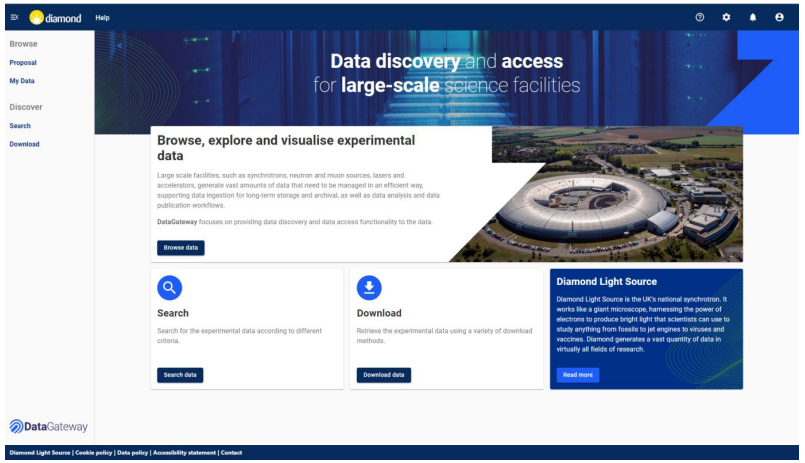
Retrieval of Data

Retrieval of Data: Summary

- Archiving Policy
- ISPyB
  - Auto-processed data retrieval
- rsync < 20 Gb Live data
  - https://www.diamond.ac.uk/Users/Experiment-at-Diamond/IT-User-Guide/Not-at-DLS/Retrieve-data/in40days/rsync.html
- Globus
  - Live data retrieval
- Data Gateway
  - Archived raw & processed Data retrieval

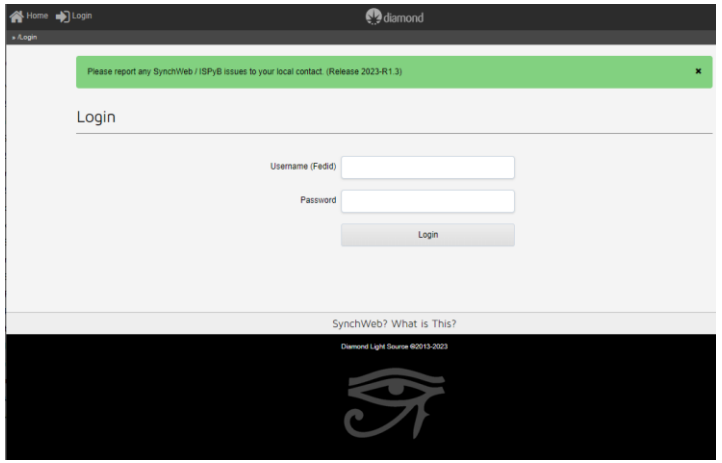
Archiving & Data Release Policy

- Raw Data (i.e. Diffraction Images) can be archived any time from **40 days** after the visit for academic users
- Archived data is stored on tape – a single non-redundant copy is kept on tape
- Retrieval after 40 days is via Data Gateway
- **3 years** after data is collected - Data can be made available on an open-access basis under a **CC-BY-4.0 licence.**
  - An extension to this period can be made by the principal investigator
- https://www.diamond.ac.uk/Home/Legal-and-Compliance/Policies/Experimental-Data-Management-Pol.html

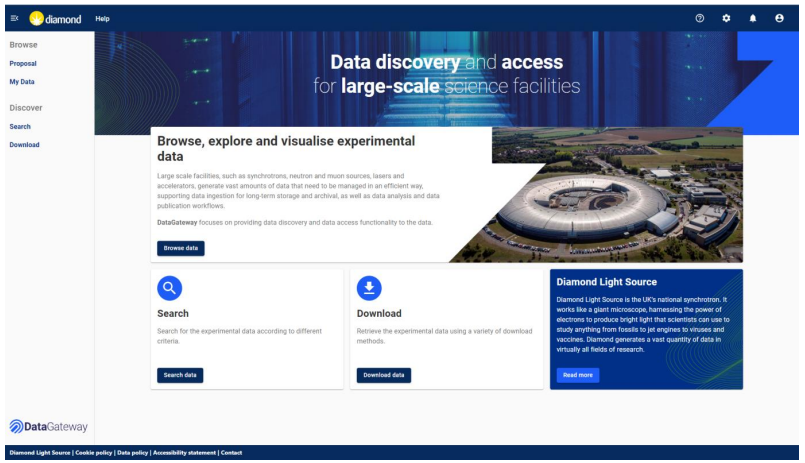


Data Gateway
https://icat.diamond.ac.uk/
Archived Raw Data

Diamond MX – Key Software Interfaces



ISPyB
https://ispyb.diamond.ac.uk/
Processed Data Files



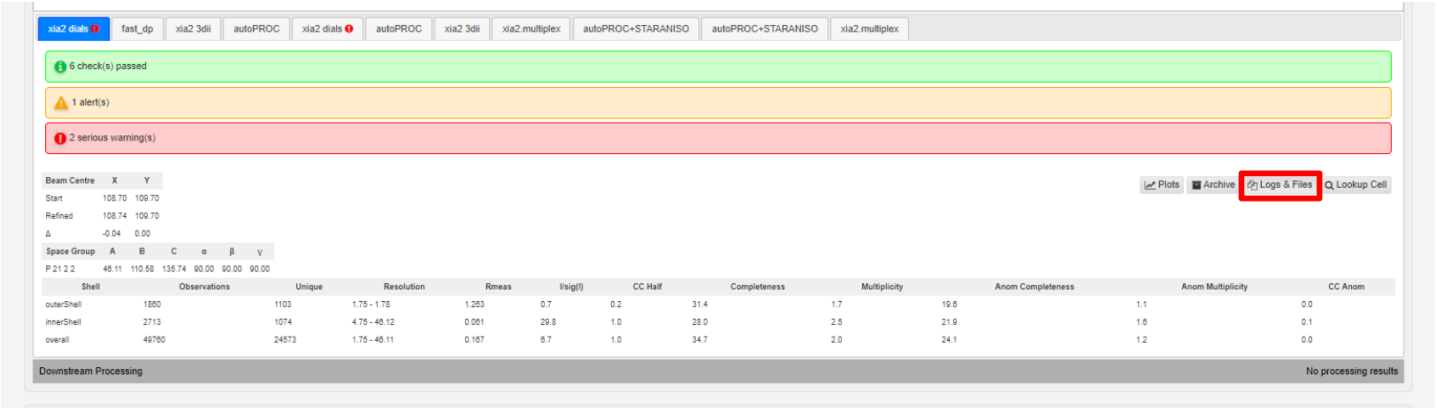
Data Gateway
https://icat.diamond.ac.uk/
Archived Raw Data

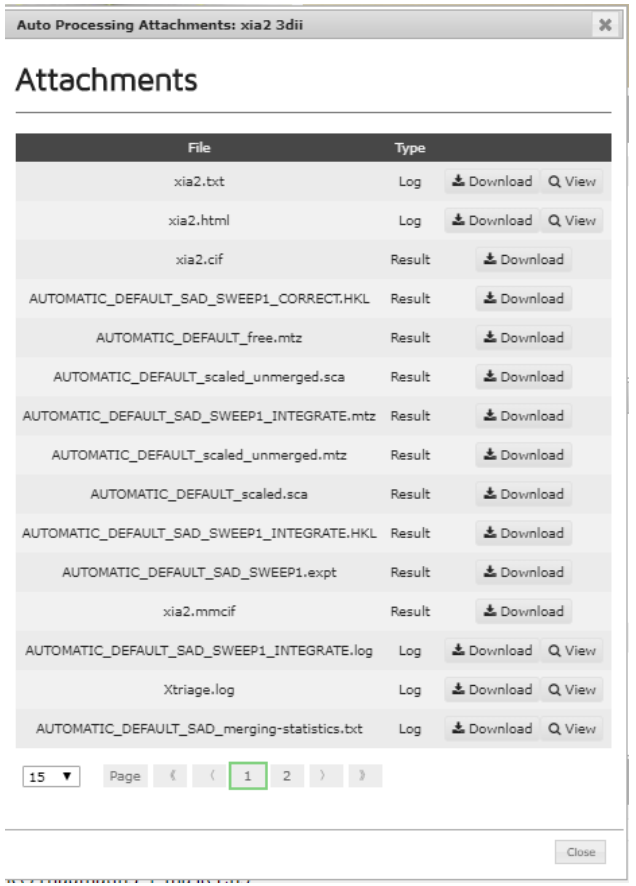
Globus
http://www.globus.org/
Live Data – Raw & Processed

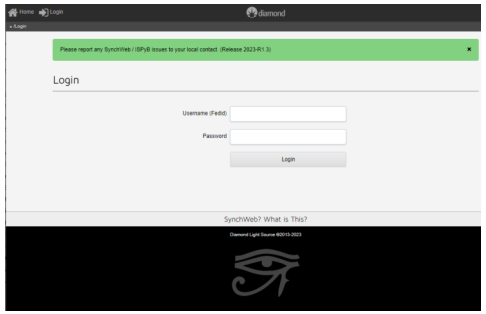
Data Retrieval

ISPyB - Logs & Files

- For each dataset, every auto-processing & downstream processing pipeline has a log and files button
- Here you can retrieve the processed mtz files, logs and other intermediate files
  - An archive of all the files can be downloaded
  - Some intermediate files are only on disk
- Raw images cannot be retrieved via ISPyB



