European XFEL users' meeting 2025

# The new Scientific Data Policy of the European XFEL

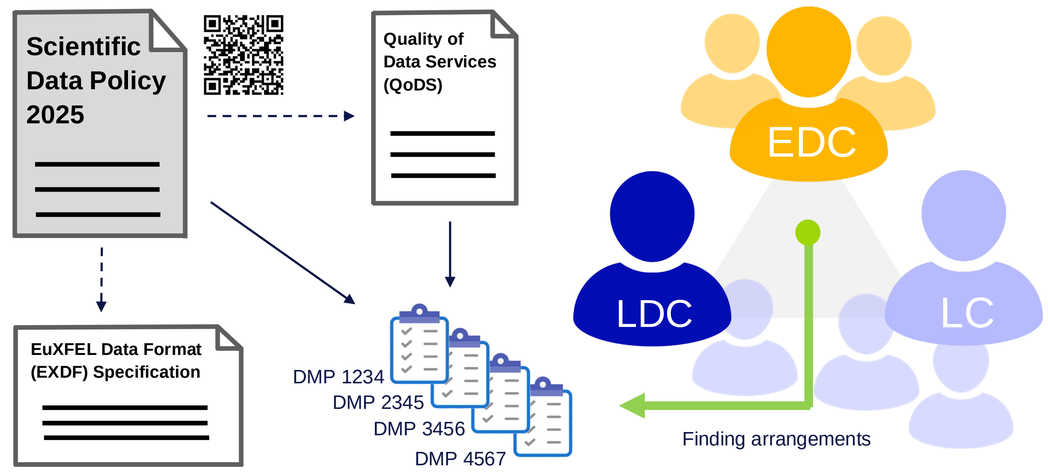
Fabio Dall'Antonia, Luca Gelisio, Janusz Malka, Philipp Schmidt, Egor Sobolev, Gundel Winterhoff and Krzysztof Wrona, on behalf of the data policy and data reduction working group.
European X-ray Free Electron Laser Facility GmbH, Schenefeld, Germany.

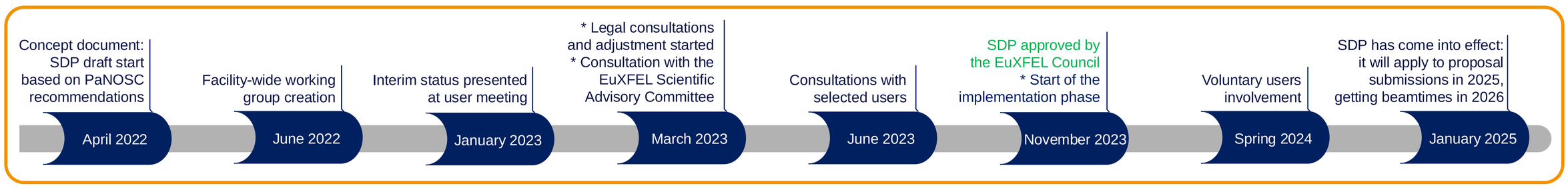
**See also:**
- F. Dall'Antonia et al. *The concept of Data Management Plans for European XFEL proposals*. **Poster #**.
- L. Maia et al. *Services for the Proposal Lifecycle at EuXFEL.* **Poster #**
- E. Sobolev et al. *Data Reduction Activities ar European XFEL.* **Poster #**

The new Scientific Data Policy (SDP) of European XFEL has taken effect with the beginning of this year. It will be the basis for an improved data management practice including data retention regulations and inclusion of FAIR principles[1], catering to the needs of our users and the facility alike.

## Major changes with respect to the previous data policy

- Adoption of recommendations for FAIR scientific data curation,[2] i. p. regarding the provision of open data with DOIs to the scientific community
- Definition of reduced data and regulations concerning its creation, preservation and publication
- Definition of data management plans (DMPs)
- More comprehensive definition of rights and responsibilities of the PI
- Definition of data service quality levels including a retention scheme in a separate document (QoDS)

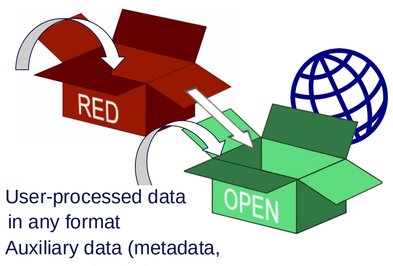



## RED data storage concept

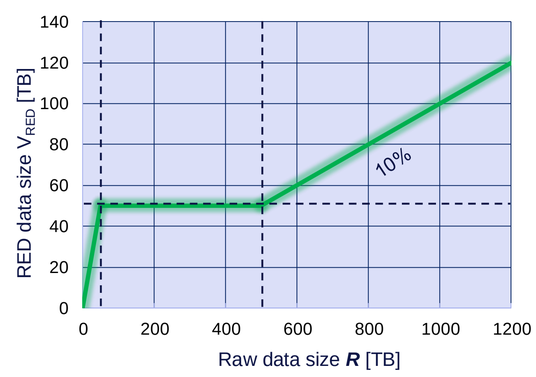
Only RED data will be retained long-term on disk. The restricted volume can have different content of your choice, for instance:

- Selected runs of high-quality raw data, portions otherwise not modified
- All runs of raw data, portions comprised of event-selected detector frames
- Only facility-processed EXDF data, likely selected and/or transformed
- A mixture of selected raw data and facility-processed EXDF data
- A mixture of selected raw data and user-processed data, if EXDF is used




- User-processed data in any format
- Auxiliary data (metadata, sample data etc.)

