

Data orientation

Thomas Kluyver

EuXFEL User Meeting, 20th January 2025



Experiment timeline



Before the experiment

- Submit the **beamtime confirmation form**
- Choose an **Experiment data contact** from your team
- Meet with the **Local contact** from the instrument group & **Local data contact** from the data group
 - Discuss what data you will capture & how to process it
- This is only the parts about data!
 - Contact the user office for queries about anything else: useroffice@xfel.eu

https://www.xfel.eu/users/quick_links/index_eng.html

During the experiment

Before

During

After



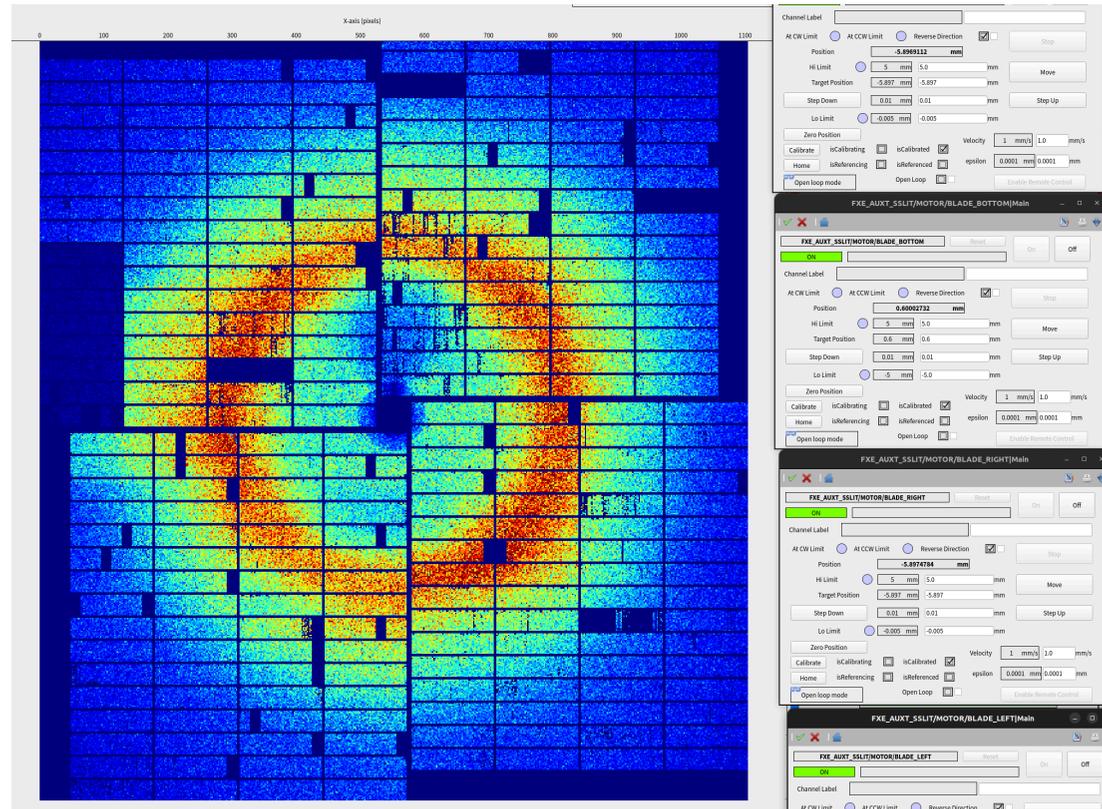
Karabo

The screenshot displays the Karabo GUI interface for controlling a motor. The main window is titled "European XFEL - Karabo GUI 2.21.0 - Topic: FXE - User: kluyvert". The interface is divided into several sections:

- System Topology / Device Topology / Projects:** A tree view on the left shows the system hierarchy. The selected device is "FXE_OGT1_PSLIT/MOTOR/BLADE1_IN_OUT".
- Service Manager:** A panel with a search bar and a "Filter" button.
- Motor Control Panel (FXE_OGT1_PSLIT/MOTOR/BLADE1_IN_OUT):**
 - Status:** A green "ON" indicator.
 - Limits:** Low Limit (HW SW) at 0 mm, High Limit (SW HW) at 51 mm.
 - Actual Position:** 26.629499 mm.
 - Encoder Position:** 26.629499.
 - Step Size:** 2.5 mm.
 - Controller:** encoder.
 - Max. Velocity:** 0.757576.
 - Buttons:** StepDown, StepUp, Move, Stop, On, Off, Reset, Reset Axis.
 - Advanced Mode:** A button to toggle advanced control.
- Parameters Table:**

Parameter	Value
Step Size	2.5 mm
Target Velocity	0.3
Target Position	26.63060009950754 ...
Enable Velocity Update	False
Actual Velocity	-0.006116670616787173
- Log Console:** A bottom panel showing system logs with timestamps and messages.
- Footer:** "fxe-rr-sys-con-gui1:44444 (FXE)"

Karabo



myMdc

The screenshot shows the myMdc web interface for proposal 900466. The browser address bar shows the URL <https://in.xfel.eu/metadata/proposals/1137#propo>. The page title is "Proposal no. 900466". Below the title, there is a status bar with the following information: Status: 2025-01-14 18:24:30 CET, Runs: 639, Calibrations: 279, Team: 17, Size: 105.97 TB (116.51 TB), Files: 11714, Dark Calibrations: 24. The main navigation menu includes: Back, Runs, Beamtime status, General, Public Information, Runs (selected), Logbook, Team, Repositories, Calibration Constants, Remote Devices, Publication, and History. The "Proposal Runs" section is active, showing a table of runs. The table has columns: Run Number (alias), Run type, Sample Name, Techniques, Start date, Run status, Data Assessment, Calibration, Run Comment, and Edit. The table contains three rows of data for runs 0338, 0337, and 0336. A sidebar on the right contains a "HIDE" button and a list of menu items: Proposals, Operational Logbooks, Data Source Groups, Techniques, Home, Statistics, Developers Information, and Legals & About.

Proposal no. 900466

Status: 2025-01-14 18:24:30 CET | Runs: 639 | Calibrations: 279 | Team: 17 | Size: 105.97 TB (116.51 TB) | Files: 11714 | Dark Calibrations: 24

Back | Runs | Beamtime status

General | Public Information | **Runs** | Logbook | Team | Repositories | Calibration Constants | Remote Devices | Publication | History

Proposal Runs

Automatically assess new runs (after being closed by DAQ) as: **To be evaluated manually**

Automatically start run processing after migration: **Yes** (Note: Calibration service will not calibrate runs with run types assessed as "Darks" or "Test experiments" types)

Run Number (alias)	Run type	Sample Name	Techniques	Start date	Run status	Data Assessment	Calibration	Run Comment	Edit
0338	Calibration - Dark HG	Vycor		2024-12-04 16:48:11 +0100	Closed	Run Quality	<input type="checkbox"/>	  	
0337	Calibration - Dark HG	Vycor		2024-12-04 16:46:02 +0100	Closed	Run Quality	<input type="checkbox"/>	  	
0336	Calibration - Dark HG	Vycor		2024-12-03 09:53:28	Closed	Run Quality	<input type="checkbox"/>	  	

<https://in.xfel.eu/metadata/>

myLog

The screenshot displays the myLog interface. On the left is a sidebar with a list of user profiles, each with a lock icon and a name: FXE_5436, FXE_5687, FXE_5725, FXE_6736, FXE_6875, FXE_6936, FXE_7296, FXE_900441_ChristopherMilne (selected), Shift Log, Liquid jet, Controls, Bugs, junk, HELIOS, mono, Laser, more topics, FXE_900491_ChristopherMilne, and ITI AD_000007_M... On the right, the main content area shows two log entries for the user FXE_900441_ChristopherMilne. The first entry, titled 'Liquid jet', is dated 11:50 and contains text about pump operations. The second entry, titled 'Shift Log', is dated NOV 14, 2024, 21:22 and describes a beam alignment session. Below this is an 'EDITED' entry from 21:27 with a link to a screenshot and a thumbnail of the screenshot. At the bottom, a partial entry 'EDITED Save SPR alignment in slot 1' is visible with a timestamp of 22:15.