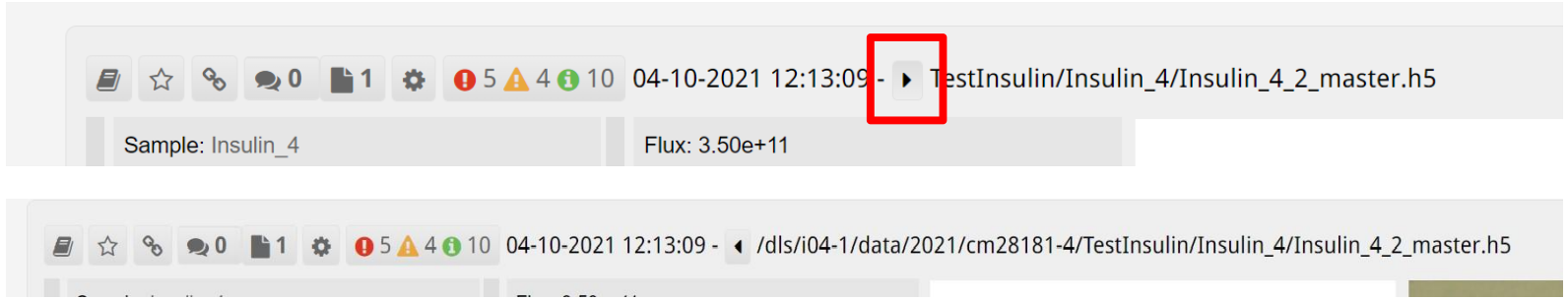




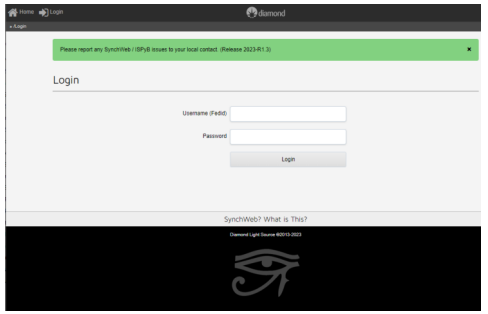
ISPyB
https://ispyb.diamond.ac.uk/
Processed Data Files

ISPyB - Logs & Files

- These can also be retrieved from the disk under the `processed` subdirectory for a visit
- The path for the raw data which can be seen on ISPyB by clicking the triangle icon on a dataset:



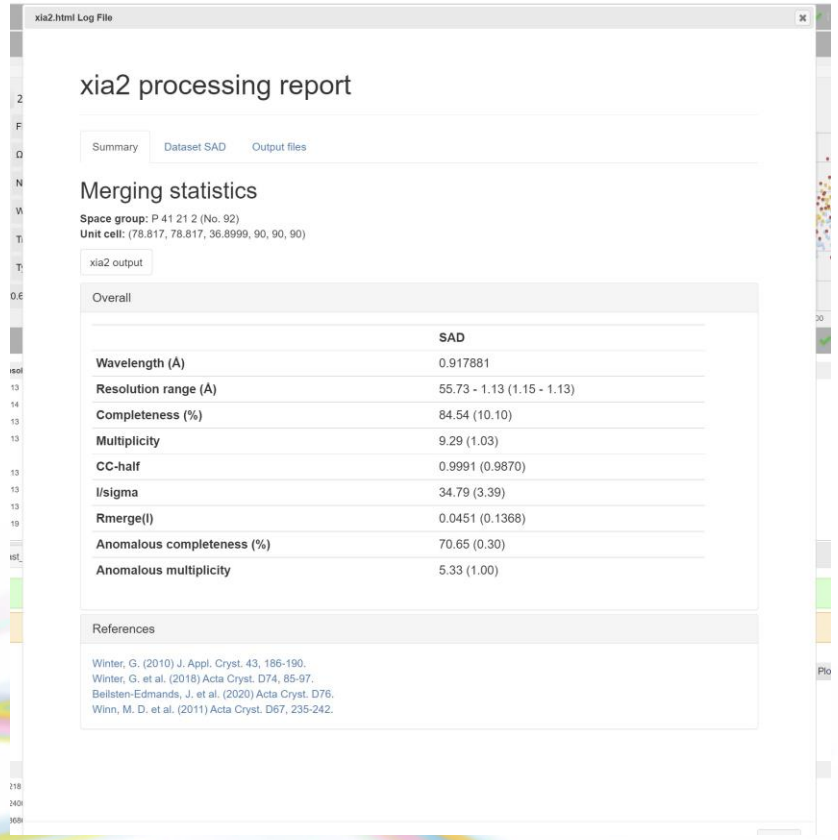
- The end of the path would need to be changed:
- `.../<visit>/processed/<protein>/<sample_name>/<dataset_name>/<processing-pipeline>/`

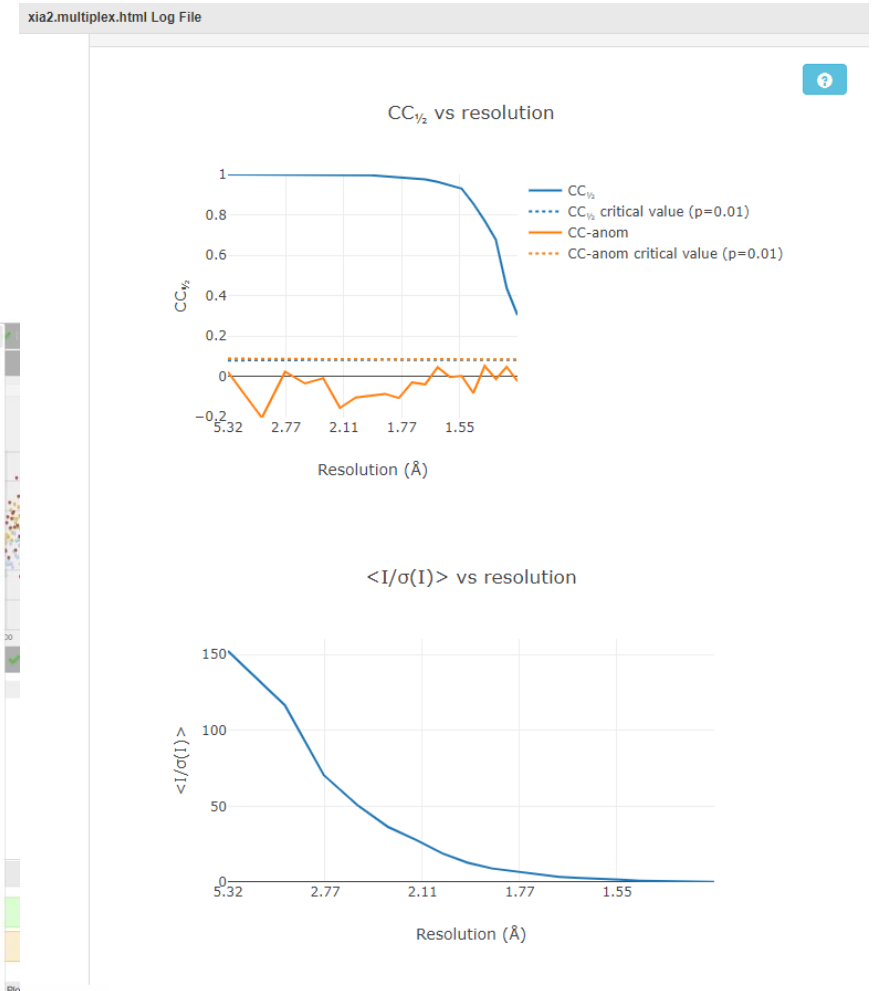


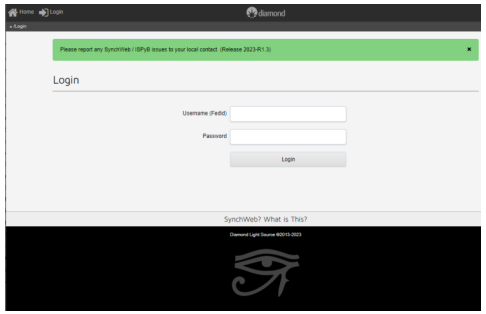
ISPyB
https://ispyb.diamond.ac.uk/
Processed Data Files

ISPyB - Logs & Files

- Reports such available from logs and files interface can have very useful graphs for assessing data quality during collection



