News and data highlights

Ivette Bermudez, on behalf of the DATA Department

Users' meeting 2025 - Data Satellite

20 January 2025

Outline

News 📢

- Data management
  - Scientific data policy - it's happening!
  - Data management portal for European XFEL users (myMdC).
  - Data lifecycle.
  - Electronic logbook (myLog).
- Software
  - Data analysis software.
  - Calibration pipeline.

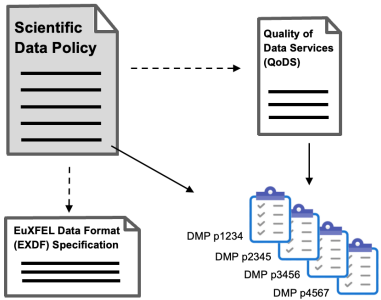
Highlights ★

- EXtra: making data analysis simpler.
- DAMNIT.
- Documentation.
- Machine learning activities.

News 

Scientific data policy (SDP) 2025+

- New SDP takes effect this year.
  - All proposals in 2025 have a data management plan (DMP).
  - One DMP pilot program per instrument in 2024.

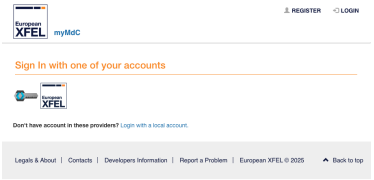


Visit talks

- "Data management plans" by **Krzysztof Wrona** at 2:45pm.
- "Data reduction tools" by **Egor Sobolev** at 3:30pm.

The Data Management portal - myMdC

- Keycloak authentication available.



- Visit talk "Data management plans" by **Krzysztof Wrona** at 2:45pm.

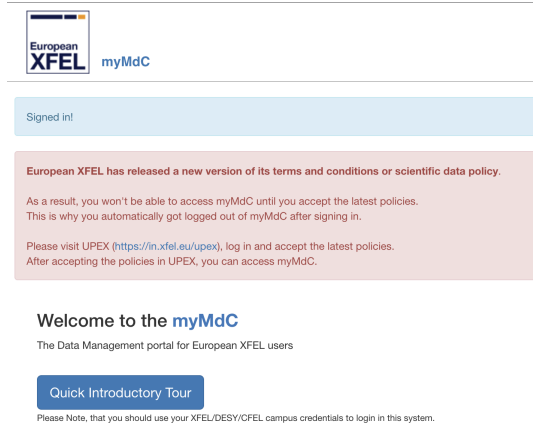
The Data Management portal - myMdC

- Keycloak authentication available.
- Creation of new roles: Experiment Data Contact (**EDC**) and Local Data Contact (**LDC**).

- Visit talk "Data management plans" by **Krzysztof Wrona** at 2:45pm.

The Data Management portal - myMdC

- Keycloak authentication available.
- Creation of new roles: Experiment Data Contact (**EDC**) and Local Data Contact (**LDC**).
- Note: **myMdC** will logout automatically anyone who has NOT accepted the SDP.

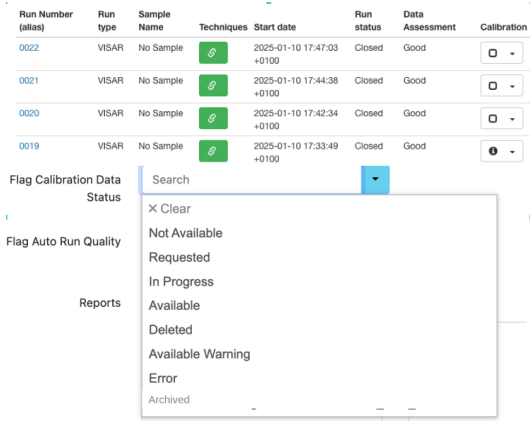


- Visit talk "Data management plans" by **Krzysztof Wrona** at 2:45pm.

The Data Management portal - myMdC

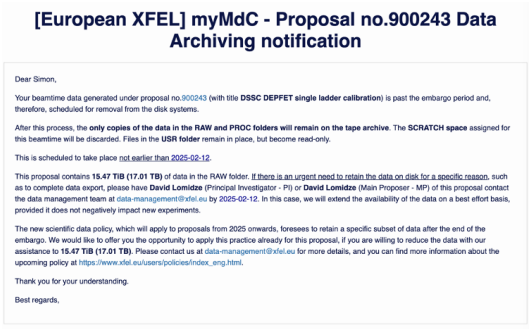
- Keycloak authentication available.
- Creation of new roles: Experiment Data Contact (**EDC**) and Local Data Contact (**LDC**).
- Note: **myMdC** will logout automatically anyone who has NOT accepted the SDP.
- More calibration status: *archived*, *warning*.
  - Field to insert more detailed feedback information on errors and warnings.

