```
$ module load imod
```

```
$ module load EMAN2
```

for exhaustive list of EM specific software installed on the diamond system.

type:

```
$ module avail EM/
```

Please, note that typing a non-specific request like

```
$ module avail
```

will take too long to run and will display very long list of software versions. Thus, more

specific requests are recommended, like
\$ module avail EMAN

- Please, note that due to increased demand, GPU is currently available to users only for data processing within Scipion automated pipeline activated via scipion-template. All other processing is restricted to CPU cluster only until further notice. **Please, speak to your LC if you have any questions or need advice regarding data processing at Diamond.**

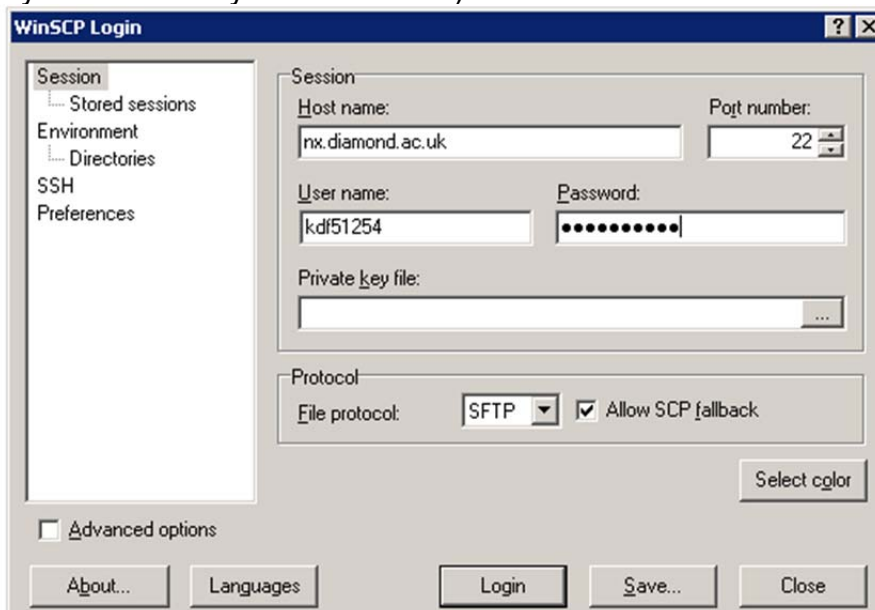
Data Transfer from Diamond

The data collected during your Diamond session will remain on the filesystem for approximately 40 days. The data is archived to a tape backup and removed from spinning disk. The archived copy remains on tape for the medium lifetime. Data can be downloaded from the archive using the topcat web interface (<https://icat.diamond.ac.uk/#/my-data/DLS>)

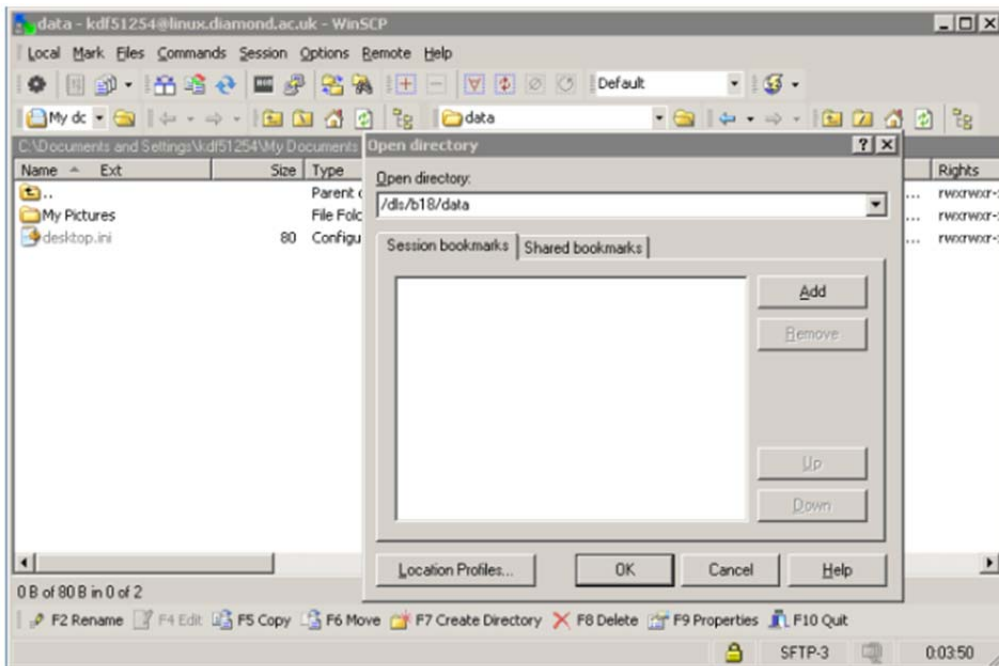
Transfer data to your home institute (SFTP – Slower than using Globus – see next section for Globus)

