Data orientation

Thomas Kluyver

EuXFEL User Meeting, 20th January 2025

Experiment timeline

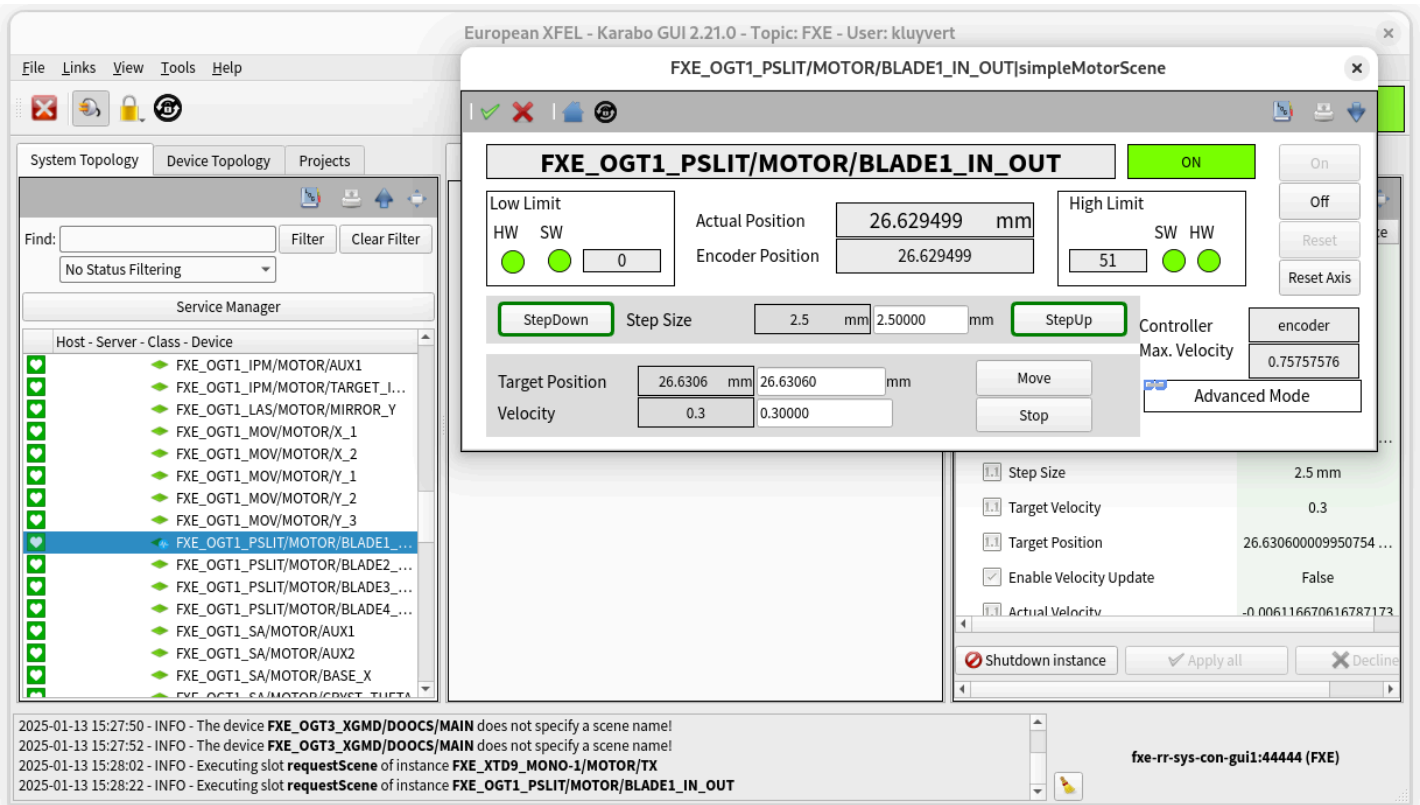
Before the experiment