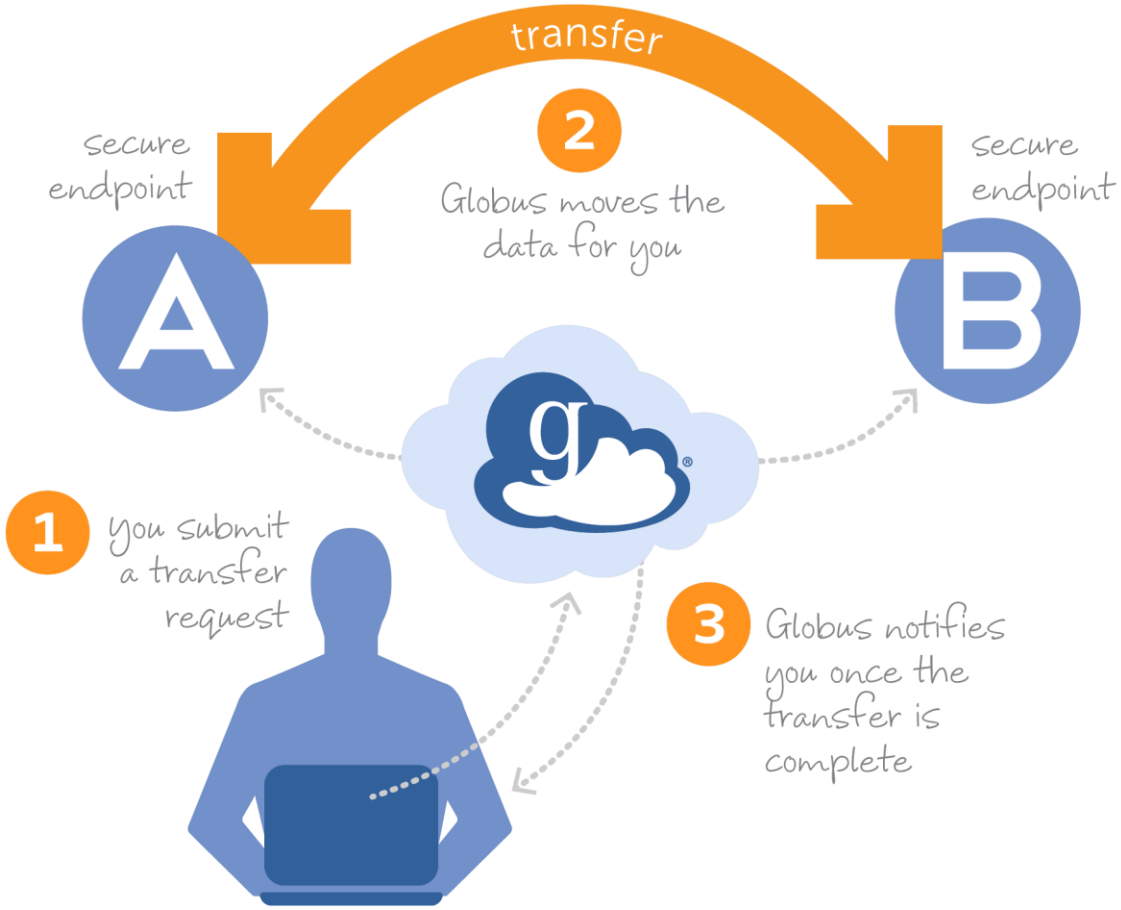




ISPyB
https://ispyb.diamond.ac.uk/
Processed Data Files

Globus

- Globus is a file transfer service free at the point of service to non-profit organisations
- Many institutions have an endpoint to which data can be transferred
- Users can create a personal endpoint to transfer to a personal/local machine
  - https://www.globus.org/globus-connect-personal
- LS-CAT & SERCAT users
  - Some data transfers may be managed on your behalf
- Globus has Python scripting interface for automated data retrieval:
  - https://docs.globus.org/cli/

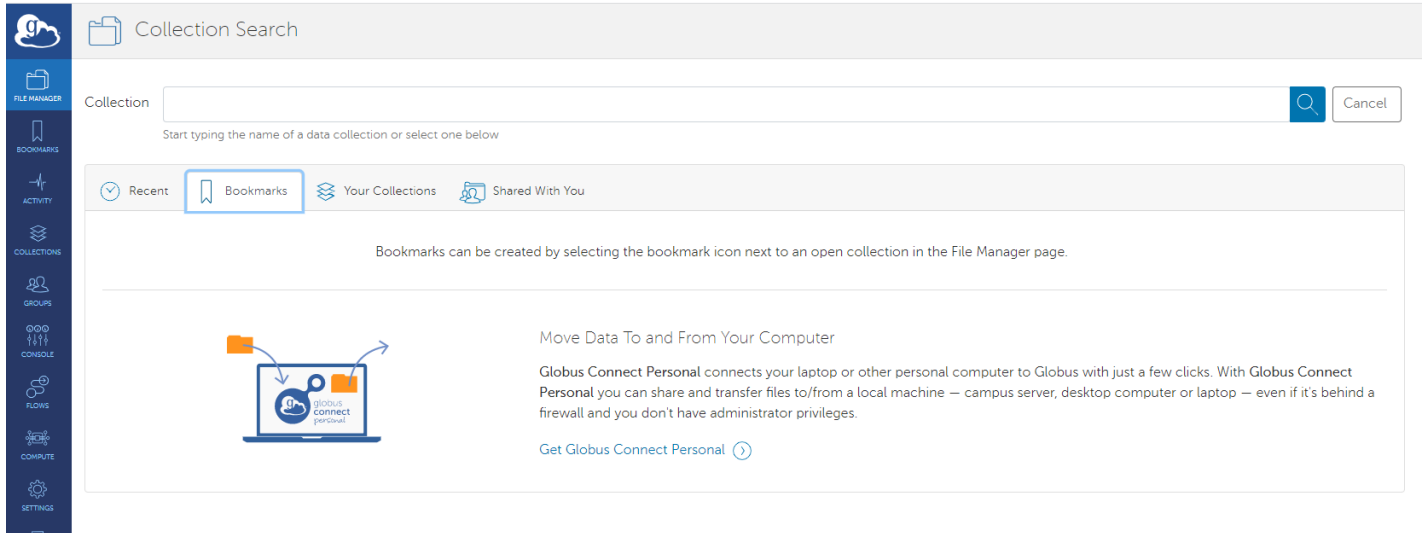


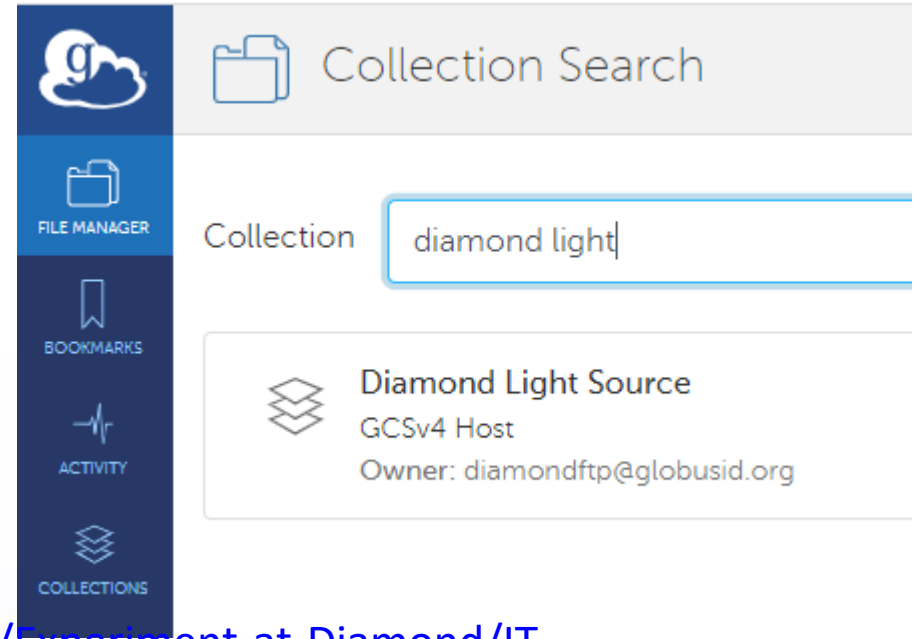
Globus
http://www.globus.org/
Live Data – Raw & Processed

https://www.diamond.ac.uk/Users/Experiment-at-Diamond/IT-User-Guide/Not-at-DLS/Retrieve-data/in40days/Globus.html

Globus

- Globus web portal can be used once logged in to an account to setup transfers
- Search for the Diamond Light Source Endpoint





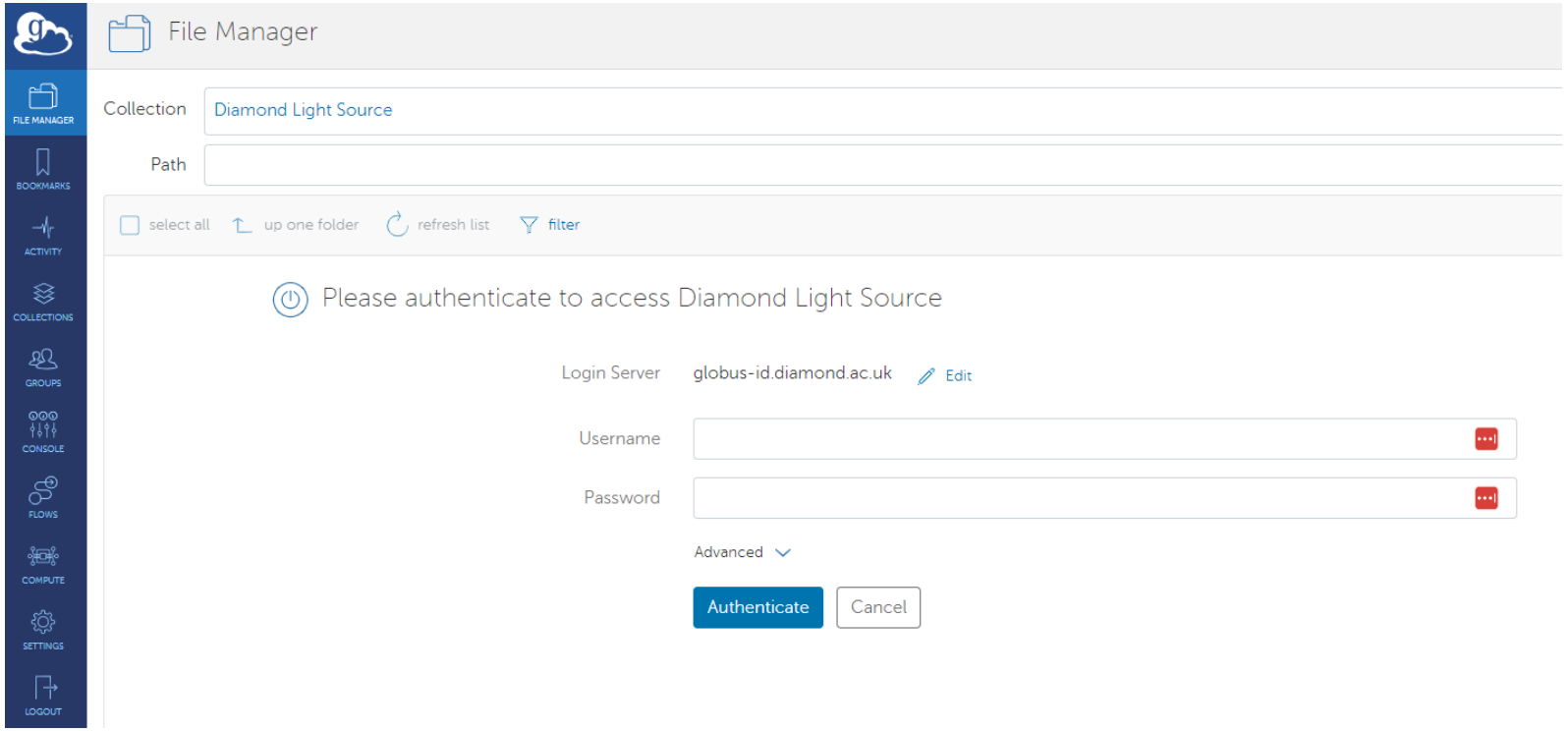
Globus
http://www.globus.org/
Live Data – Raw & Processed