Visit talk "Data management plans" by **Krzysztof Wrona** at 2:45pm.



Data lifecycle

- Automatic email to users informing about data stop being available in our infrastructure.
  - Email sent on the day embargo is expected to end.
  - Data may be removed **6 weeks** after notification.
  - This time period may be reduced during 2025 in case of urgent need.



[European XFEL] myMdC - Proposal no.900243 Data Archiving notification

Dear Simon,

Your beamtime data generated under proposal no.900243 (with title **DSSC DEPFET single ladder calibration**) is past the embargo period and, therefore, scheduled for removal from the disk systems.

After this process, the **only copies of the data in the RAW and PROC folders will remain on the tape archive**. The **SCRATCH space** assigned for this beamtime will be discarded. Files in the **USR folder** remain in place, but become read-only.

This is scheduled to take place not earlier than 2025-02-12.

This proposal contains **15.47 TiB (17.01 TB)** of data in the RAW folder. If there is an urgent need to retain the data on disk for a specific reason, such as to complete data export, please have **David Lomidze** (Principal Investigator - PI) or **David Lomidze** (Main Proposer - MP) of this proposal contact the data management team at data-management@xfel.eu by 2025-02-12. In this case, we will extend the availability of the data on a best effort basis, provided it does not negatively impact new experiments.

The new scientific data policy, which will apply to proposals from 2025 onwards, foresees to retain a specific subset of data after the end of the embargo. We would like to offer you the opportunity to apply this practice already for this proposal, if you are willing to reduce the data with our assistance to **15.47 TiB (17.01 TB)**. Please contact us at data-management@xfel.eu for more details, and you can find more information about the upcoming policy at https://www.xfel.eu/users/policies/index_eng.html.

Thank you for your understanding.

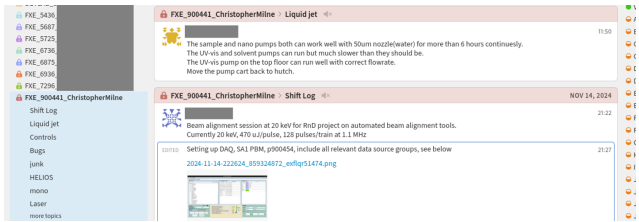
Best regards,

Visit

- Poster "Proposal Lifecycle Services at the European XFEL" by **Luis Maia**.
- Poster "Data Management Infrastructure for European XFEL" by **Janusz Malka**.
- Talk "Data management plans" by **Krzysztof Wrona** at 2:45pm

Electronic logbook myLog

- Logbook can be created once the proposal is in myMdC (upon submission of the *beamtime confirmation form*).
- Move old logbooks from Elog to myLog.
- Aiming to enable direct messaging functionality in myLog.
  - Meant for less formal communication.
  - Shorter retention (e.g. few months). Important information must go to the official stream.



Visit poster "MyLog: the new Electronic Logbook of European XFEL" by **Luis Maia**.

Data analysis software

Software environments

- New cycle environment in 2025 on Python 3.11.
- Check the documentation to see a list of changes.
- Contact us `da-support@xfel.eu` for support if needed.

Extra-data

- Tab-completing in IPython and Jupyter for source and key names.
  - Example: `run['FXE<TAB>` → list of sources starting with FXE.
- New `.masked_data()` method to load detector data with mask.

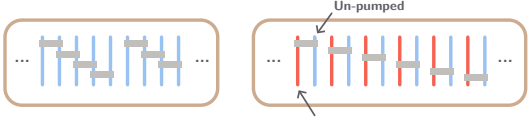
Extra-geom

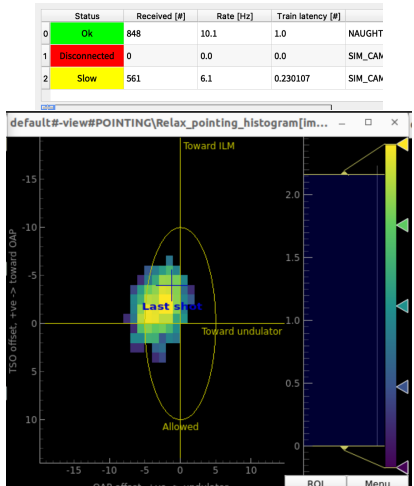
- Update AGIPD detector geometry based on motor positions.