- 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`

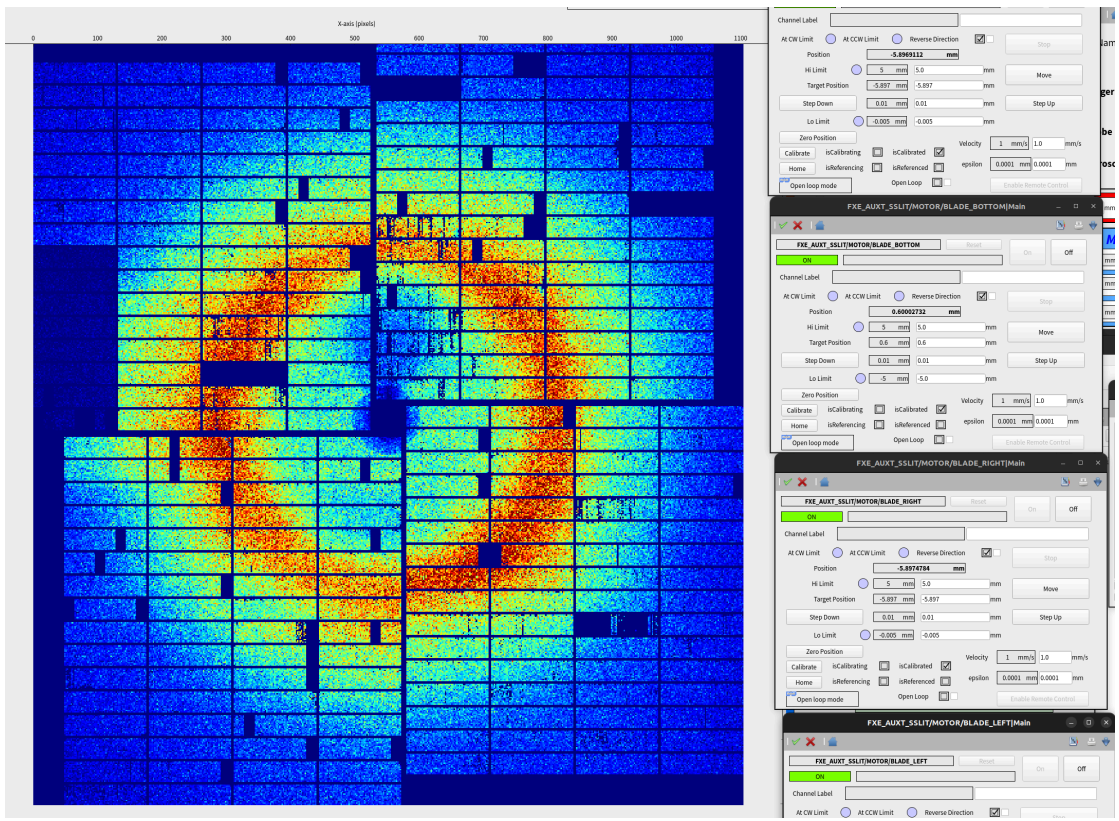
https://www.xfel.eu/users/quick_links/index_eng.html