Windows:

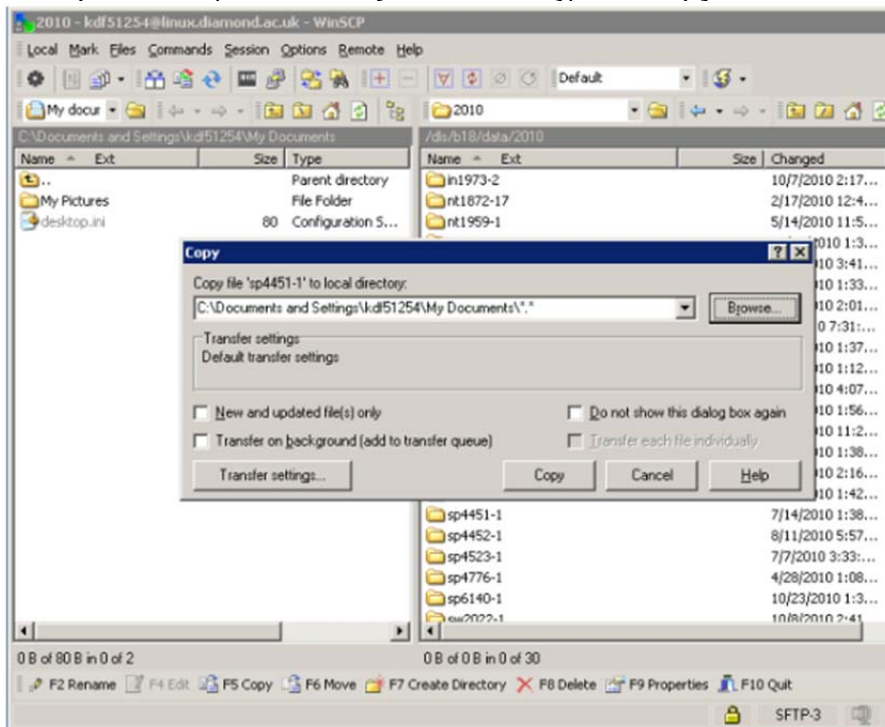
Start WinSCP and connect to the server 'nx.diamond.ac.uk' with your FedID (maybe as CLRC\FedID if your FedID only does not work).