- The size of raw data (***R***) determines the retained volume.
- Draft quota specified in QoDS:
  $V_{RED} = \max(10\% R; \min(50TB; R))$
  10% of your data, but at least 50 TB
- Reduced data will become open data after the end of embargo – or portions thereof earlier, upon PI request for publication – with extra user data added



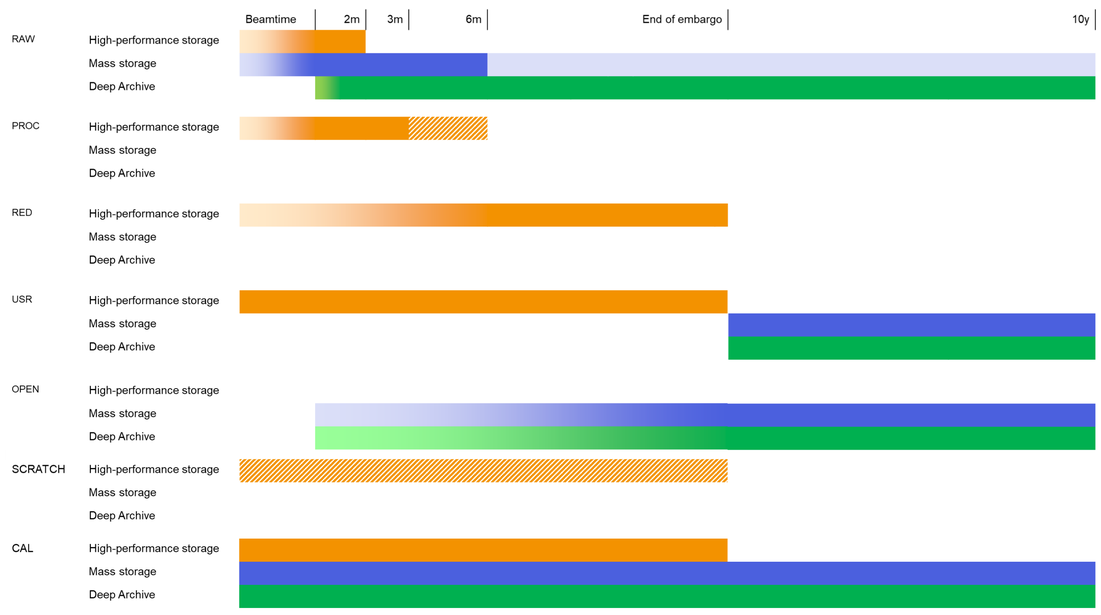

## Data reduction methods and services

- RED data assignment requires user expertise and knowledge about the scientific content/value – users decide about their data
- On the technical level, European XFEL will provide tools and services to assist users with e. g. the following reduction tasks:

| Selection | Transformation |
|---|---|
| ▶ By experiment setup: runs | ▶ Temporal averaging (e.g. trains) |
| ▶ Temporal: trains, pulse pattern | ▶ Spatial integration (e.g. azimuthal) |
| ▶ By event: "hit" frames | ▶ Conversion to photon counts |
| ▶ Spatial: ROIs, modules | ▶ Compression (lossless / lossy) |

For a closer description and discussion please refer to a recent publication[3]

## Draft of the data retention scheme in QoDS




## Summary, conclusions & outlook

- Aim of the new SDP is to ensure successful beamtimes and data analysis up to the publication phase, making data more FAIR throughout;
- Data reduction tools and services are being implemented; data reduction at the instruments has started;
- DMPs have gone into a pilot phase in early 2024, expanded for the upcoming runs of 2025, and will become mandatory in 2026.

**References**

[1] M. D. Wilkinson et al. (2016). *The FAIR Guiding Principles for scientific data management and stewardship*. DOI: 10.1038/sdata.2016.18

[2] A. Götz et al. (2021). *PaNOSC Guidelines on best practices implementing a research data policy*. DOI: 10.5281/zenodo.4899344

[3] E. Sobolev et al. (2024). *Data reduction activities at European XFEL: early results.* Front. Phys. 12, DOI: 10.3389/fphy.2024.1331329

European XFEL GmbH, Fabio Dall'Antonia, Holzkoppel 4, 22869 Schenefeld, Germany, Phone +49 40 8998-6411, fabio.dall.antonia@xfel.eu
**www.xfel.eu**

ENLIGHTENING SCIENCE

**Acknowledgements**

We would like to thank more than 50 colleagues being part of the data policy and data reduction working group who contributed actively to the shaping of the new data policy draft.

Part of the work was funded by the PaNOSC project (H2020 Call: **INFRAEOSC**-04-2018) and LEAPS-INNOV (grant agreement no. 101004728).