During the experiment

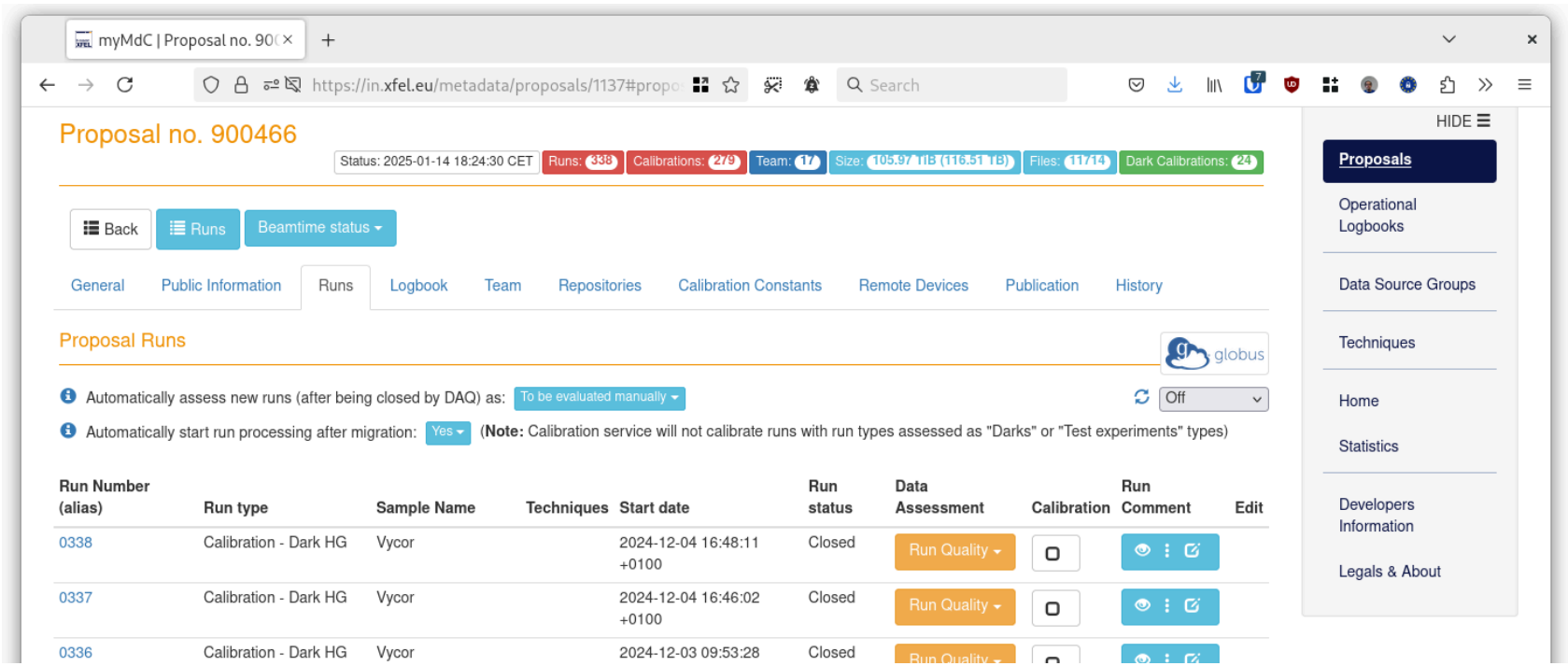
Karabo


European XFEL - Karabo GUI 2.21.0 - Topic: FXE - User: kluyvert
File Links View Tools Help
System Topology
Device Topology
Projects
Find:
Filter
Clear Filter
No Status Filtering
Service Manager
Host - Server - Class - Device
FXE_OGT1_IPM/MOTOR/AUX1
FXE_OGT1_IPM/MOTOR/TARGET_I...
FXE_OGT1_LAS/MOTOR/MIRROR_Y
FXE_OGT1_MOV/MOTOR/X_1
FXE_OGT1_MOV/MOTOR/X_2
FXE_OGT1_MOV/MOTOR/Y_1
FXE_OGT1_MOV/MOTOR/Y_2
FXE_OGT1_MOV/MOTOR/Y_3
FXE_OGT1_PSLIT/MOTOR/BLADE1_...
FXE_OGT1_PSLIT/MOTOR/BLADE2_...
FXE_OGT1_PSLIT/MOTOR/BLADE3_...
FXE_OGT1_PSLIT/MOTOR/BLADE4_...
FXE_OGT1_SA/MOTOR/AUX1
FXE_OGT1_SA/MOTOR/AUX2
FXE_OGT1_SA/MOTOR/BASE_X
FXE_OGT1_PSLIT/MOTOR/BLADE1_IN_OUT|simpleMotorScene
FXE_OGT1_PSLIT/MOTOR/BLADE1_IN_OUT
ON
On
Off
Reset
Reset Axis
Low Limit
HW SW
0
Actual Position
26.629499 mm
Encoder Position
26.629499
High Limit
SW HW
51
StepDown
Step Size
2.5 mm
2.50000 mm
StepUp
Controller
encoder
Max. Velocity
0.75757576
Target Position
26.6306 mm
26.63060 mm
Move
Velocity
0.3
0.30000
Stop
Advanced Mode
Step Size
2.5 mm
Target Velocity
0.3
Target Position
26.630600009950754 ...
Enable Velocity Update
False
Actual Velocity
-0.006116670616787173
Shutdown instance
Apply all
Decline
2025-01-13 15:27:50 - INFO - The device FXE_OGT3_XGMD/DOOCS/MAIN does not specify a scene name!
2025-01-13 15:27:52 - INFO - The device FXE_OGT3_XGMD/DOOCS/MAIN does not specify a scene name!
2025-01-13 15:28:02 - INFO - Executing slot requestScene of instance FXE_XTD9_MONO-1/MOTOR/TX
2025-01-13 15:28:22 - INFO - Executing slot requestScene of instance FXE_OGT1_PSLIT/MOTOR/BLADE1_IN_OUT
fxe-rr-sys-con-gui1:44444 (FXE)