Data analysis software

Extra-metro

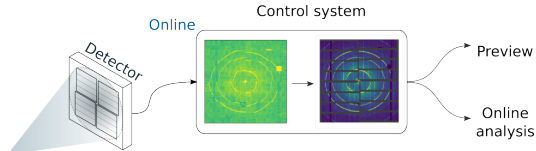
- Detailed diagnostics of incoming data streams.
- Views can run concurrently if they don't depend on each other.
- Improvements to annotations and plotting.
- Analysis on multiple trains in a single event (previously only one).





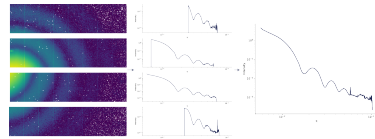
Online calibration processing pipeline

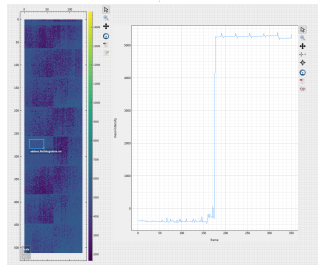
- Online processing system corrects full detector data in real time.
- Extensibility: integrate your code towards a more complete online analysis.
- Execute custom kernels on full data stream.
- Filter out irrelevant frames train by train.
- Example use case: SFX online indexing
  - Apply user-defined mask.
  - Run peakfinding during correction step.
  - Two-way bridge to CrystFEL for indexing.



Online calibration: recent integrations

- Azimuthal integration (new)
- Detector saturation monitoring
- Autocorrelation
- ROI integration
- AGIPD geometry with motor tracking
- More flexible frame selection: combine multiple selection kernels



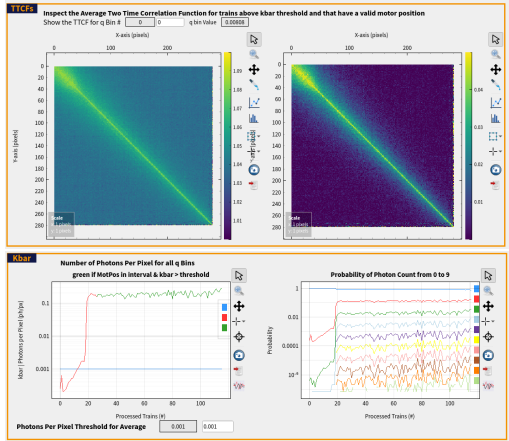


These are advanced features; please contact your local data contact. Feel free to suggest additional useful (pre-)processing steps to upstream.

- Visit talk "Data reduction tools" by **Egor Sobolev** at 3:30pm.

Online XPCS

- Real time feedback for XPCS experiments with AGIPD at MID.
  - Several ROIs, Two-time correlation functions, photon count statistics, g_2 functions.
- Output saved to DAMNIT.
- Enable better use of beamtime, further auto- and optimization of experimental setup.



Visit

- Poster "A high throughput data pipeline for MHz XPCS: Online analysis" by **Mads Jakobsen**.
- Talk "Automated analysis workflows" by **James Wrigley** at 1:45pm.

Highlights 

EXtra: Making analysis of data simpler.

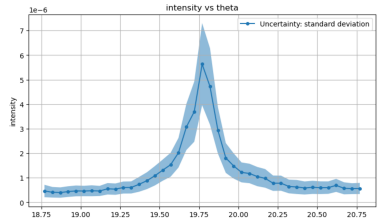
EXtra is a single entry-point for data analysis.

- Access to our other libraries.
- Tools for finding and loading detector calibration constants.
- High-level components that abstract low-level Karabo devices.
- Implementations of specific analysis techniques.

```
scantool = Scantool(run)
```

```
scantool.info()
```

```
Scantool (MID_RR_SYS/MDL/KARABACON) configuration:
  Scan type: dscan
  Acquisition time: 1.0s

Motors:
  SSRY (MID_EXP_SAM/MOTOR/SSRY): -1.0 -> 1.0, 40 steps
```