FXE_900441_ChristopherMilne > Liquid jet 11:50

The sample and nano pumps both can work well with 50um nozzle(water) for more than 6 hours continuesly.
The UV-vis and solvent pumps can run but much slower than they should be.
The UV-vis pump on the top floor can run well with correct flowrate.
Move the pump cart back to hutch.

FXE_900441_ChristopherMilne > Shift Log NOV 14, 2024 21:22

Beam alignment session at 20 keV for RnD project on automated beam alignment tools.
Currently 20 keV, 470 uJ/pulse, 128 pulses/train at 1.1 MHz

EDITED Setting up DAQ, SA1 PBM, p900454, include all relevant data source groups, see below 21:27

[2024-11-14-222624_859324872_exflqr51474.png](#)

EDITED Save SPR alignment in slot 1 22:15

<https://mylog.connect.xfel.eu/>

See poster session

Online analysis options

Analysis devices
in Karabo



EXtra-foam

Metro



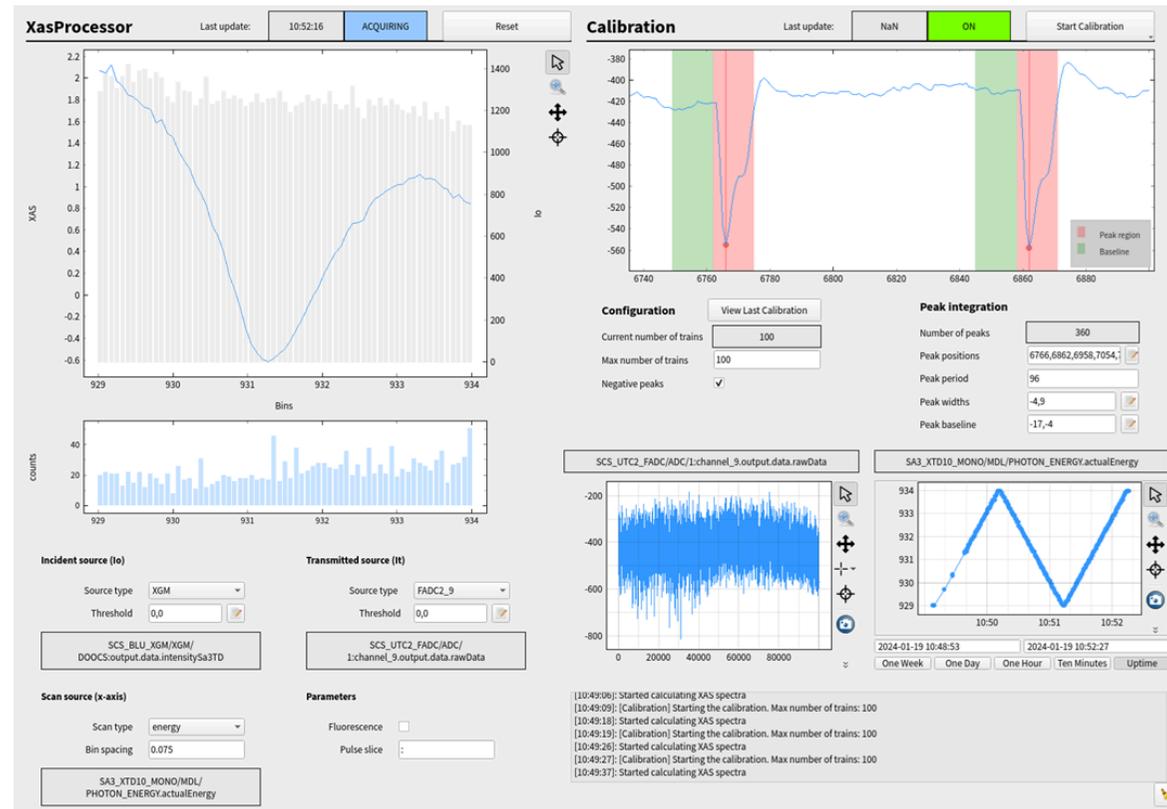