Karabo

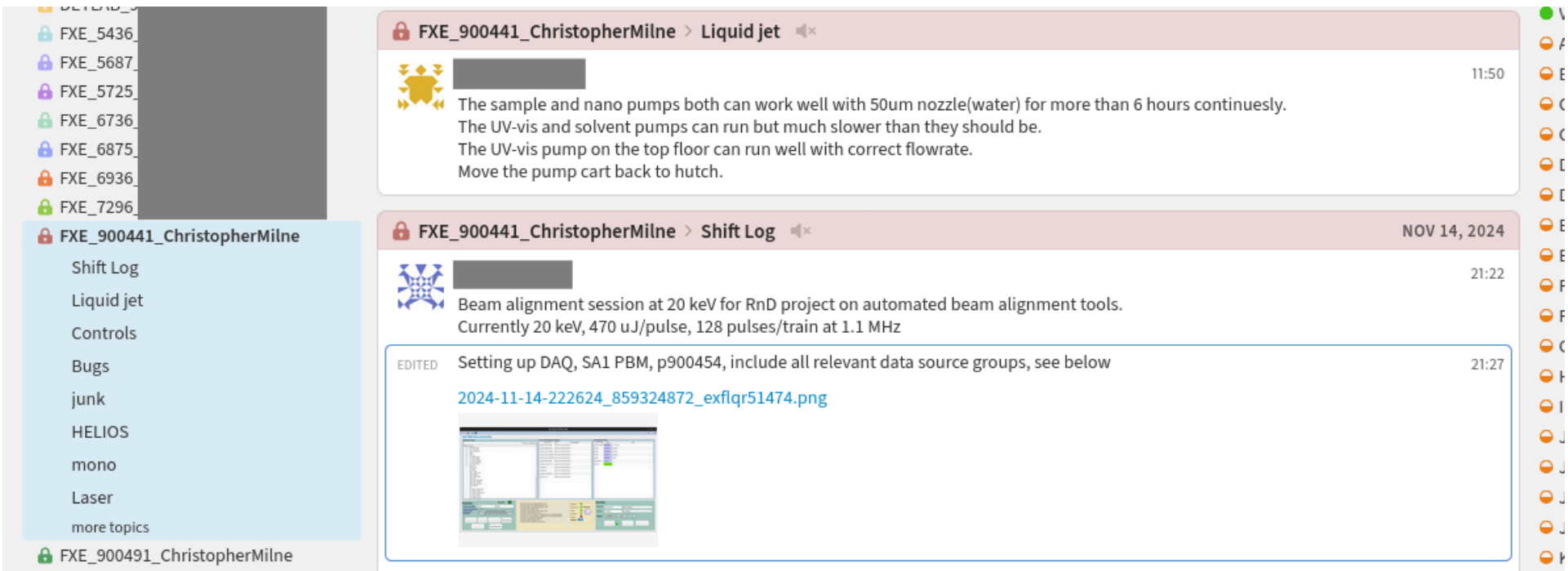

X-axis [pixels]
FXE_AUXT_SSLIT/MOTOR/BLADE_BOTTOM|Main
FXE_AUXT_SSLIT/MOTOR/BLADE_RIGHT|Main
FXE_AUXT_SSLIT/MOTOR/BLADE_LEFT|Main

