



26 October 2023

Scientific Data Policy of the European X-Ray Free-Electron Laser Facility GmbH

(as approved by the Council at its 39th meeting
on 15-16 November 2023)

1	Preface.....	2
2	Definitions	3
3	General principles	6
4	Principal investigator	9
5	Data management plan.....	11
6	Persistent identifiers.....	11
7	Raw data and associated metadata	12
	7.1 Curation of raw data and associated metadata.....	12
	7.2 Access to raw data and associated metadata.....	13
8	Processed data, reduced data, and auxiliary data	14
	8.1 Curation of processed data, reduced data, and auxiliary data	14
	8.2 Access to processed data, reduced data, and auxiliary data	15
	8.3 Intellectual property rights of processed data	16
9	Warranty and liability regarding scientific data	16
10	Good practices.....	17
11	Termination of custodianship or metadata catalogue.....	19
12	Effective date	19

1 Preface

The Scientific Data Policy regulates the responsibilities and rights of the involved parties with respect to scientific data from public research collected and/or stored by the European XFEL GmbH. It pertains to curation, archiving, and access to scientific data from public research. It also introduces the data management plan. In line with FAIR¹ data principles, this Scientific Data Policy ensures that the scientific data are Findable, Accessible, Interoperable, and Reusable.

This Scientific Data Policy follows the recommendations of the Strategic Working Group of PaN-data Europe, laying out a common framework for scientific data management at photon and neutron facilities (Deliverable D2.1, PaN-data Europe, co-funded by the European Commission under the 7th Framework Programme), and updated data policy recommendations developed within the H2020 project, PaNOSC (Deliverable D2.1)².

This Scientific Data Policy does not apply to scientific data generated through proprietary research done through purchased (commercial) access.

¹ M. Wilkinson, M. Dumontier, I. Aalbersberg et al.; “The FAIR Guiding Principles for scientific data management and stewardship”, *Sci. Data* **3**, 160018 (2016)

[doi:10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18)

² H2020-EU.1.4. – EXCELLENT SCIENCE – Research Infrastructures

Grant agreement ID: 823852

[doi:10.3030/823852](https://doi.org/10.3030/823852)

2 Definitions

For the purposes of this Scientific Data Policy:

- 2.1. The term **European XFEL GmbH** shall mean European X-Ray Free-Electron Laser Facility GmbH.
- 2.2. The term **European XFEL Facility** shall mean the complete research facility of European XFEL GmbH located in Hamburg (Bahrenfeld, Osdorfer Born) and in Schenefeld, Schleswig-Holstein, Germany, including underground areas.
- 2.3. The term **European XFEL instruments** shall mean all instruments designed to perform scientific experiments and located at the European XFEL Facility, including those contributed by third-party organizations and user consortia.
- 2.4. The term **public research** shall mean any peer-reviewed experiments, any in-house research or commissioning activities as well as any experiments allocated based on management's contingency as described more detailed in the Policy for the allocation of beamtime at the European XFEL Facility (for non-proprietary research; www.xfel.eu/organization/leadership/council/basic_documents/).
- 2.5. The term **proprietary research** refers to research done through purchased (commercial) access.
- 2.6. The term **user** shall mean a person registered in the User Portal to the European XFEL (UPEX).
- 2.7. The term **proposal** shall mean a specific set of information submitted through UPEX and describing an experiment to be potentially performed at the European XFEL Facility.
- 2.8. The term **principal investigator (PI)** pertains to the person identified as such on the submitted proposal. For experiments outside of UPEX, the PI is the person who initiated the proposal.

- 2.9. The **proposal team** includes the PI and any other user of the European XFEL Facility who is defined as a member of the proposal team in UPEX.
- 2.10. The term **experiment team** includes the PI and any other user of the European XFEL Facility to whom the PI designates the right to access resultant scientific data from the corresponding proposal. The experiment team is defined in the metadata catalogue.
- 2.11. The term **beamtime** shall mean the time period allocated by European XFEL GmbH to the experiment team in a context of a single proposal to perform scientific experiments at a European XFEL instrument. Every beamtime has a well-defined start date and end date. Multiple beamtimes can be allocated to a single proposal.
- 2.12. The term **detector** shall, in this context, mean any measuring device which contributes to the data of the beamtime.
- 2.13. The term **scientific data** shall mean data generated in the context of or related to public research performed at European XFEL instruments. Scientific data include raw data, processed data, reduced data, and auxiliary data as well as all their respective subcategories.
- 2.14. The term **dataset** shall mean a collection of scientific data that is complete and self-consistent, e.g. by including all the scientific data necessary to reproduce the results of the corresponding journal publication, or those related to a certain scientific question.
- 2.15. The term **raw data** shall mean a category of scientific data that is recorded during experiments, that cannot be derived from other persistent data, and that is registered in the metadata catalogue.
- 2.16. The term **reduced raw data** shall mean a subcategory of raw data that pertains to a specifically identified and scientifically meaningful subset of raw data.
- 2.17. The term **processed data** shall mean a category of scientific data that is derived from raw data.