https://www.diamond.ac.uk/Users/Experiment-at-Diamond/IT-User-Guide/Not-at-DLS/Retrieve-data/in40days/Globus.html

diamond

Globus

- Log into the endpoint with diamond fedid and password

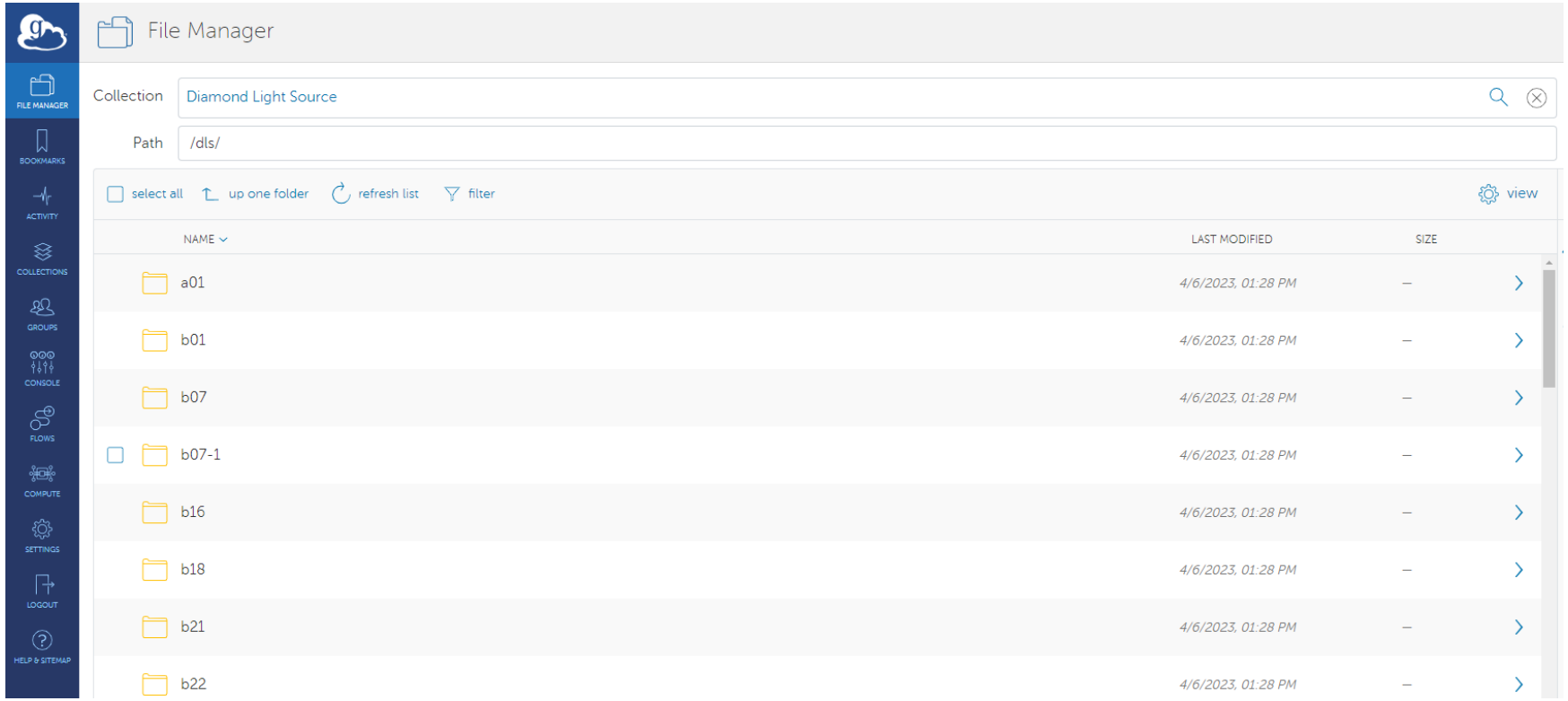


Globus
http://www.globus.org/
Live Data – Raw & Processed

https://www.diamond.ac.uk/Users/Experiment-at-Diamond/IT-User-Guide/Not-at-DLS/Retrieve-data/in40days/Globus.html

Globus

- You can then navigate the live files on disk



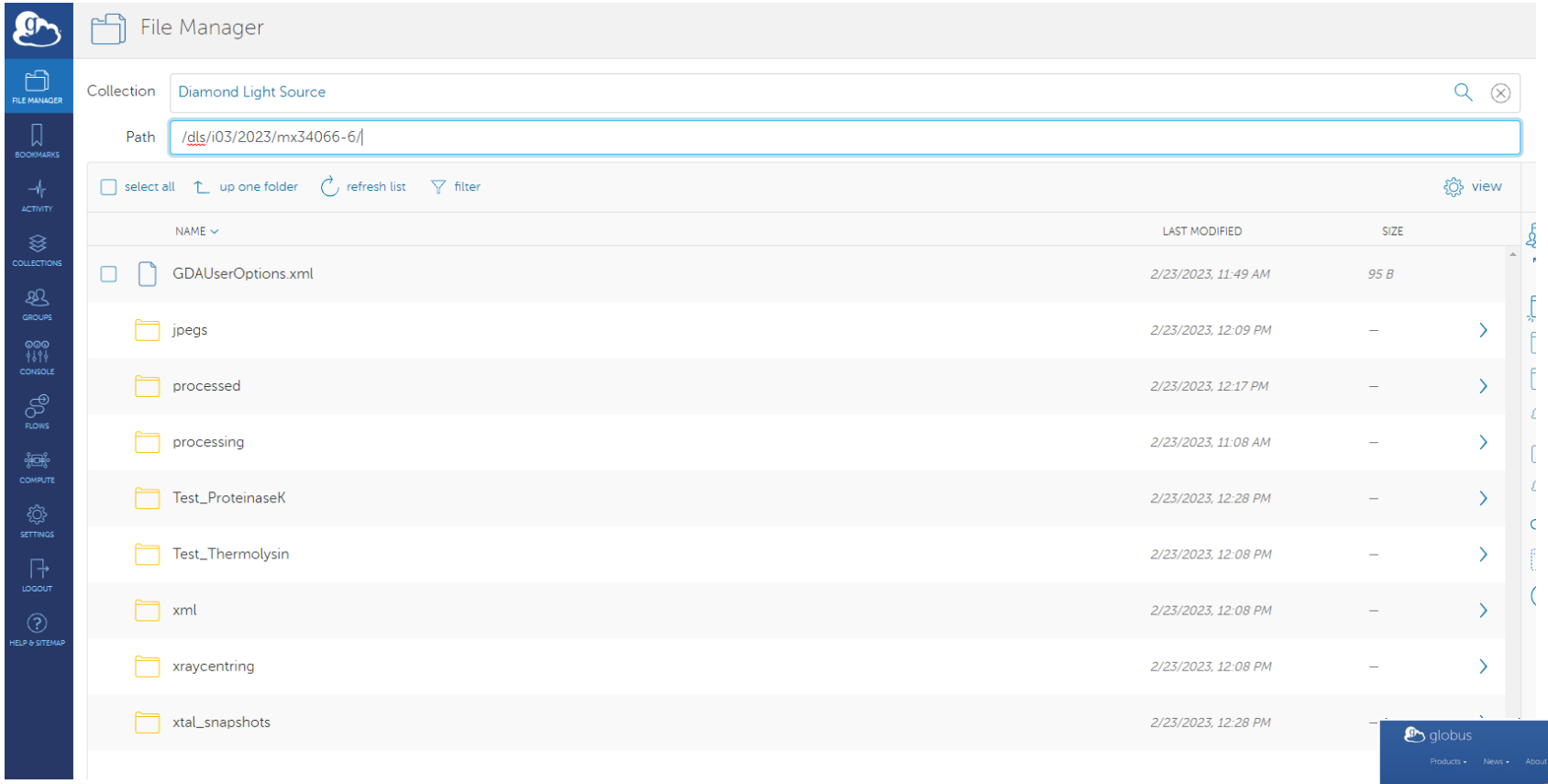
Globus
http://www.globus.org/
Live Data – Raw & Processed

https://www.diamond.ac.uk/Users/Experiment-at-Diamond/IT-User-Guide/Not-at-DLS/Retrieve-data/in40days/Globus.html