myMdC



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

myLog



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

See poster session

Online analysis options

Analysis devices in Karabo

EXtra-foam

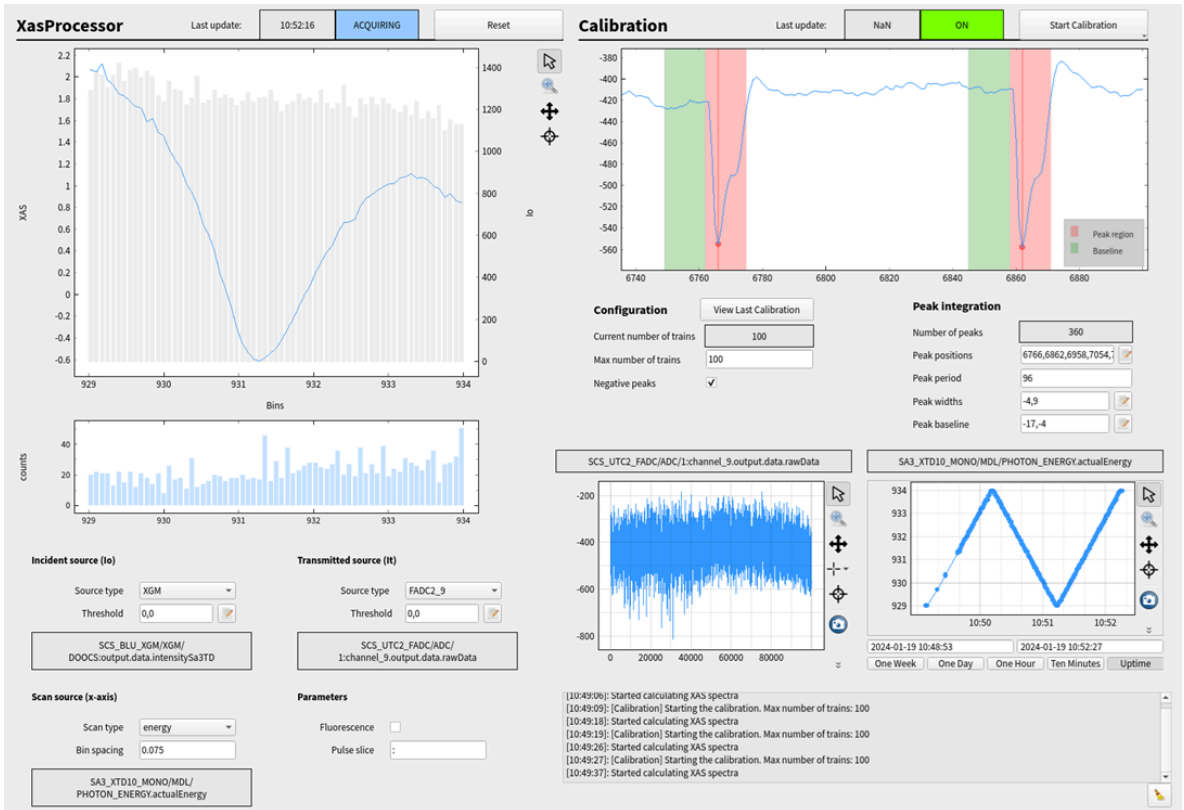
Metro

Custom code with Karabo bridge

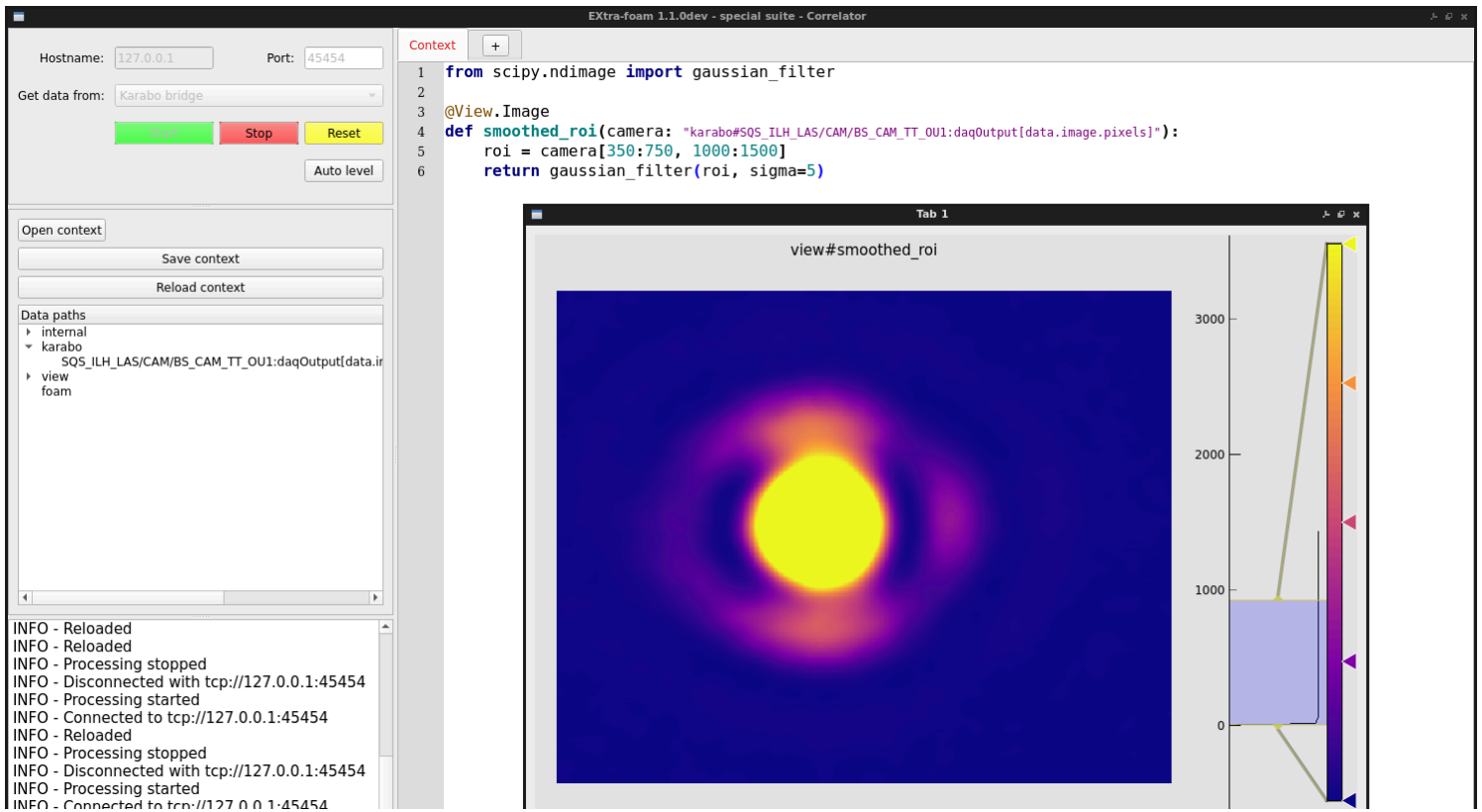
Graphical interface
Simple to use

Programmatic interface
Flexible

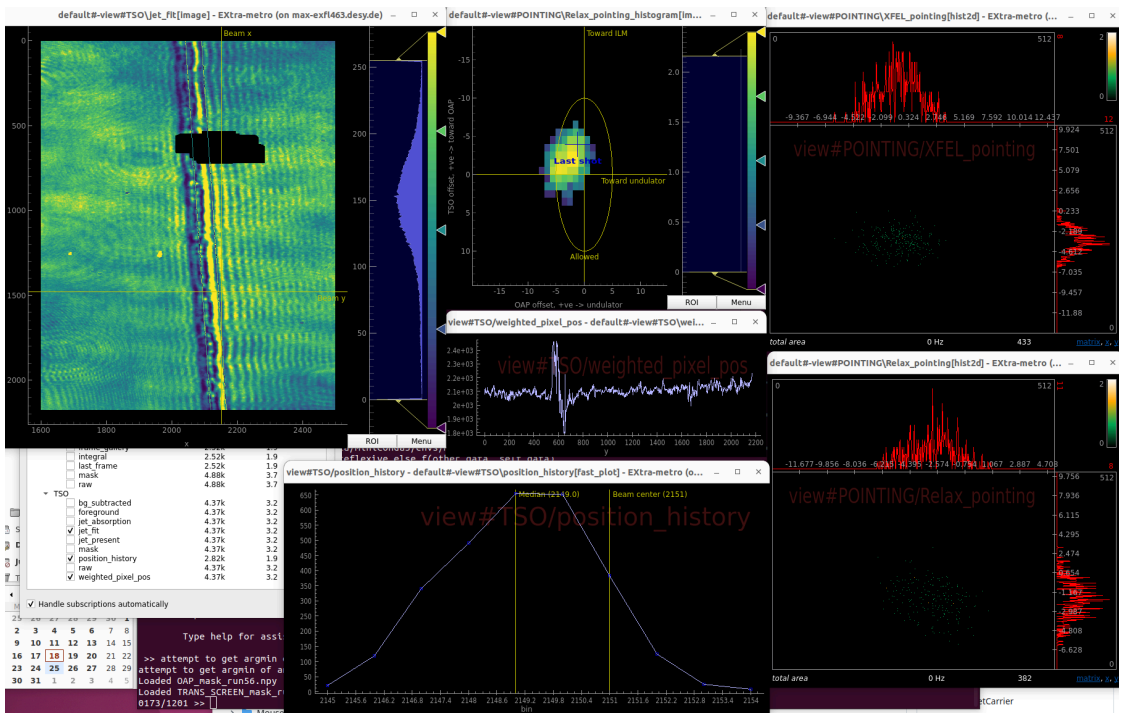
Analysis in Karabo: XasProcessor



EXtra-metro


EXtra-foam 1.1.0dev - special suite - Correlator
Hostname: 127.0.0.1
Port: 45454
Get data from: Karabo bridge
Start
Stop
Reset
Auto level
Open context
Save context
Reload context
Data paths
internal
karabo
SQS_ILH_LAS/CAM/BS_CAM_TT_OU1:daqOutput[data.ir
view
foam
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
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)
Tab 1
view#smoothed_roi
3000
2000
1000
0