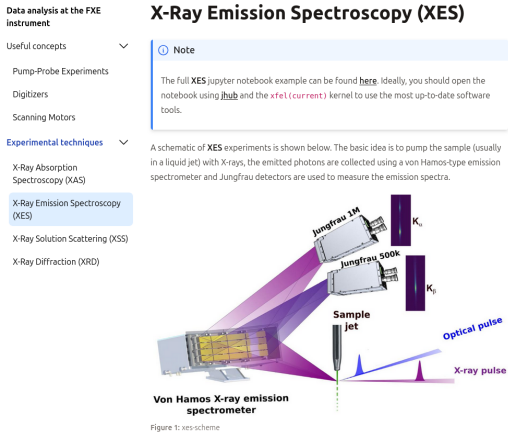



```
scan.plot_bin_by_steps(peak.intensity)
```

Visit talk "High-level analysis building blocks" by **Philipp Schmidt** at 1pm.

Experimental technique oriented documentation

- Collaboration between DA, instrument scientists and users - talk to us!
- Documentation based on the analysis steps of an experimental technique.
- Explanation of useful concepts from a data analysis perspective.
- Jupyter notebook with all the analysis steps for each technique.



DAMNIT: API & WEB

Python API

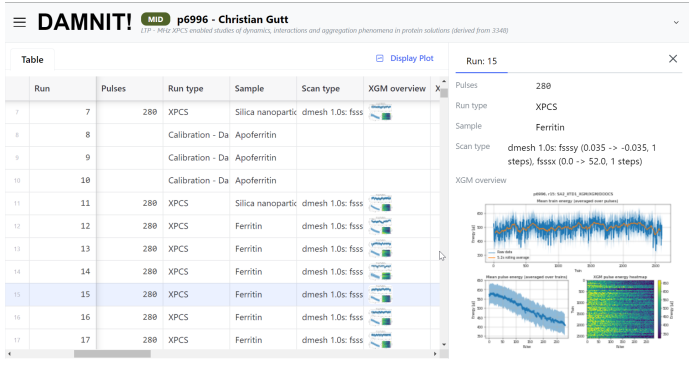
- Your go-to for accessing the DAMNIT database.

```python
from damnit import Damnit

db = Damnit(1234) # This would also work: Damnit("/my/path/to/amore")
```

```python
run_vars = db[100] # Index by run number
run_vars.keys()    # Get all available variables for this run

myvar = db[100, "myvar"] # Equivalent to run_vars["myvar"]
data = myvar.read()
summary = myvar.summary()
```

Web interface

- Improving accessibility. From Desktop app ⇒ web interface.
- Currently only available in the internal network. Publicly available soon™.



Visit poster "Experiment overview and automated data analysis with DAMNIT" by **Thomas Michelat**.

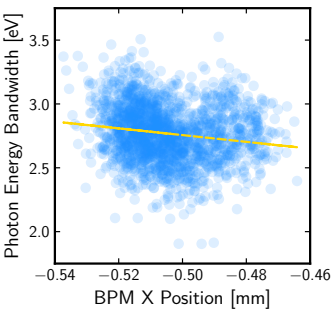
Highlights ★

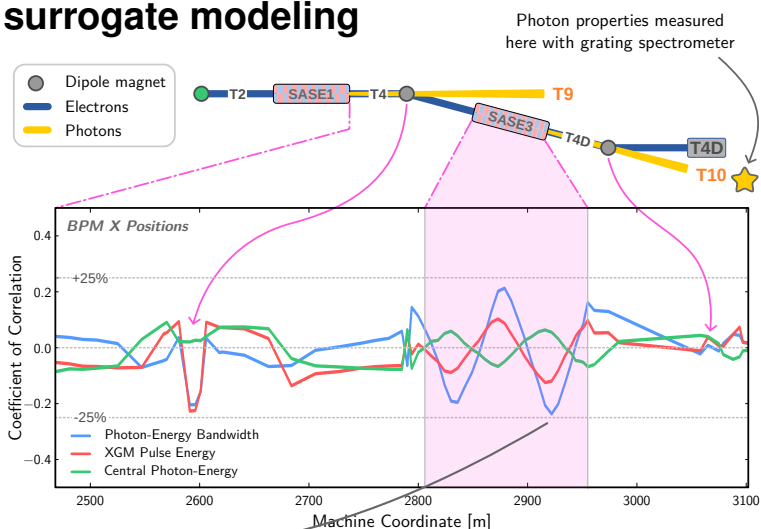
Machine learning activities

Virtual diagnostics: surrogate modeling

Spectral properties at MHz repetition rate, non-invasive

- **Machine tuning**: real-time feedback for operators
- More diagnostics for you!