diamond

Globus

- You can then navigate the live files on disk



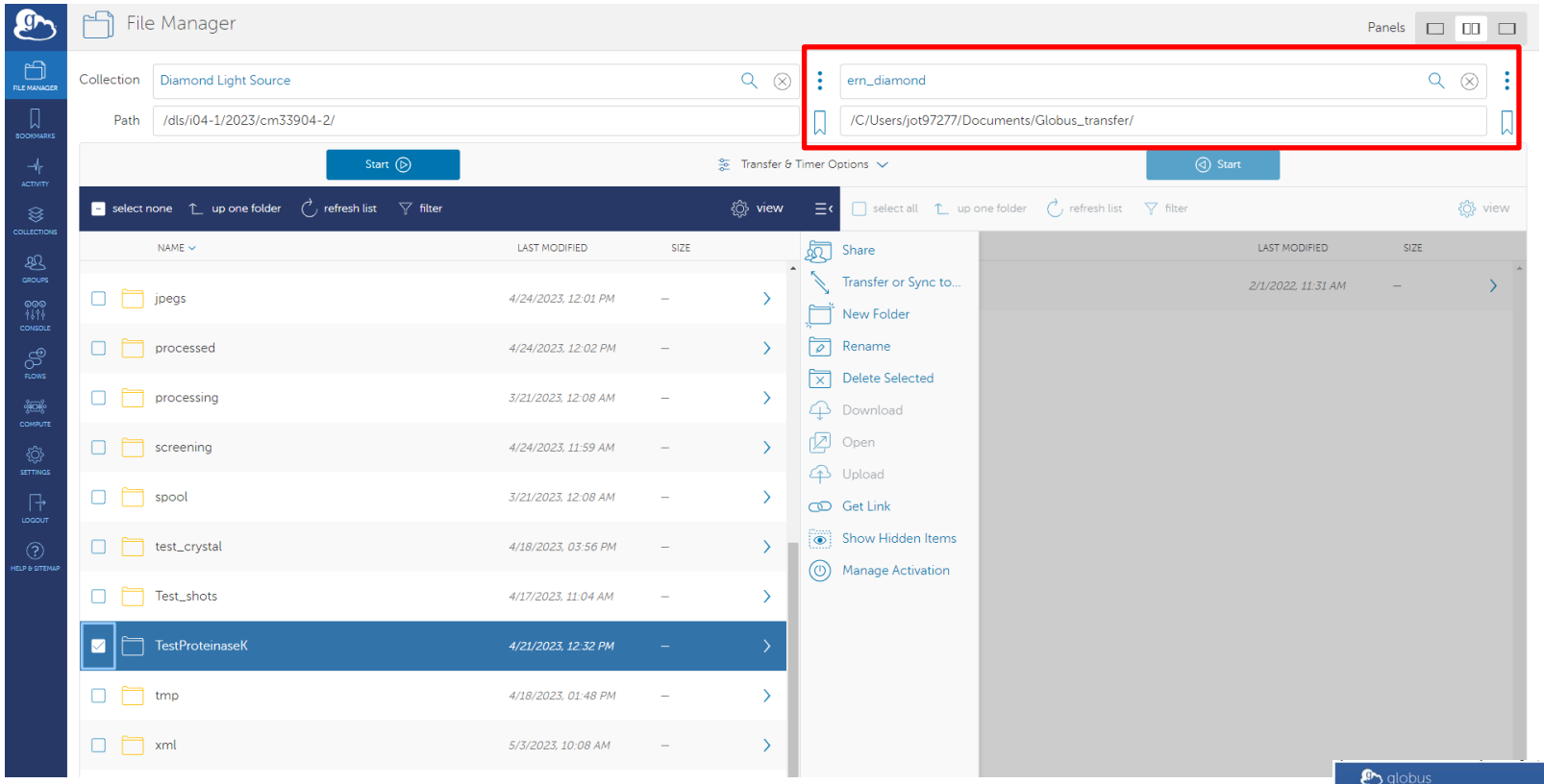
Globus
http://www.globus.org/
Live Data – Raw & Processed

https://www.diamond.ac.uk/Users/Experiment-at-Diamond/IT-User-Guide/Not-at-DLS/Retrieve-data/in40days/Globus.html

diamond

Globus

- Select a personal or institution endpoint to transfer to

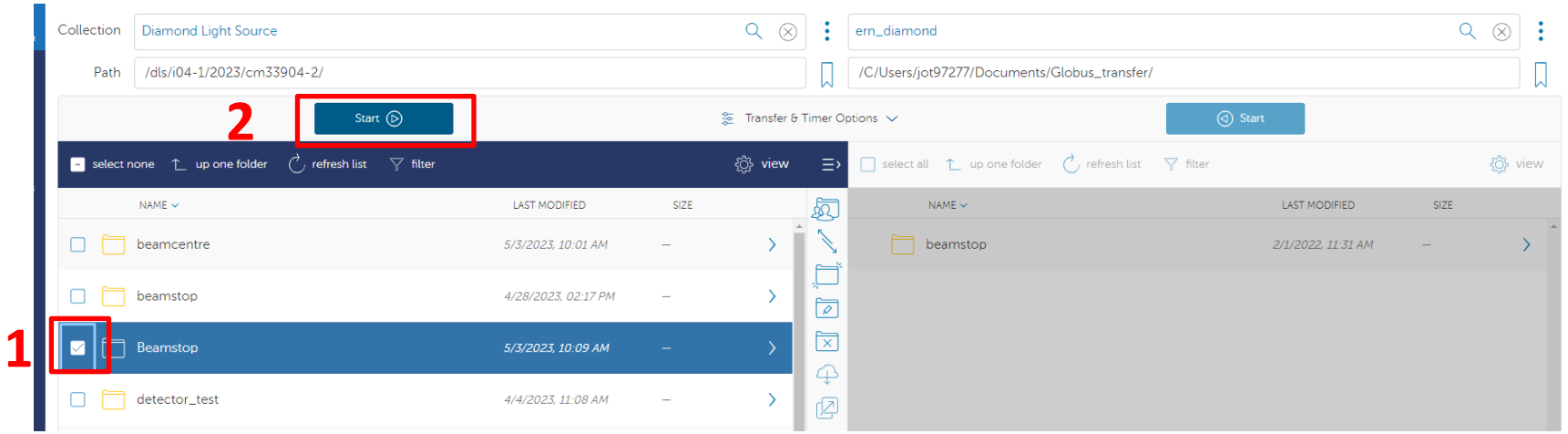


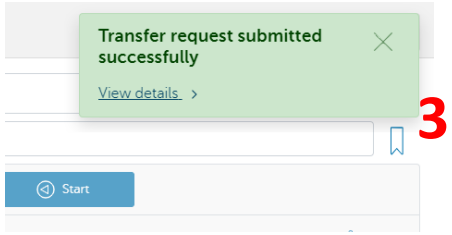
Globus
http://www.globus.org/
Live Data – Raw & Processed

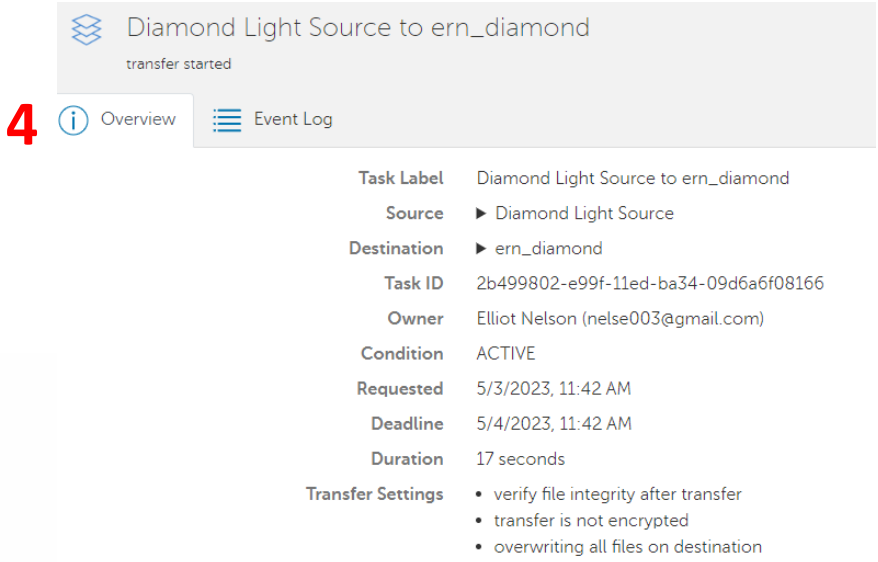
https://www.diamond.ac.uk/Users/Experiment-at-Diamond/IT-User-Guide/Not-at-DLS/Retrieve-data/in40days/Globus.html

Globus

- Once data is selected initiate transfer with start command
- You can review progress of a transfer







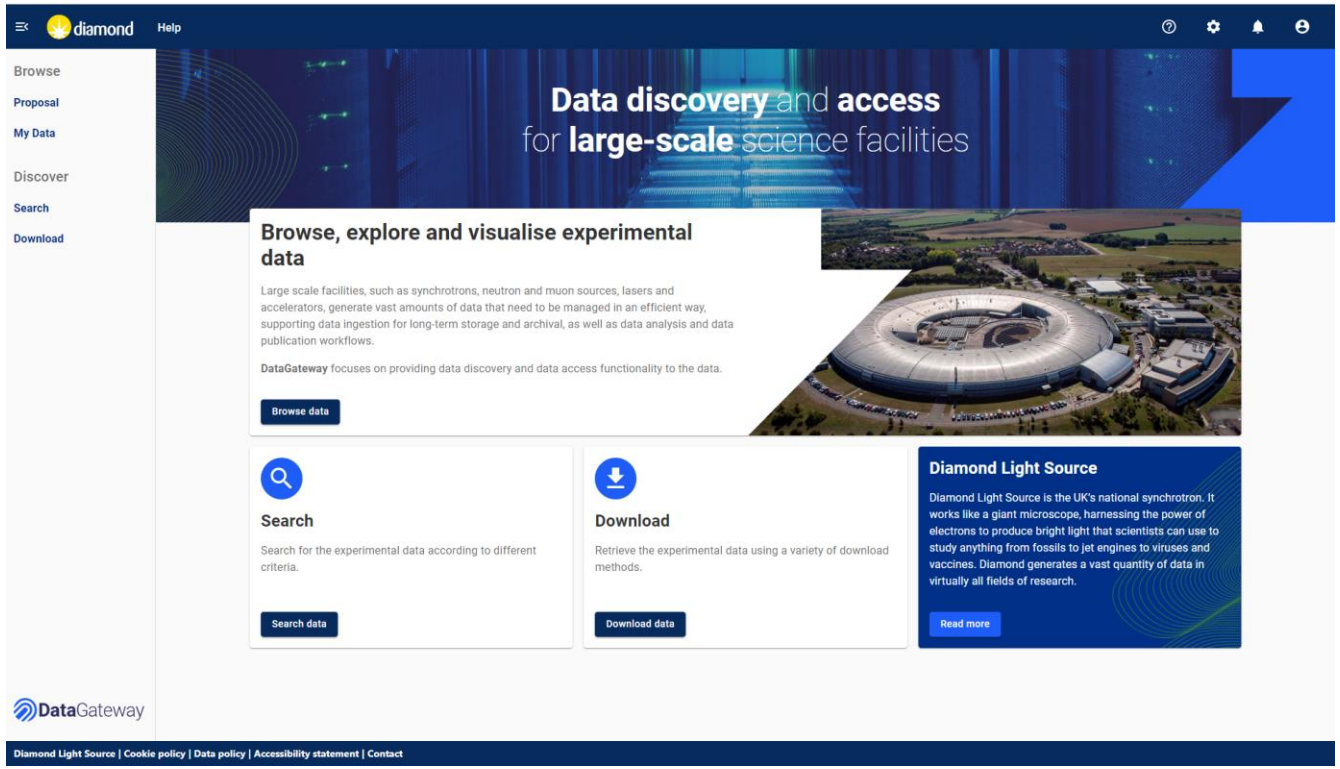
Globus
http://www.globus.org/
Live Data – Raw & Processed