Once connected, open the data directory on the beamline /dls/{beamline}/data/2017/{visit number}



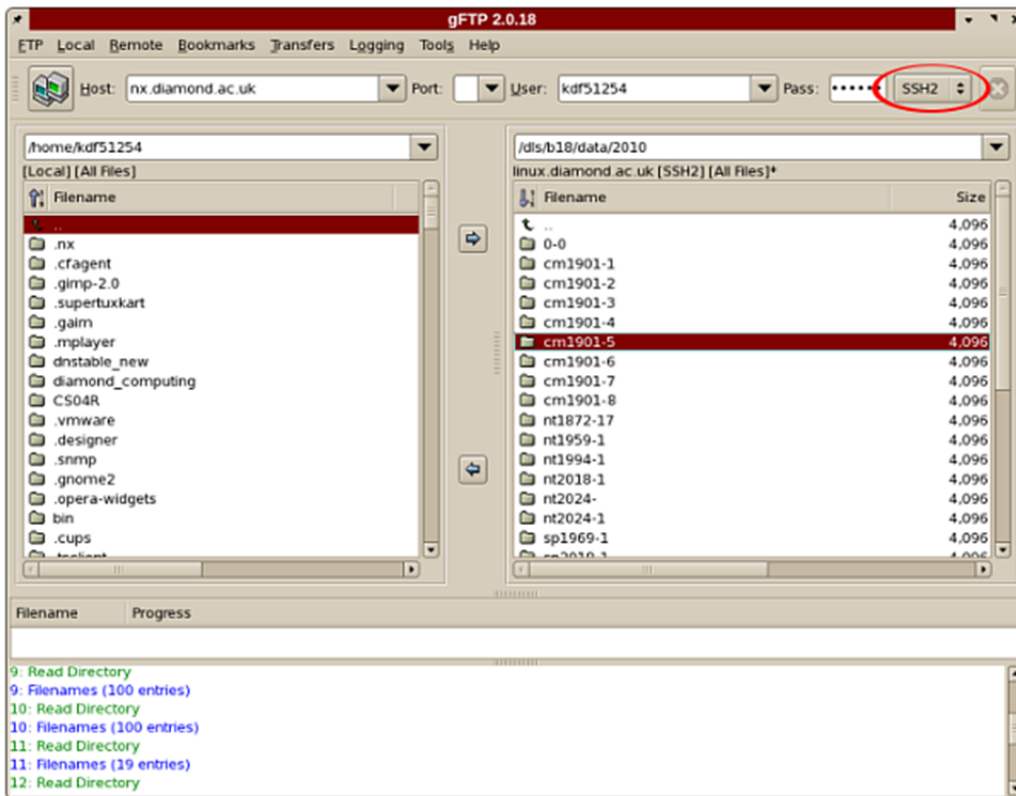
navigate to your experiment (data/2017/{visit number}) and copy the folder



Linux:

There are numerous remote transfer clients for Linux; the shown examples are gFTP (using 'SSH2' as the protocol; the default FTP will not work), and command-line rsync.

You will have to log in using your FedID (if this does not work immediately then try adding CLRC before the FedID). Note that this drops you into your home directory (/home/{FedID}), from where you will have to navigate into the data directory to which you want to go. The data directory can be found as /dls/{beamline}/data/{year}/{visit number} and you can transfer data via drag and drop.



Alternatively, transfer data from your experiment using command-line rsync in a terminal window. Move to the area of your local disk where you will put the data then type `rsync -av -e ssh {fedID}@nx.diamond.ac.uk:/dls/{beamline}/data/{year}/{visit number}/raw .`