- 2.18. The term **facility-processed data** shall mean a subcategory of the processed data, which is obtained using services provided by the European XFEL Facility, e.g. corrections of detector raw data. This data must adhere to a format supported by the European XFEL GmbH, as defined in the document “European XFEL data and metadata formats”.
- 2.19. The term **user-processed data** shall mean a subcategory of the processed data that is obtained by the experiment team.
- 2.20. The term **reduced data** shall mean a category of scientific data that is a selection of reduced raw data and/or facility-processed data and/or user-processed data in the European XFEL format defined in the document “European XFEL data and metadata formats”.
- 2.21. The term **auxiliary data** shall mean a category of scientific data that provides contextual information regarding the experiment. This may include data generated inside or outside the European XFEL Facility, such as detector geometry data, electronic logbook records, and other logistical information, as well as information about provenance and preparation of the sample, its images, or data processing scripts, software resources and versions used, etc
- 2.22. The term **calibration data** shall mean a subcategory of the auxiliary data that describes detector correction factors.
- 2.23. The term **metadata** shall mean a subcategory of auxiliary data and pertains to information collected in relation to the scientific data.
- 2.24. The term **metadata catalogue** pertains to a computer database of metadata maintained and provided by the European XFEL GmbH containing, e.g. links to proposals and data files, which can be accessed by a variety of methods, including but not limited to web-based browsers.
- 2.25. The term **long-term** shall mean a duration of a minimum 10 years.
- 2.26. The term **storage period** shall mean the time interval for which the European XFEL GmbH stores data. The start of the storage period is defined by the beginning of the first beamtime associated with a given proposal.

- 2.27. The term **quality of data services** shall mean a strategy for storing and accessing scientific data according to their categories and usage purpose, determining their availability.
- 2.28. The term **embargo period** shall mean a period starting with the beginning of the beamtime and ending three years after the end of the beamtime allocated to a given proposal. In case multiple beamtimes are allocated to the same proposal, the embargo period is automatically extended to cover a period of three years after the end of the last beamtime. The embargo period protects the experiment team's right to exclusively draw scientific insight within the context of the proposal, from the scientific data associated with it.
- 2.29. The term **open access** shall mean free availability for use by anyone without fees or patent restrictions as long as attribution is given to the creator. With respect to the European XFEL GmbH, open access shall mean that scientific data will be published under CC-BY (<http://creativecommons.org/licenses/by/4.0/legalcode>) thus ensuring correct attribution. To enable automated harvesting of machine accessible metadata by third parties an exception will be made for high-level metadata, which will be published under CC0 (<https://creativecommons.org/publicdomain/zero/1.0/>).
- 2.30. The term **custodian** in the context of this Scientific Data Policy refers to the European XFEL GmbH, which will store, curate, and manage access to scientific data in accordance with this Scientific Data Policy.
- 2.31. The term **data management plan** (DMP) shall mean the strategies and measures for collecting and handling scientific data and metadata during their lifecycle.

3 General principles

- 3.1. The present Scientific Data Policy pertains to the curation, archiving, and access to scientific data from public research collected at the European

XFEL instruments and/or stored by the European XFEL GmbH. The European XFEL GmbH may use subcontractors to perform its obligations under this Scientific Data Policy. These subcontractors act as data processors in case personal data are concerned.

- 3.2. Persons who want to become users of the European XFEL Facility must personally register at UPEX and accept all required policies and conditions. For details regarding the collection and processing of personal data in connection with UPEX, the UPEX Privacy Policy applies³. For details regarding the collection and processing of personal data in connection with the use of the European XFEL Facility, the data handling provisions of the Terms and Conditions for the non-proprietary user access to the European XFEL Facility⁴ apply. Persons who want to become users of the European XFEL Facility must ensure that their employer agrees with the person's registration in UPEX.
- 3.3. Acceptance of this Scientific Data Policy is a condition for the award of beamtime.
- 3.4. Users shall not attempt to access, exploit, or distribute scientific data unless they are entitled to do so under the terms of this Scientific Data Policy.
- 3.5. As a matter of precaution (and without prejudice to the question of ownership), all members of the experiment team shall hereby grant the European XFEL GmbH the unlimited, unrestricted, irrevocable, non-exclusive, sublicensable right to use the scientific data to the extent necessary to curate, store, and make available said scientific data in

³ The UPEX Privacy Policy is available under <https://in.xfel.eu/upex/docs/upex-privacy-policy.pdf>

⁴ The Terms and Conditions for the non-proprietary user access to the European XFEL Facility are still under negotiations. A final version is not yet available. In case the name of these Terms and Conditions will change, the Council herewith allows the respective amendment of the wording of the Scientific Data Policy without further approval.

accordance with this Scientific Data Policy and to make available said scientific data on an open access basis (as defined in Art. 2.29. of this Scientific Data Policy), as well as to use said scientific data during and after the embargo period to improve the European XFEL Facility's processes and performance, and to support further development that directly enhances the scientific user programme of the European XFEL Facility.