Custom code with
Karabo bridge



Graphical interface
Simple to use

Programmatic interface
Flexible

Analysis in Karabo: XasProcessor



EXtra-metro

EXtra-foam 1.1.0dev - special suite - Correlator

Context +

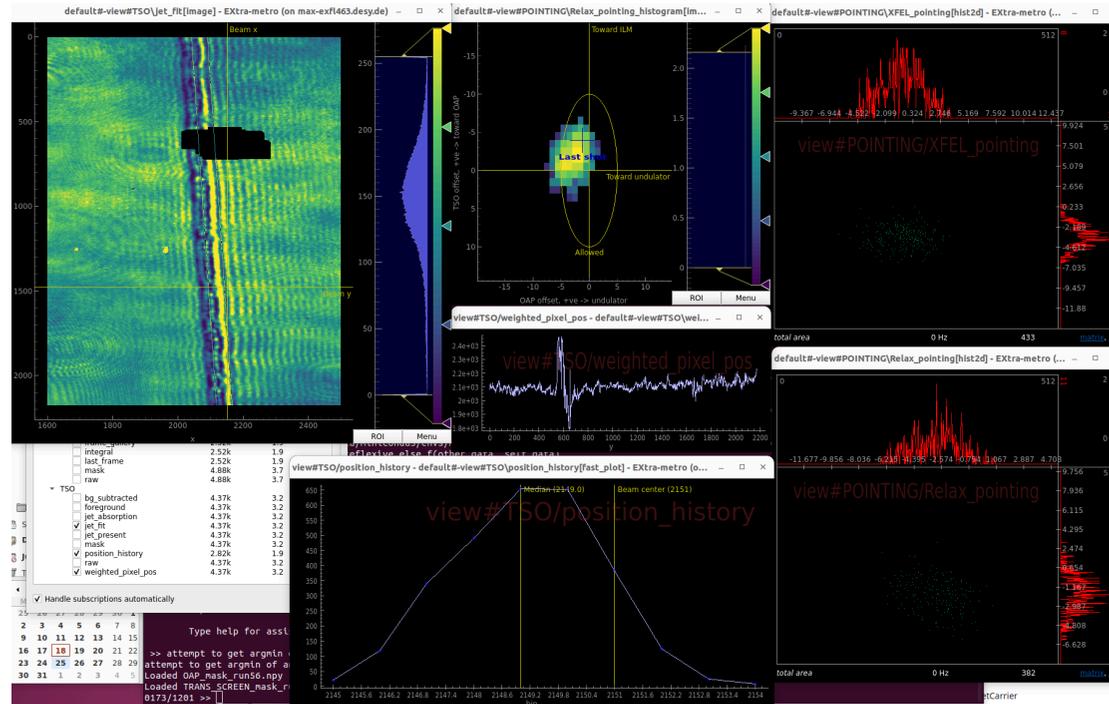
```
1 from scipy.ndimage import gaussian_filter
2
3 @View.Image
4 def smoothed_roi(camera: "karabo#SQS_ILH_LAS/CAM/BS_CAM_TT_OU1:daqOutput[data.image.pixels]"):
5     roi = camera[350:750, 1000:1500]
6     return gaussian_filter(roi, sigma=5)
```

view#smoothed_roi

3000
2000
1000
0

INFO - Reloaded
INFO - Reloaded
INFO - Processing stopped
INFO - Disconnected with tcp://127.0.0.1:45454
INFO - Processing started
INFO - Connected to tcp://127.0.0.1:45454
INFO - Reloaded
INFO - Processing stopped
INFO - Disconnected with tcp://127.0.0.1:45454
INFO - Processing started
INFO - Connected to tcp://127.0.0.1:45454