EXtra-metro



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

Karabo bridge

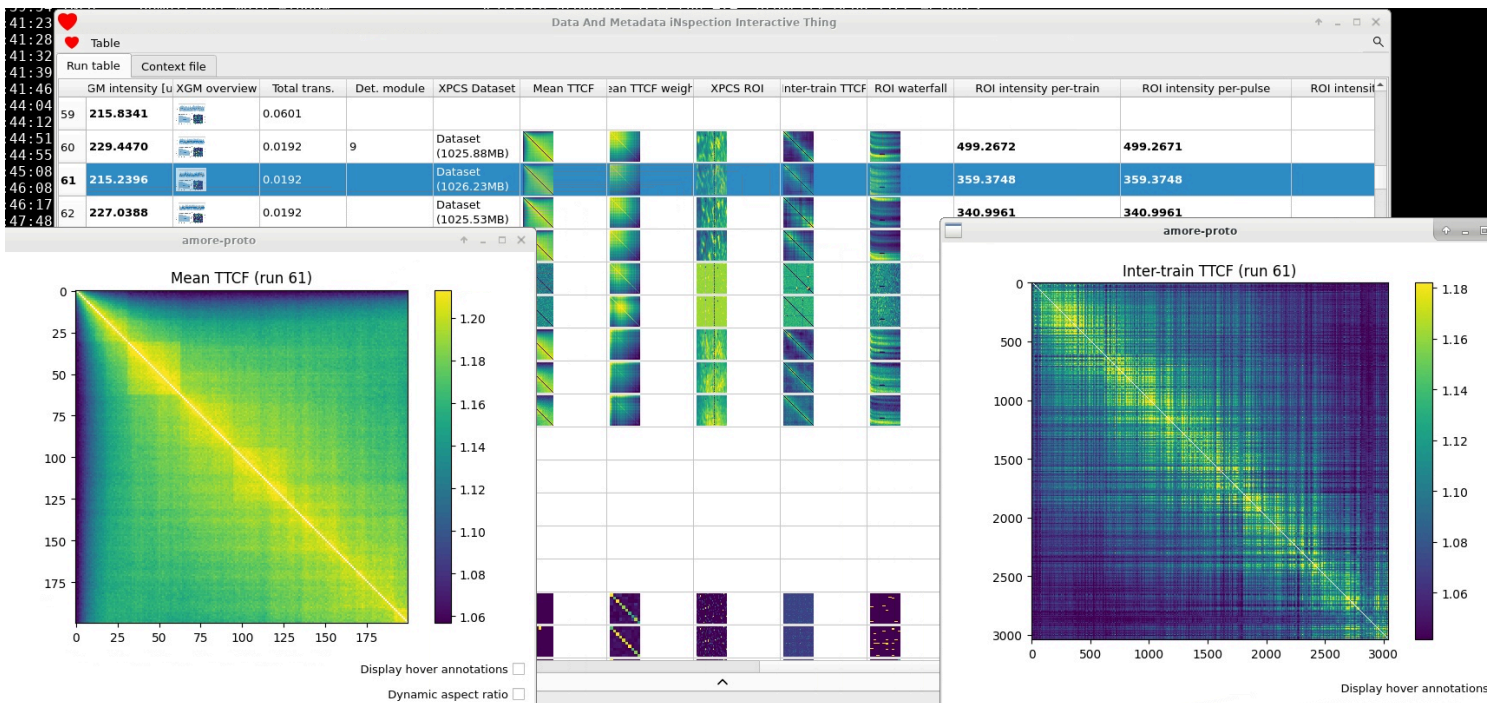
```
from karabo_bridge import Client
bridge = Client("tcp://153.0.55.21:12345")

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

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
```

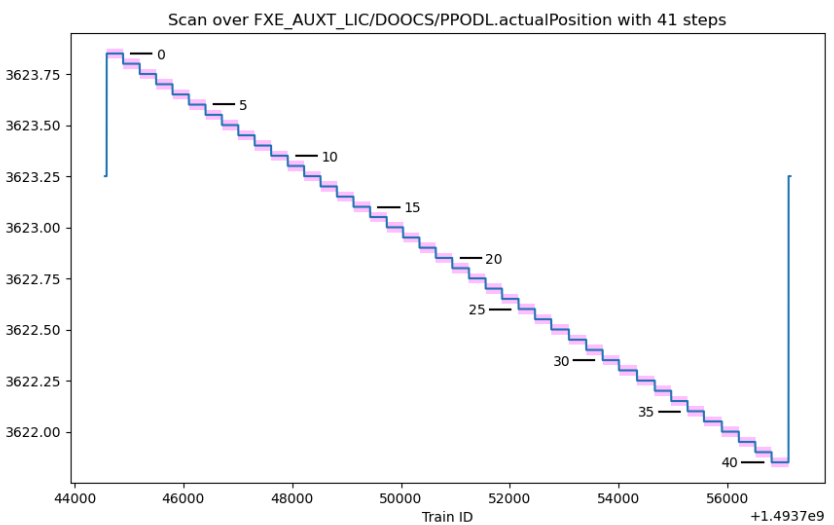
- `raw`: saved runs
- `proc`: corrected detector data
- `red`: put reduced data here
- `usr`: analysis code, results, plots, etc. (5 TB)
- `scratch`: temporary intermediate results (will be cleaned up)