https://www.diamond.ac.uk/Users/Experiment-at-Diamond/IT-User-Guide/Not-at-DLS/Retrieve-data/in40days/Globus.html

diamond

Data Gateway

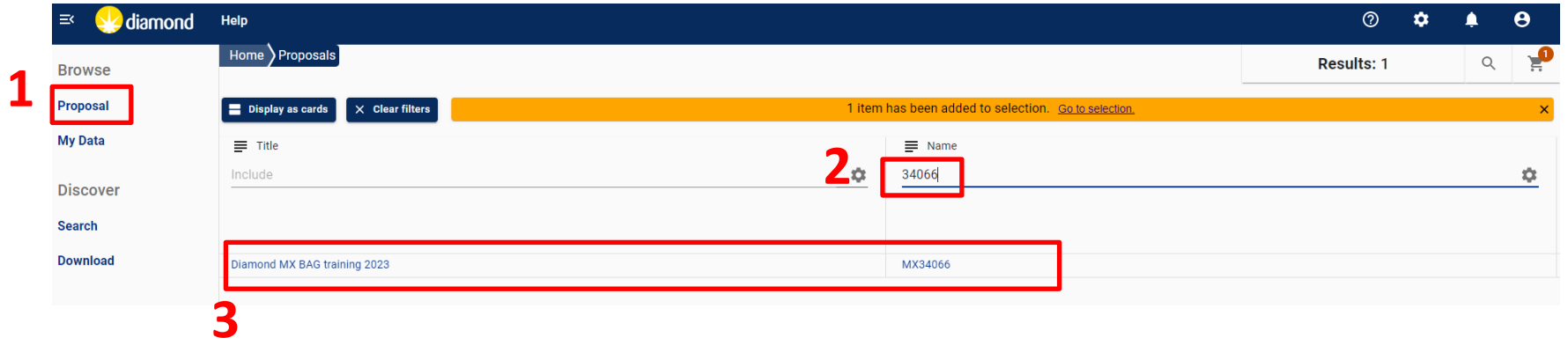
- Data gateway allows access to archived academic data
  - Image data > 40 days old
- Sign in is via Fedid & password

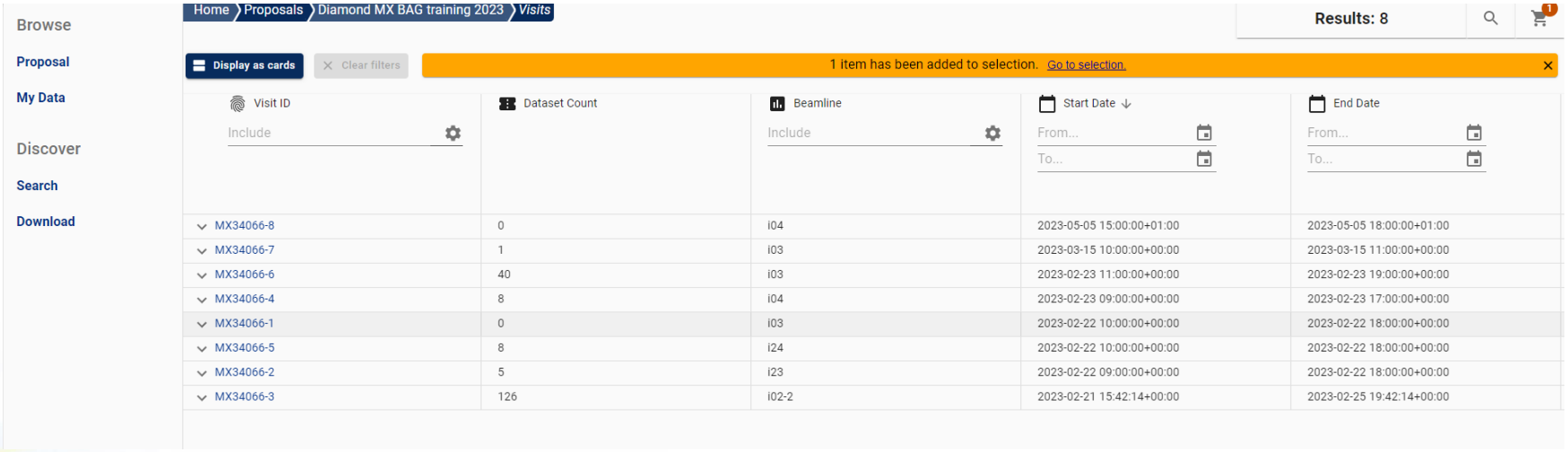


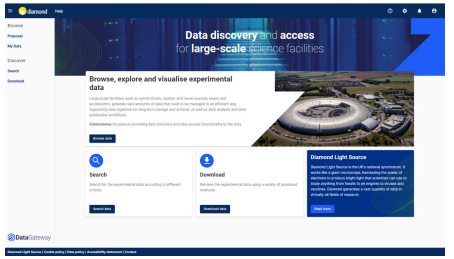
Data Gateway
https://icat.diamond.ac.uk/
Archived Raw Data

Data Gateway

- Select the proposal of interest
- Then visit(s)



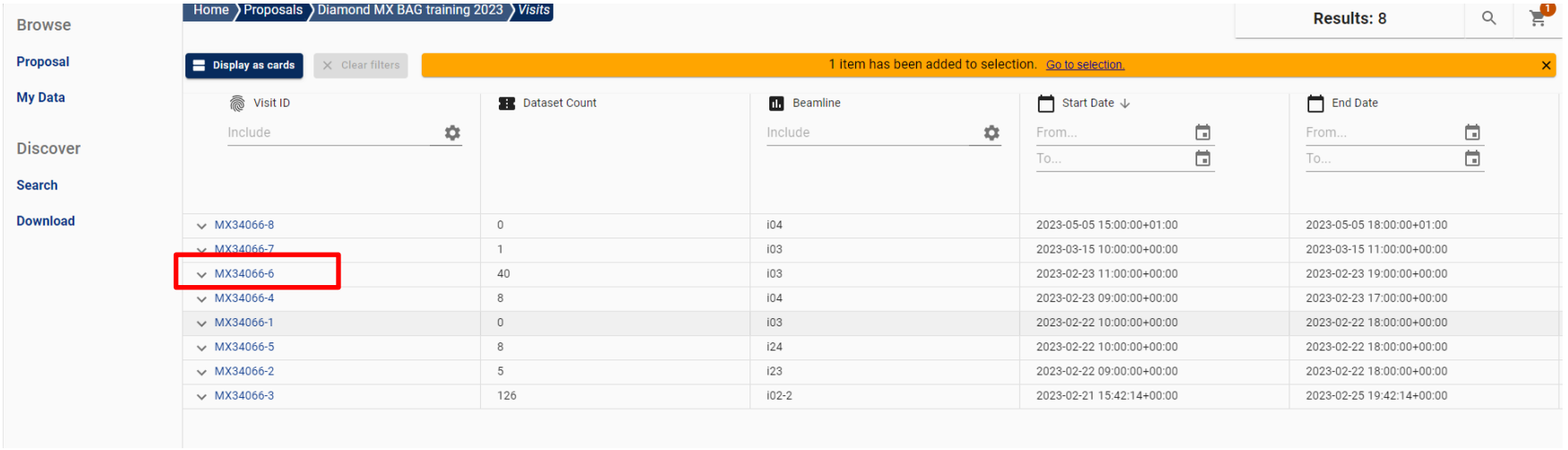


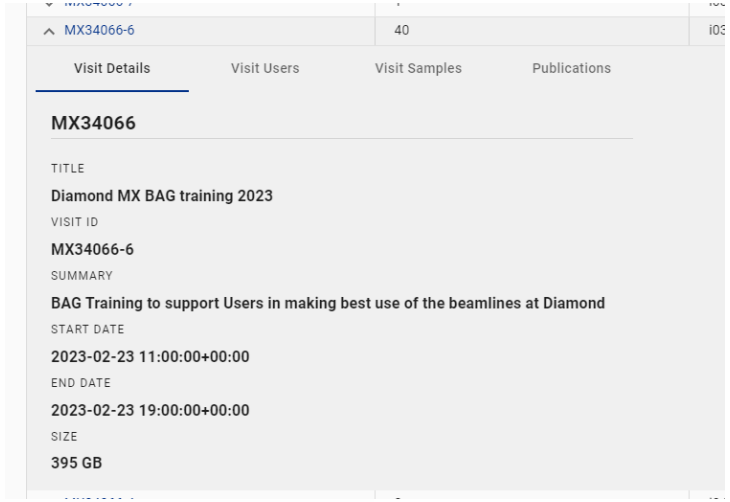


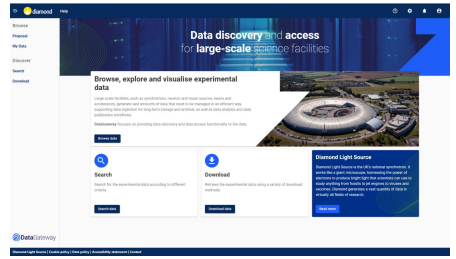
Data Gateway
https://icat.diamond.ac.uk/
Archived Raw Data