EXtra-metro



<https://rtd.xfel.eu/docs/metropc/en/latest/>

Karabo bridge

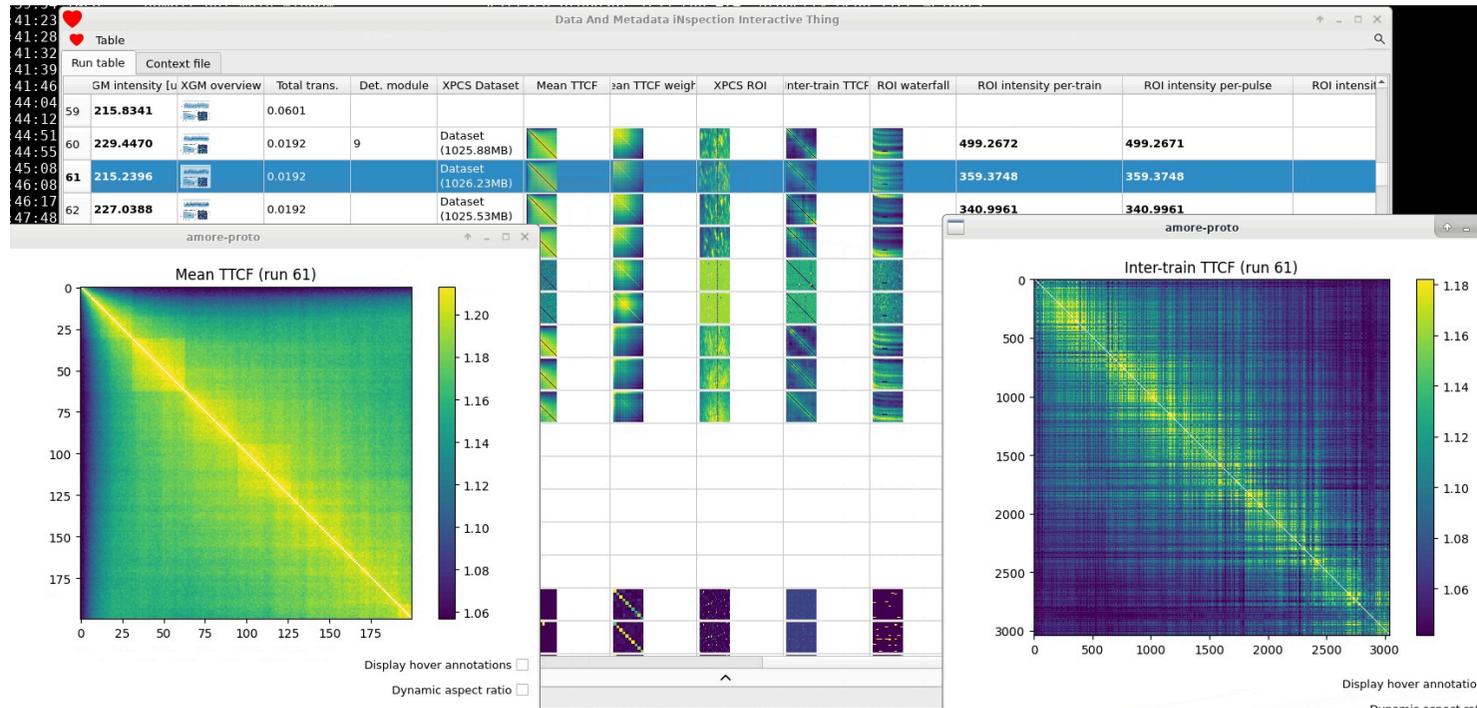
```
from karabo_bridge import Client
bridge = Client("tcp://153.0.55.21:12345")
```

```
data, metadata = bridge.next()
```

```
#           Source                               Property
image = data['FXE_XAD_JF500K/DET/JNGFR03:daqOutput']['data.adc']
```