Visit poster "Electron-Photon Correlations: Towards X-Ray Pulse Diagnostics at MHz Repetition Rates" by **Fady Bishara**.

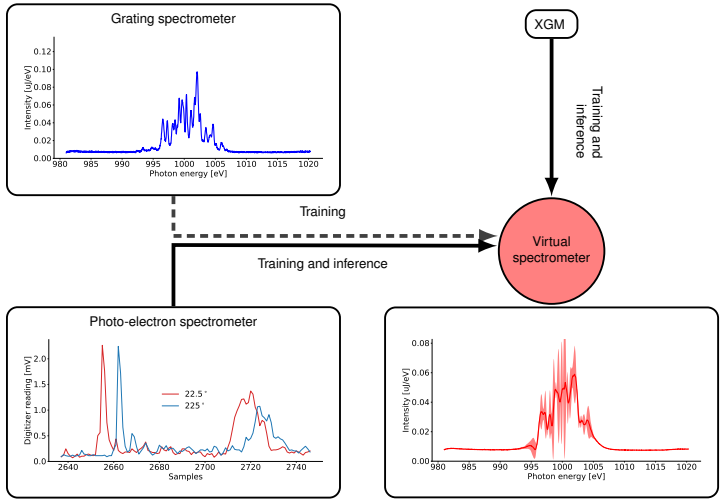
Virtual diagnostics: enhanced spectral characterization

Grating Spectrometer

- Invasive.
- Train-resolved.
- Simple calibration.
- High resolution.

Photoelectron spectrometer

- Non-invasive.
- Pulse-resolved.
- Complex calibration.
- Low resolution.



Virtual diagnostics: enhanced spectral characterization

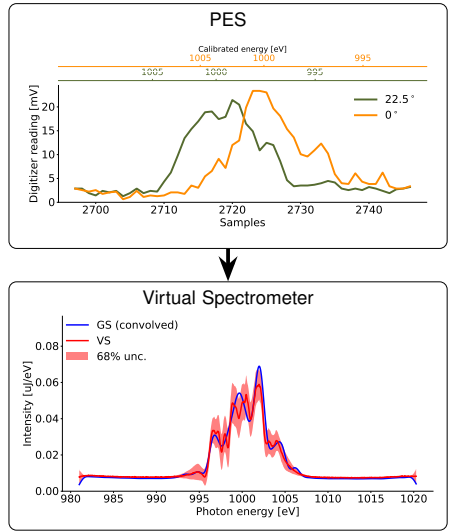
Virtual spectrometer

- Non-invasive.
- Pulse-resolved.
- Simple calibration.
- Improved resolution.

Further work on temporal diagnostic tools.

Results published in “Machine-learning-enhanced automatic spectral characterization of x-ray pulses from a free-electron laser”.

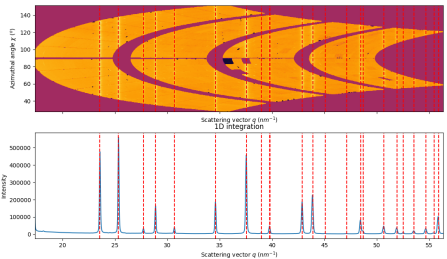
Visit talk "Machine-learning-enhanced characterization of x-ray pulses" by **Danilo E. Ferreira de Lima** and posters.

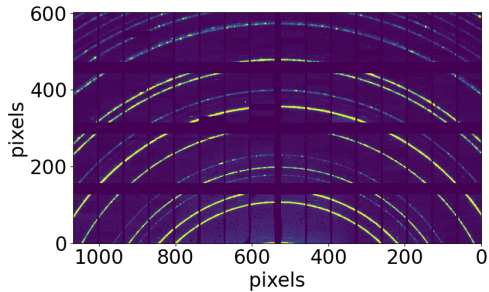


Automation: Multi-modular geometry tuning

Misalignment on module positions of multi-module detectors.

- Manual alignment: requires lots of time.
- Powder diffraction data are often the starting point for techniques requiring high-precision.
- Powder diffraction-based methods require many parameters and manual tuning.

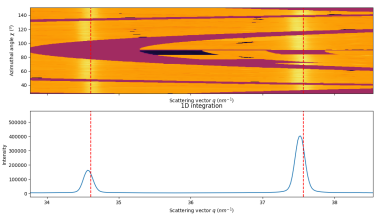




Automation: Multi-modular geometry tuning.