Data Gateway

- A summary of the visit size and data is availiable



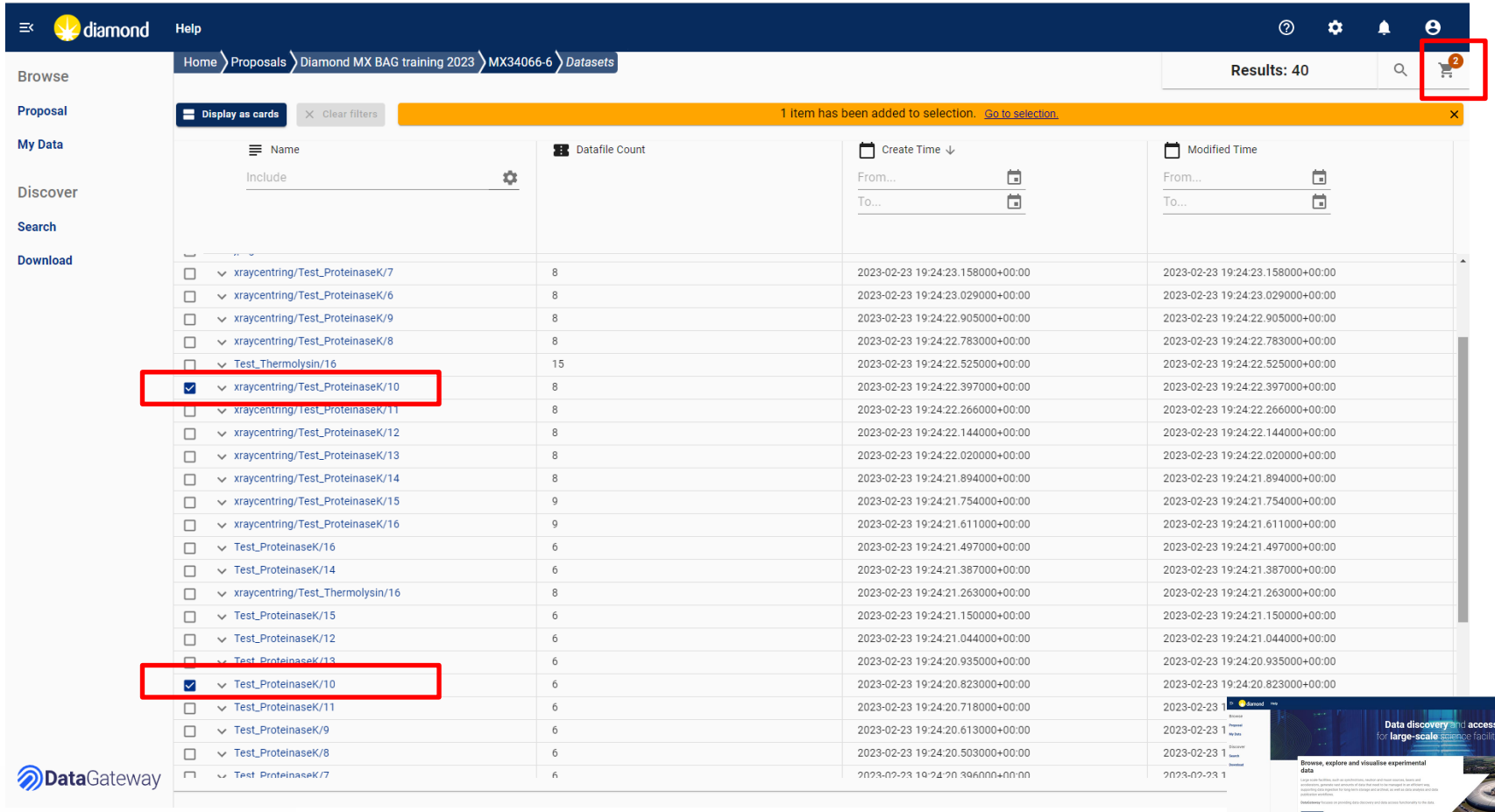


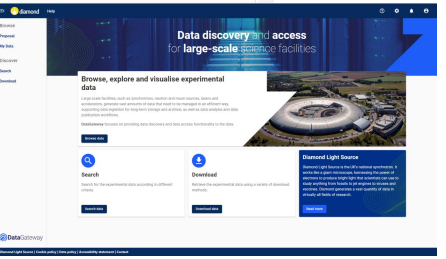


Data Gateway
https://icat.diamond.ac.uk/
Archived Raw Data

Data Gateway

- Within a visit the data is structured as it was on disk
- Hence for a single sample you may want to get centring information, processed data and raw images