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()
```



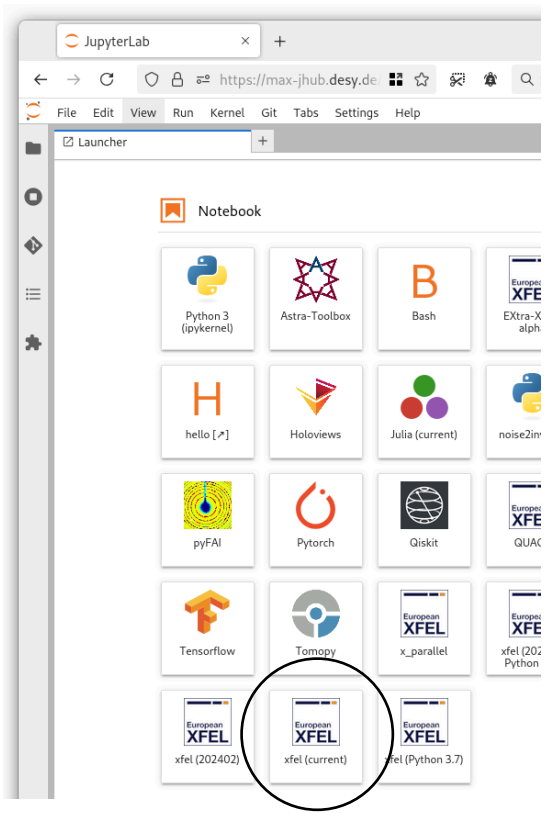
https://extra.readthedocs.io/

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`



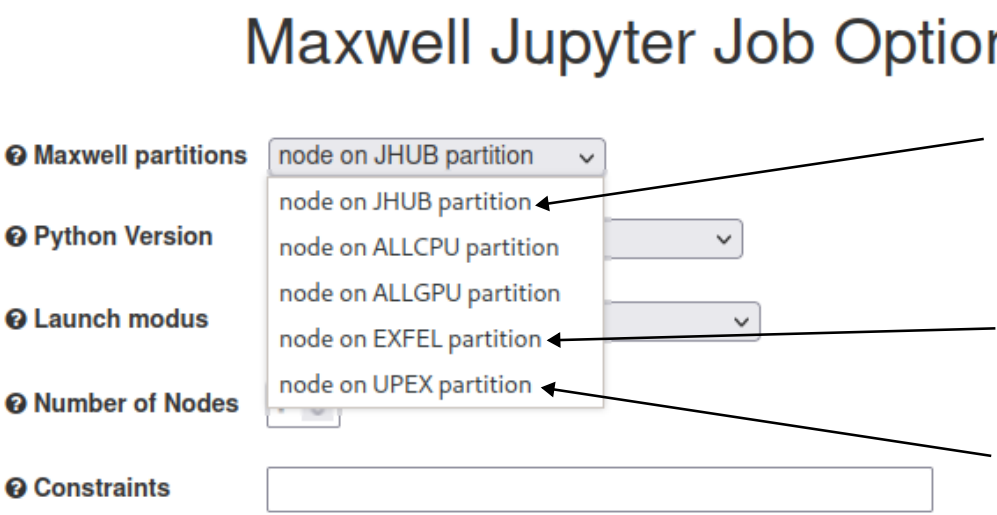
Offline analysis: pipelines

- Serial crystallography: EXtra-Xwiz
  - Wraps CrystFEL
- XPCS
- SPI
- XAS, XES

Variable coverage

See talk *Automated analysis workflows* & poster session

Using the cluster: Dedicated node on max-jhub



Shared node, 1 week session

Dedicated node (EuXFEL staff)

Dedicated node (users)

8 hour sessions
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