Then enter your password and the transfer will start

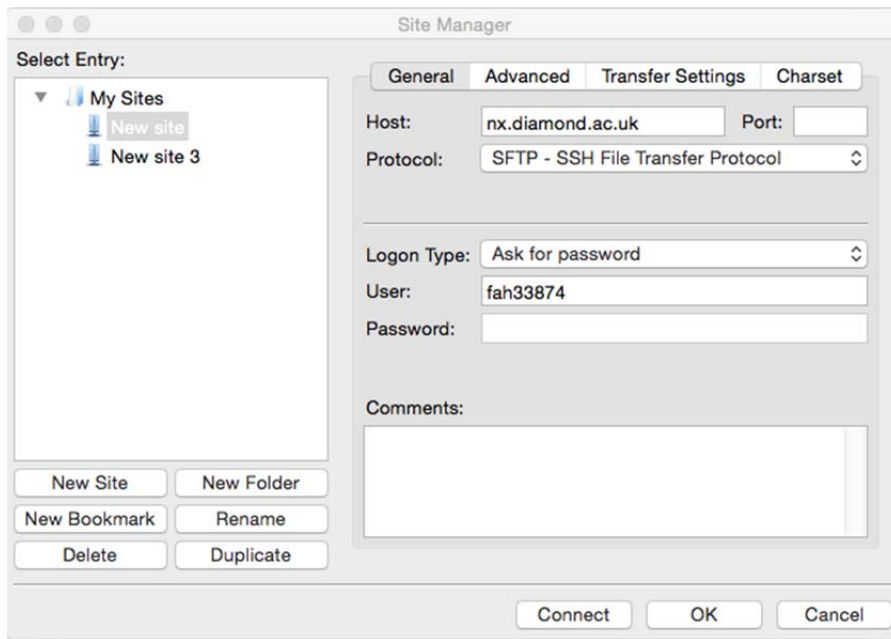
```
kakapo@pw012:~$ rsync -av -e ssh kdf51254@nx.diamond.ac.uk:/dls/i02/data/2010/cm1891-6 /data/test/
kdf51254@nx.diamond.ac.uk's password:
receiving file list ... done
cm1891-6/
cm1891-6/.workspace/
cm1891-6/processing/
cm1891-6/spool/
cm1891-6/xml/

sent 50 bytes  received 179 bytes  26.94 bytes/sec
total size is 0  speedup is 0.00
kakapo@pw012:~$ ls -l /data/test/
total 4
drwxrws--- 6 kakapo kakapo 4096 Oct 26 00:33 cm1891-6
kakapo@pw012:~$ ls -l /data/test/cm1891-6/
total 12
dr-xrws--- 2 kakapo kakapo 4096 Oct 26 00:33 processing
drwxrws--- 2 kakapo kakapo 4096 Oct 26 00:33 spool
drwxrws--- 2 kakapo kakapo 4096 Oct 26 00:33 xml
kakapo@pw012:~$
```

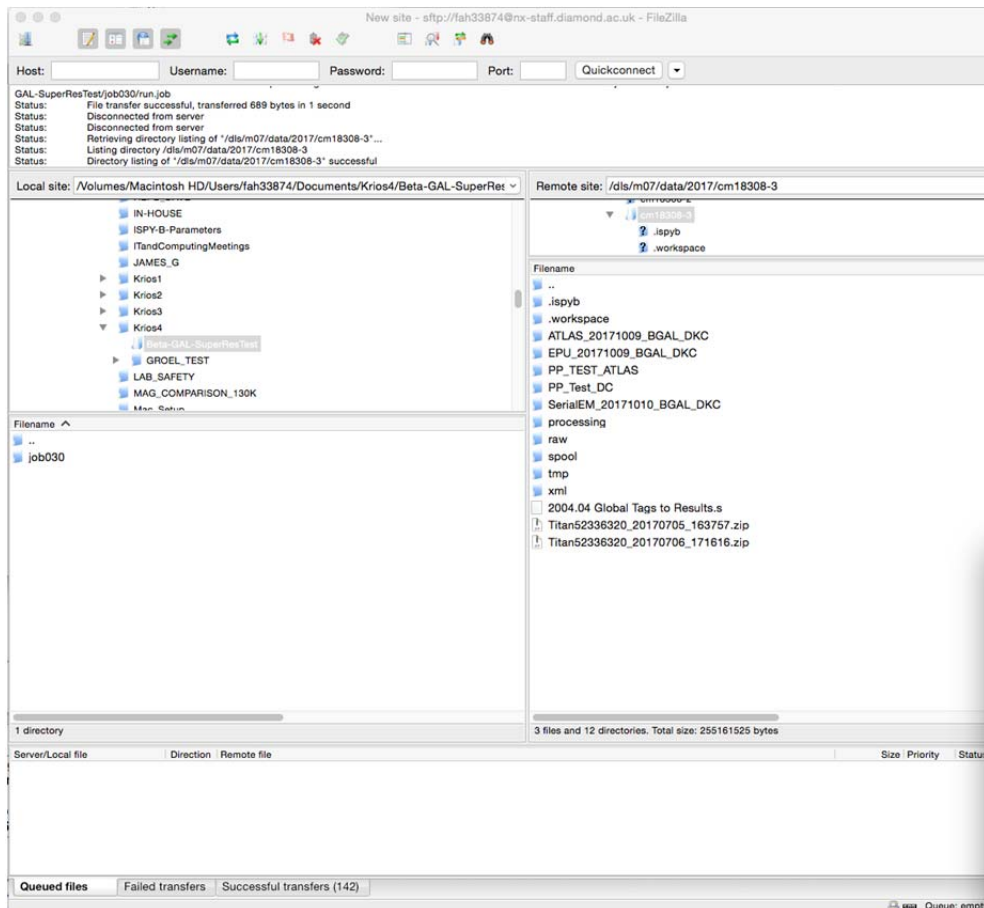
Mac:

There are numerous remote transfer clients for Mac; the shown example is FileZilla. The rsync transfer shown in the linux section will also work using mac terminal.

Open Filezilla and click on the site manager icon and select new site. Then enter `nx.diamond.ac.uk` in the host and use the SFTP protocol. Then put your FedID in as the user and then select Ask for password as logon type. Then click connect and enter your password.



Navigate to your visit folder `/dls/{beamline}/data/{year}/{visit number}` in the left hand side of the panel and then drag the directories you require into the area on your local file system



Data transfer using Globus (Much faster than SFTP)

1 Introduction

This is how researchers can download large datasets from Diamond using Globus Online, a service designed to optimise the file transfer process.

1. Log in to a web interface called Globus Online
2. Log in to Diamond with your usual FedID
3. You will see a folder browser with all mounted beamlines
4. Select the files you want to transfer to your institute's Globus server
5. Click the transfer button, the file transfer is managed by Globus Online

- You may close the window and shut down your laptop - the transfer between servers will proceed and you be emailed when the transfer is complete

2 Setting up Globus

To send files using Globus Online you will need: 1. A free Globus Online account 2. Globus Connect endpoint software

- Globus Connect might already be set up at your institution
- You can set up a Globus Connect endpoint on your own server, workstation or laptop

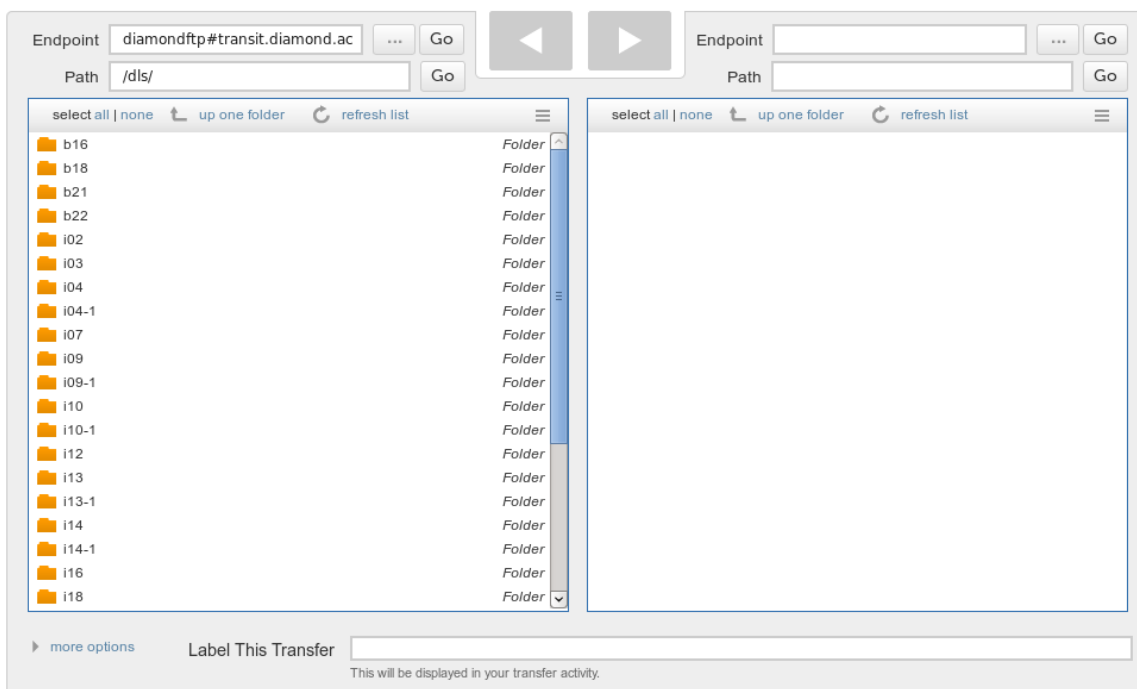
3 Getting started

Follow steps 1, 2, 3 and 4 of the Globus 'Getting Started guide' at <https://www.globus.org/researchers/getting-started>

4 Seeing your files at Diamond

Diamond's Endpoint is called `diamondftp#transit.diamond.ac.uk`

1. Go to the Globus Transfer Files page
2. Go to the left box
3. In the 'endpoint' put `diamondftp#transit.diamond.ac.uk` and click Go
4. In the Username and Password boxes put your Diamond FedID and password.
5. When you click 'Authenticate' you should see a list of beamlines on the left-hand side.



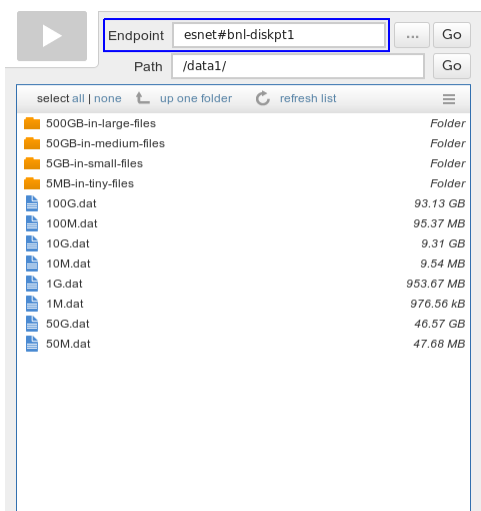
5 Finding your institution

Many campuses, research labs, and cross-site projects have already set up Globus server endpoints on their servers.

5.1 Your institute already has a Globus endpoint

If your institution has already set up a Globus endpoint:

1. Enter the institute's endpoint name on the Globus Transfer Files page and click Go
2. Put in your username and password for your home institute and click Authenticate

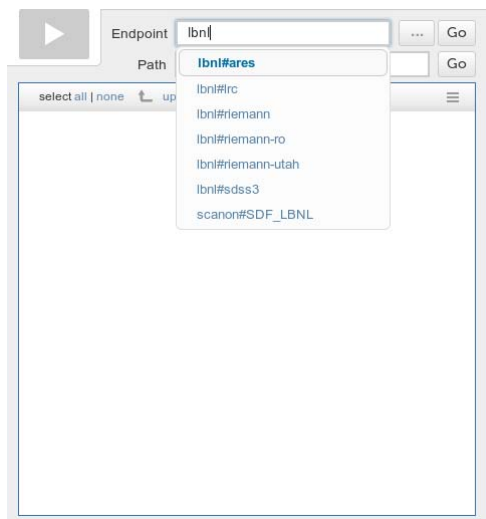


5.2 Searching for your institution's endpoint

The endpoint name is a Globus-specific name that identifies a server inside the Globus Online system.