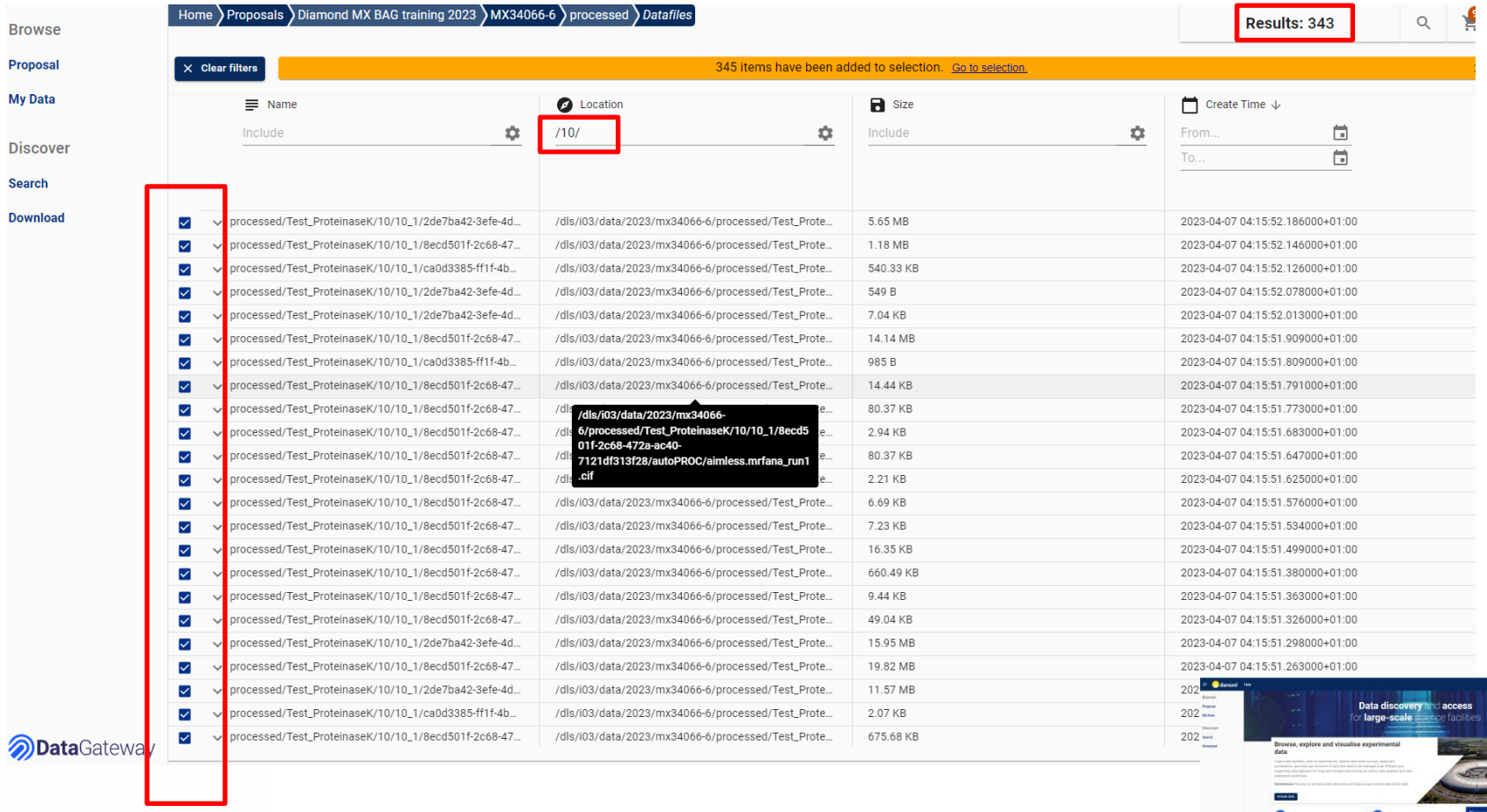




Data Gateway
https://icat.diamond.ac.uk/
Archived Raw Data

Data Gateway

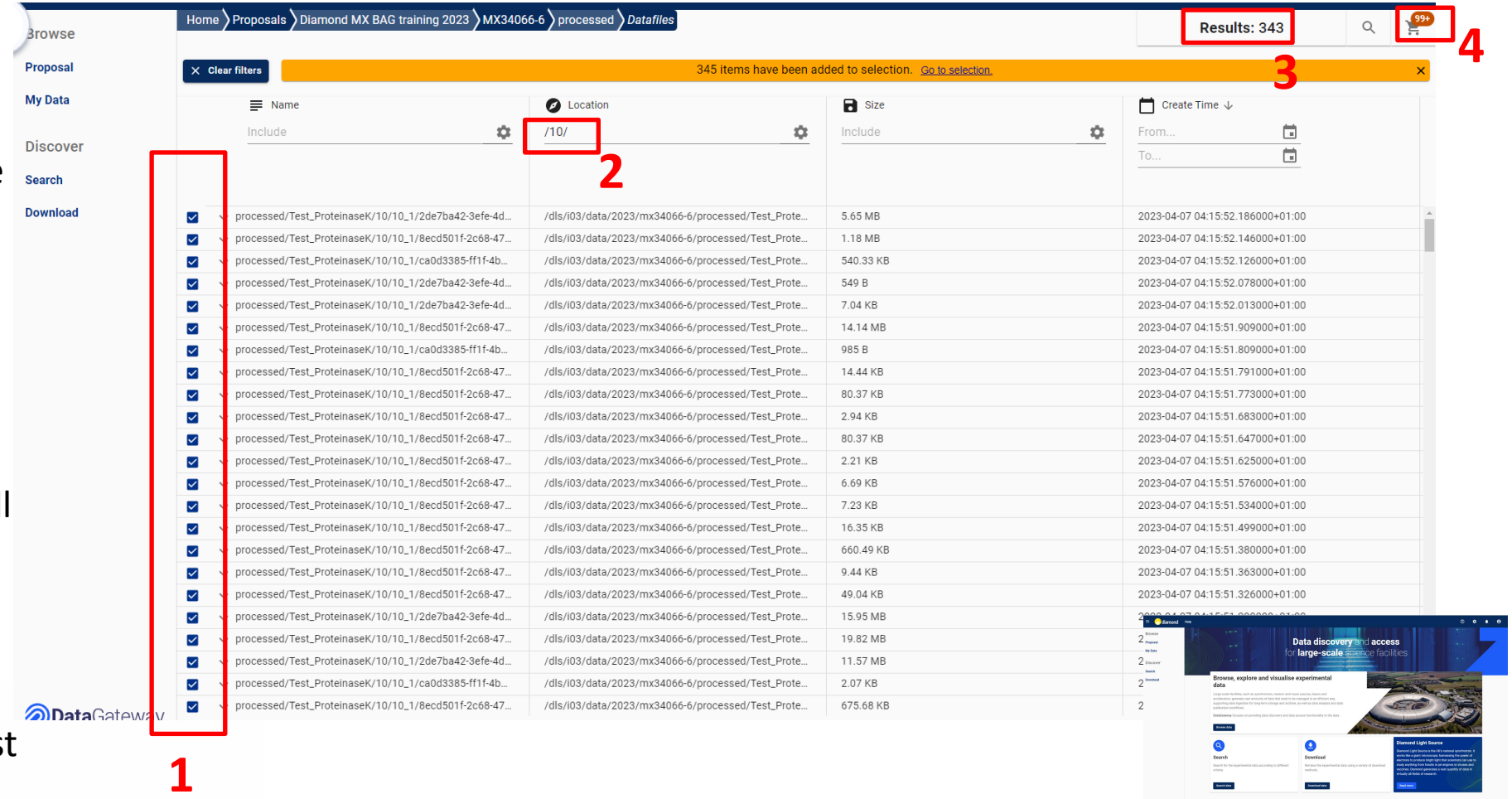
- Within a visit the data is structured as it was on disk
- Hence for a single sample you may want to get centring information, processed data and raw images
- SHIFT can be used to select multiple rows
- Filtering in this case by `/10/` can give just the processed files for the test data

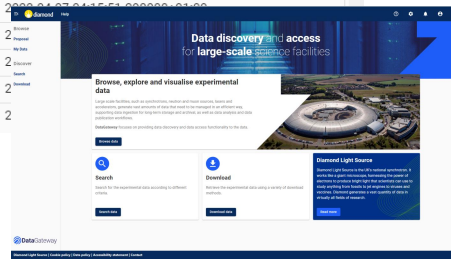


Data Gateway
https://icat.diamond.ac.uk/
Archived Raw Data

Data Gateway

- Within a visit the data is structured as it was on disk
- Hence for a single sample you may want to get centring information, processed data and raw images
- SHIFT can be used to select multiple rows
  - Select top then Scroll to bottom and shift slick to select all
- Filtering in this case by `/10/` can give just the processed files for the test data
- The checkout icon 4 brings you to a summary of items to download

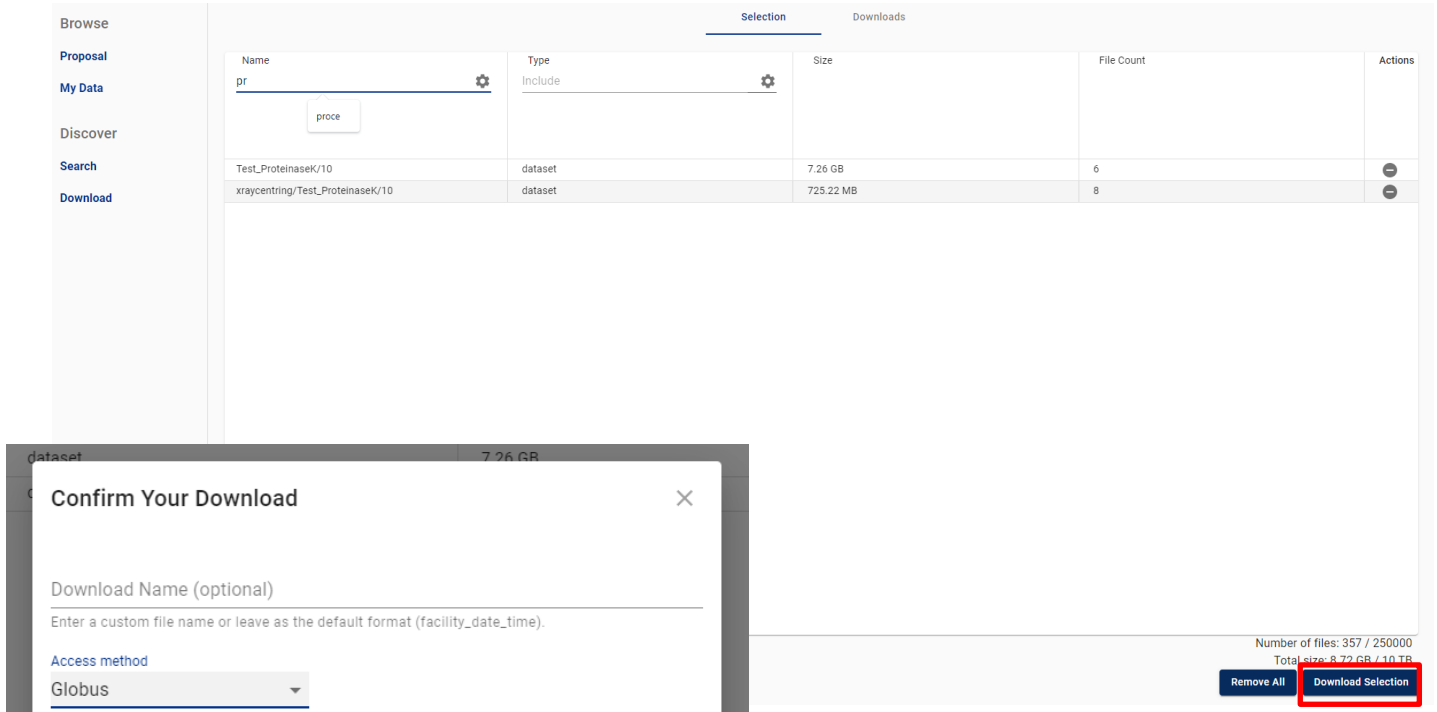


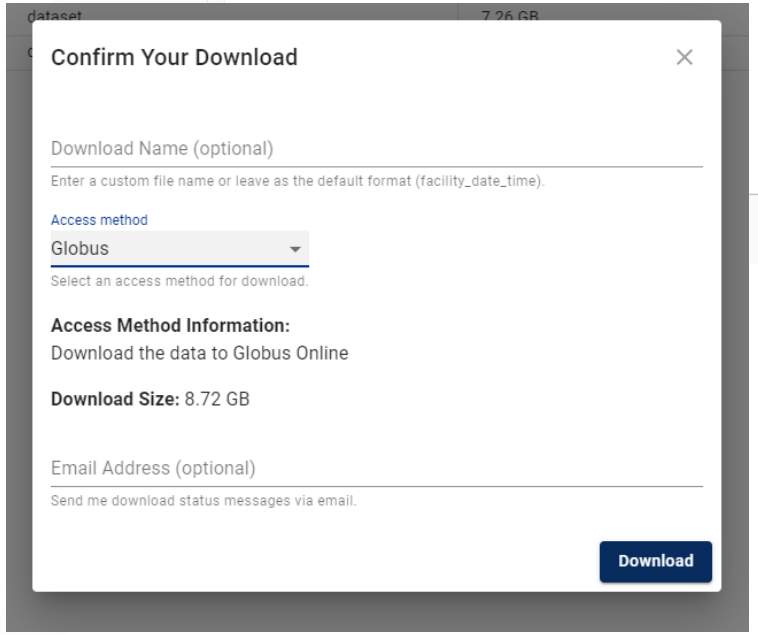


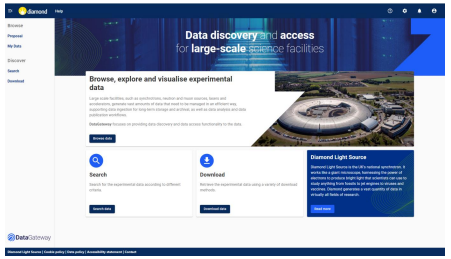
Data Gateway
https://icat.diamond.ac.uk/
Archived Raw Data

Data Gateway

- A summary of the dataset
  - 8.72GB
  - Excluding the processed files by using the gear icon and typing processed
- Up to 10TB can be retrieved at once
- Data can be downloaded, restored to Diamond disk, sent to globus





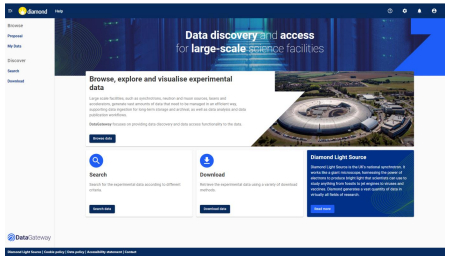


Data Gateway
https://icat.diamond.ac.uk/
Archived Raw Data

Data Gateway

When restored to Diamond disk:

- Users will receive an email when data is available in the DLS file system. Data will be available in one of the following directory structures:
  - **`/dls/staging`**`/dls/$instrument/data/$year/$visit`
  - **`/dls/staging`**`/dls/$village/data/$proposal/$visit`
- The user, beamline staff and visit users will have full write/read access to the directory that has been created at DLS and contains the data requested in DataGateway.
- No further archiving will take place.
- Users will have 15 days to process and transfer their data.
  - After that period, data will be deleted from DLS filesystem.
  - Information on when data will be deleted can be found in README/DATA_EXPIRATION_DATE file in the restored visit directory.
  - It is the responsibility of the user to ensure that data has been transferred before deletion, processed data will not be archived.



Data Gateway
https://icat.diamond.ac.uk/
Archived Raw Data