<https://rtd.xfel.eu/docs/data-analysis-user-documentation/en/latest/online/#streaming-from-karabo-bridge>

Near online analysis: DAMNIT



<https://damnit.readthedocs.io/>

After the experiment

Before



During



After



The offline cluster: Maxwell

Saved data is migrated within minutes to the Maxwell cluster.

You can log into this in a few ways:

■ **Remote desktop** (FastX) for graphical applications

■ <https://max-exfl-display.desy.de:3389/>

■ **JupyterHub** for Python code in notebooks

■ <https://max-jhub.desy.de/>

■ **SSH** for command-line access

■ `ssh myusername@max-exfl-display.desy.de`

The proposal folder

```
$ module load exfel cmdxfel  
$ findxfel 6640  
/gpfs/exfel/exp/FXE/202405/p006640
```

- raw: saved runs
- proc: corrected detector data
- red: put reduced data here
- usr: analysis code, results, plots, etc. (5 TB)
- scratch: temporary intermediate results (will be cleaned up)

Offline analysis: EXtra toolkit

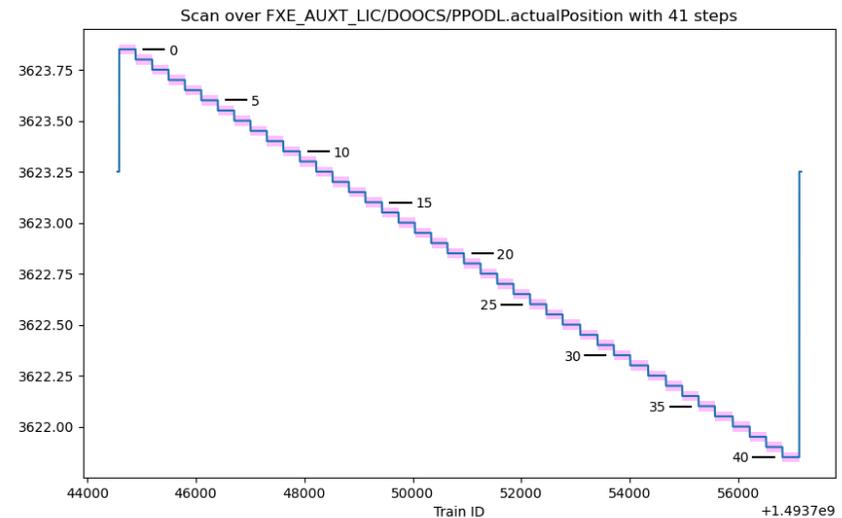
Python libraries to simplify working with EuXFEL data.

```
from extra.data import open_run
from extra.components import Scan

run = open_run(proposal=6060, run=123)
Scan(run.alias['delay']).plot()
```