Globus endpoints are named using the format <globus-username>\#<resource-name>. For example, the XSEDE project has a Globus account under the username 'xsede' and server endpoints named xsede#stampede (for the Stampede system at Texas Advanced Computing Center).

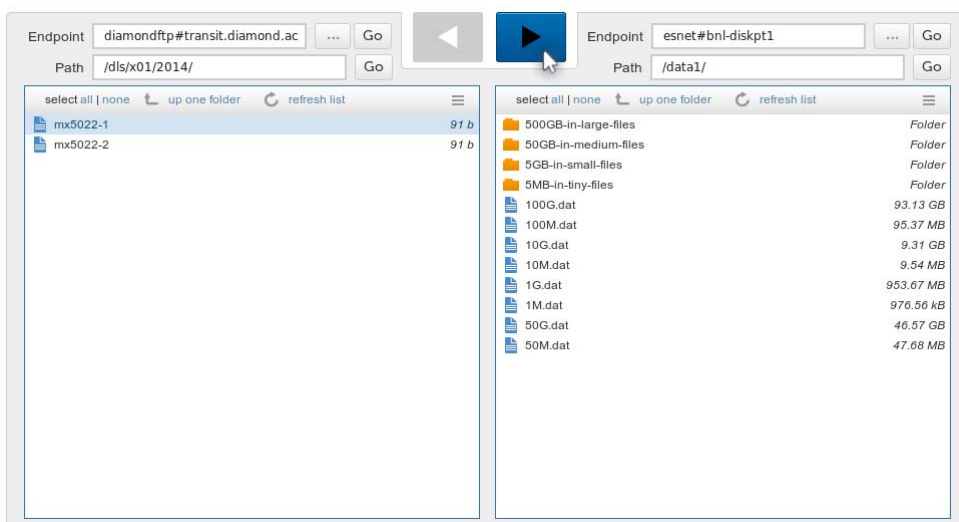
If you don't know your institute's endpoint name, you can use the Globus Transfer Files page to search. Just enter a likely name in the Endpoint field and look for the resource in the scroll-down list. For example, 'umich', 'nersc', 'xsede', . . .



5.3 Your institute is not found in the list!

The Globus Online software needs to be running at your institution so that we can send files to you. See the Appendix 'Installing the Globus Online software' to install Globus Online on a server, workstation or laptop.

6 Sending files



1. On the Transfer Files page Use Globus Online to log in to Diamond on one side, and log in to your institution on the other side
2. You will see a folder browser on the left with all available beamlines
3. Use the folder browser to find your visit

4. Select the files or folder you want to transfer to your institute
5. Click the transfer button to start the transfer
6. The file transfer is managed by Globus Online
 - You may close your web browser; the transfer will carry on between servers
 - You will get an email when the transfer is finished

Appendix A Installing the Globus Online software

If your institution is not set up as a Globus endpoint, you need to set up a Globus Connect endpoint on your own server, laptop or desktop.

- We recommend installing on a dedicated server if your institution has experience working with Linux systems

A.1 Installing on your own desktop or laptop

This tells you how to install Globus Connect Personal on your desktop or laptop You will need:

1. A machine running either a Windows, Linux or MacOS
2. Administrator permissions to install software on the machine
3. Access to enough free disk space for your research data

Follow the guide at: <https://www.globus.org/globus-connect-personal>

A.2 Setting up a dedicated Globus server at your institution

This tells you how to install Globus Connect Server on a dedicated server You will need a server with:

1. Access to enough free disk space for your research data
2. A recent version of Linux:
 - (Red Hat Enterprise Linux (RHEL), CentOS, Scientific Linux, Fedora, SuSE Linux Enterprise Server, Ubuntu or Debian)
3. A reliable, fast network connection Follow the guide at: <https://www.globus.org/globus-connect-server> and click 'Install Globus

Connect Server'