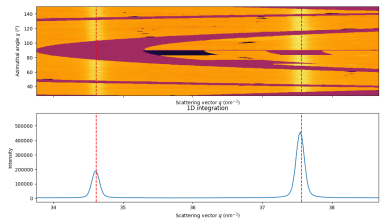
- Only first step in a long pipeline due to the limited experimental method resolution.
- Validation tools available ⇒ quality metrics.
- Jupyter Notebook tool widely adopted by users last year. Available for JUNGFRAU, AGIPD and LPD detectors.
- Implemented in DAMNIT.

Before tuning



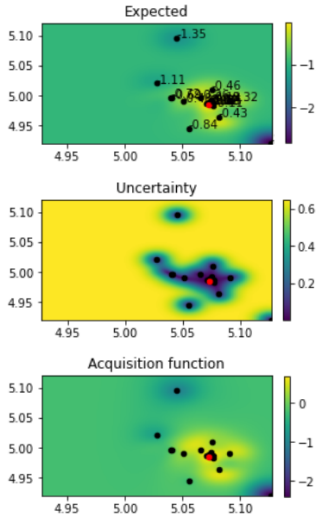
Minimize
$\mathrm{MI}(\mathrm{r}, \phi)$

After tuning



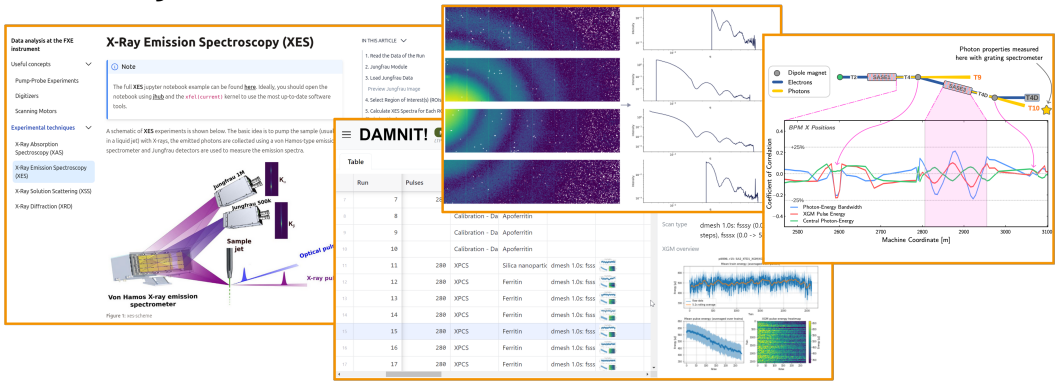
Automation: Alignment of facility sub-systems

- Optical laser alignment (SXP)
  - Align laser to the center of the camera, maintain the alignment using Bayesian optimization (BO).
- NKB mirror alignment (SPB/SFX)
  - Auto-focusing of the beam with NKB mirrors.
- Beamline alignment (upstream FXE)
  - Align beam adjusting multiple optical components.



Visit poster "Automation of facility sub-systems" by **Sarlota Birnsteinova**.

Summary



We'd like to collaborate with you.
Contact us with ideas or requests at da@xfel.eu.
Thank you!

Our posters - Come visit us!

- Danilo - Interpretable Machine Learning at the European XFEL.
- Danilo - Enhancing spectral and temporal diagnostics at European XFEL.
- Egor - Data reduction activities at European XFEL.
- Fabio - The concept of Data Management Plans for European XFEL proposals.
- Fabio - The forthcoming Scientific Data Policy at European XFEL.
- Fady - Electron-photon correlations: towards x-ray pulse diagnostics at MHz repetition rate.
- Ivette - Data Analysis at the European XFEL.
- Mads - A high throughput data pipeline for MHz XPCS: Online analysis.
- Mads - A high throughput data pipeline for MHz XPCS: Offline analysis.
- Matheus - DAPHNE4NFDI at European XFEL.
- Oleksii - Automatic data processing and results overview during SFX experiments.
- Sarlota - Automation of facility sub-systems.
- Thomas M - Experiment overview and automated data analysis with DAMNIT.
- Tim - Fluctuation x-ray scattering data analysis.
- Janusz - Data Management Infrastructure for European XFEL.
- Luis - Proposal Lifecycle Services at the European XFEL -Managing the Proposal Lifecycle from idea to Open Data.
- Luis - MyLog: the new Electronic Logbook of European XFEL.