- 3.6. Access to the full experiment proposal and experiment reports will be provided only to the proposal team, to the experiment team, to the members of the proposal review panels, and to authorized European XFEL GmbH staff, guests, and subcontractors, unless otherwise agreed with the PI.
- 3.7. The experiment team and the European XFEL GmbH shall ensure that scientific data are generated with accurate metadata in the spirit of the FAIR principles. For this purpose, the European XFEL GmbH will define a minimum subset of metadata.
- 3.8. Users shall follow good practices as described in Art. 10 of this Scientific Data Policy.
- 3.9. The European XFEL GmbH provides the experiment team members with data services, including access to computing and storage resources for the purpose of processing scientific data related to the performed experiment within the embargo period.
- 3.10. Authorized European XFEL GmbH staff, guests, and subcontractors have access to any scientific data for European XFEL GmbH-related purposes and for providing services to the experiment team. The European XFEL GmbH will ensure that access to such data is restricted in accordance with this Scientific Data Policy during the embargo period.
- 3.11. Within the framework of this Scientific Data Policy, the management of the European XFEL GmbH decides upon further technical and organizational implementation. In particular, the management of the European XFEL GmbH decides on the quality of data services (e.g. maximum data volumes, definition of storage classes, levels of quality

of the data services, and procedures for providing and making the data accessible), as outlined in the document “Quality of Data Services”, which may be subject to change.

- 3.12. Culpable and repeated non-culpable infringements of the Scientific Data Policy may lead to denial of access to scientific data and/or denial of future beamtimes at the European XFEL Facility, as well as actions of the European XFEL GmbH in the court of law.
- 3.13. If and to the extent scientific data include personal data, the data protection legislation of the Federal Republic of Germany and the European Union, respectively, will be applicable. Please be aware that personal data that is included in the published scientific data will also be published in accordance with this Scientific Data Policy.
- 3.14. This Scientific Data Policy is governed by and constructed in accordance with the law of the Federal Republic of Germany. The exclusive place of jurisdiction is Hamburg, Germany.

4 Principal investigator

- 4.1. Within the scope of this Scientific Data Policy, the PI represents the proposal and experiment teams and is responsible for communication with the European XFEL GmbH.
- 4.2. The PI is responsible for ensuring the correctness of information pertaining to the proposal, e.g. referencing the correct proposal numbers and providing a sample description as well as compliance with FAIR principles by the experiment team.
- 4.3. It is the responsibility of the PI to ensure that the metadata collected meets the minimum requirements defined by the European XFEL GmbH.
- 4.4. The PI must ensure that the material uploaded by the experiment team does not violate any license, intellectual property rights, or data protection law.

- 4.5. The PI has the right to compose and modify the proposal and experiment teams within the default or extended embargo period.
- 4.6. The PI has the right to define datasets from the scientific data of the proposal and register associated persistent identifiers through the European XFEL service.
- 4.7. The PI has the right to distribute scientific data outside the experiment team.
- 4.8. A PI who wishes to extend the embargo period shall submit a written request, specifying the reasons for the proposed prolongation, to the management board of European XFEL GmbH, which decides at its own discretion on the request.
- 4.9. During the embargo period, experiment team members must obtain agreement from the PI to distribute scientific data outside the experiment team. After the embargo period, the same applies to any scientific data that is not defined as open access, e.g. the experiment logbook.
- 4.10. The PI has the possibility to transfer the totality of his or her data-related rights and responsibilities to another person registered in UPEX as a user of the European XFEL GmbH. The transfer must be documented in writing and submitted to the European XFEL GmbH.
- 4.11. The PI has the possibility to delegate parts or the totality of his or her rights under this Scientific Data Policy to another person registered in UPEX as a user of the European XFEL GmbH. The delegation must be documented in UPEX.
- 4.12. The PI is encouraged to collaborate with researchers who aim to analyse open access data.
- 4.13. In case of either culpable or non-culpable inability of the PI to adhere to the defined responsibilities, the management board of the European XFEL GmbH reserves the right to decide on the inheritance of the PI's rights and responsibilities in the context of this Scientific Data Policy.