<https://extra.readthedocs.io/>

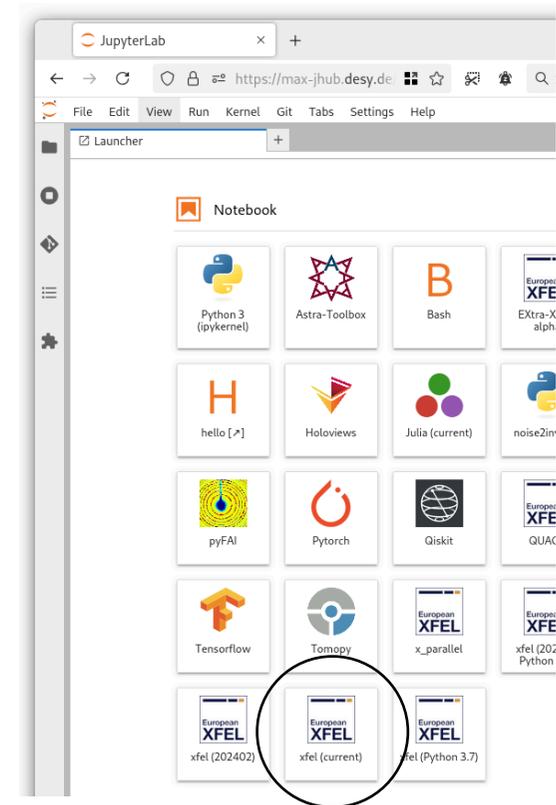
See talk *High-level analysis building blocks*



Offline analysis: Python environments

We provide a Python environment with many libraries preinstalled, created at the start of each cycle.

- In Jupyter: select **xfel (current)** kernel
- Command line: `module load exfel exfel-python`
- To go back to old environments:
 - **xfel (202402)** kernel
 - `module load exfel exfel-python/202402`



Offline analysis: pipelines

 Serial crystallography: EXtra-Xwiz

 Wraps CrystFEL

 XPCS

 SPI

 XAS, XES

Variable coverage

See talk *Automated analysis workflows* & poster session

   **European XFEL**

Using the cluster: Dedicated node on max-jhub

Maxwell Jupyter Job Options

The screenshot shows the 'Maxwell partitions' dropdown menu with the following options:

- node on JHUB partition (selected)
- node on ALLCPU partition
- node on ALLGPU partition
- node on EXFEL partition
- node on UPEX partition

Annotations point to these options:

- 'node on JHUB partition' is annotated as 'Shared node, 1 week session'.
- 'node on EXFEL partition' is annotated as 'Dedicated node (EuXFEL staff)'.
- 'node on UPEX partition' is annotated as 'Dedicated node (users)'.

Additional text on the right side of the image states: '8 hour sessions' and 'May be pre-empted for higher priority jobs'.

Node and GPU availability						
Partition	# nodes	# avail	# GPUs avail	# P100 avail	# V100 avail	# A100 avail
jhub	4	4	0	0	0	0
allcpu	480	240	0	0	0	0
allgpu	166	8	8	1	7	0
exfel	354	114	0	0	0	0

Using the cluster: Batch jobs with Slurm

Use **Slurm** to make full use of the processing power on the cluster.

E.g. to run a script:

```
sbatch -p upex -t 8:00:00 myscript.sh
```

To see your running & queued jobs:

```
squeue --me
```

```
$ squeue --me
```

JOBID	PARTITION	NAME	USER	ST	TIME	NODES	NODELIST(REASON)
12498410	exfel	wrap	kluyvert	PD	0:00	1	(Priority)

Embargoes, data retention & data reduction

New scientific data policy will apply for proposals submitted in 2025 & beyond.

■ Experimental data will be reduced for long term storage

■ 50 - 500 TB: reduce to 50 TB

■ > 500 TB: reduce to 10%

■ Un-reduced data will be archived after 3 - 6 months

■ Data becomes open after embargo period of 3 years (or earlier on request)

■ EuXFEL will help with the processes of reducing & opening data

See talks *Data management plans* & *Data reduction tools*, and the poster session

Resources

- Presentations from this meeting: <https://indico.desy.de/event/47497/overview>
- Last year's meeting (inc. tutorials): <https://indico.desy.de/event/42904/>
- Data analysis docs: <https://rtd.xfel.eu/docs/data-analysis-user-documentation/en/latest/>
- DESY Maxwell docs: <https://docs.desy.de/maxwell/>
- Contact Data Analysis: da@xfel.eu