5 Data management plan

- 5.1. A data management plan (DMP) will enhance the communication between the European XFEL GmbH and the proposal and experiment teams with respect to data management.
- 5.2. A DMP must be specifically defined for each proposal.
- 5.3. A DMP will be created at the proposal submission stage and the European XFEL GmbH will ask the users to confirm or update the DMP at different stages until the end of the embargo period. The European XFEL GmbH reserves the right to reject a DMP and any of its updates on the grounds that the DMP is neither comprehensible nor accurate.
- 5.4. The DMP will clarify all aspects of data management and will, in particular, document agreements reached between the PI and European XFEL GmbH, with respect to the provision of data storage, transfers, processing and analysis, computing resources, data retention periods, and data disposal.
- 5.5. Users must apply their best knowledge and due diligence when completing the DMP and are required to follow the guidelines provided by the European XFEL GmbH on what constitutes good data management and guidelines for completing DMPs.

6 Persistent identifiers

- 6.1. The persistent identifier implementation for scientific data and persons is the Digital Object Identifier (DOI) system (www.doi.org) and ORCID (orcid.org), respectively.
- 6.2. Persistent identifiers will be generated by a facility-maintained automated system for open access data and selected associated metadata.
- 6.3. The PI will be able to create or request a persistent identifier for one or more specific datasets belonging to a given proposal.

- 6.4. Any party providing data with the same identifier must ensure the copy is identical to the data in the facility repository.
- 6.5. Any party publishing results based on open access data must quote the same identifier (and related publications if available and required) and follow all necessary provisions pertaining to open access.
- 6.6. High-level metadata, e.g. proposal title, authors, and instrument name, will be made public following the completion of the beamtime. This information will be available via the persistent identifier landing page on the web.

7 Raw data and associated metadata

7.1 Curation of raw data and associated metadata

- 7.1.1. The European XFEL GmbH is the custodian of the raw data and the associated metadata during the storage period.
- 7.1.2. All raw data and associated metadata will be curated in the format defined by the European XFEL GmbH, which also provides the means of reading the data.
- 7.1.3. The European XFEL GmbH raw data file format is HDF5. The structure and content of HDF5 files are defined in the document “European XFEL data and metadata formats”.
- 7.1.4. Raw data and associated metadata are recorded through the services provided by the European XFEL GmbH.
- 7.1.5. In exceptional cases when these services cannot be used and, if agreed upon in the DMP, raw data and associated metadata may be recorded by other means, provided that the experiment team converts such raw data and associated metadata to the corresponding European XFEL GmbH formats and deposit them in the European XFEL GmbH systems within three months after the end of the corresponding beamtime.

- 7.1.6. The European XFEL GmbH will preserve raw data, or where applicable reduced raw data, and associated metadata for the long-term period. In cases where a proposal is allocated additional beamtime to take place at a later date, the storage period is extended to ensure that the long-term storage period uses the end date of the final beamtime. However, in such cases, the European XFEL GmbH reserves the right to restrict the storage periods of datasets.
- 7.1.7. Metadata associated with raw data that are automatically captured by European XFEL instruments will be curated either within the raw data files, within an associated metadata catalogue, or within both.
- 7.1.8. The European XFEL GmbH will provide means for users to store the metadata generated by software not integrated with the European XFEL GmbH services. The metadata must be provided in a format supported by the European XFEL GmbH, as defined in the document “European XFEL data and metadata formats”, and only as such will be accepted as an addition to the existing scientific data.
- 7.1.9. Raw data will be read-only for the duration of its storage period.

7.2 Access to raw data and associated metadata

- 7.2.1. During the embargo period and subject to Art. 3.10. of this Scientific Data Policy, access to raw data and associated metadata will be restricted to the experiment team. However,
- a) high-level metadata, according to Art. 6.5. of this Scientific Data Policy, and
 - b) datasets that are made open public by the PI before the end of the embargo period, according to Art. 8.2.4. of this Scientific Data Policy,
- are already publicly accessible before the end of the embargo period.
- 7.2.2. After the embargo period, metadata associated with raw data obtained as a result of public research will be made available in the searchable

metadata catalogue as open access. Due to the volume of raw data and other technical, operational, and sustainability aspects, it is not feasible to provide general and direct access to the full set of raw data. For this reason, a selection of raw data is available as reduced data, according to Art. 8.1.2. and Art. 8.2.3. of this Scientific Data Policy. The metadata catalogue contains provisions regarding access to such data.

8 Processed data, reduced data, and auxiliary data

8.1 Curation of processed data, reduced data, and auxiliary data

- 8.1.1. Unless otherwise stated in this Scientific Data Policy, the processed data and the associated metadata hereto will not be curated for the long term by the European XFEL GmbH.
- 8.1.2. The reduced data and the associated metadata up to the volume defined by the European XFEL GmbH in the document “Quality of Data Services” will be preserved for the long term. Datasets of the reduced data must be finalized on the decision of the PI within a maximum of six months after the end of the corresponding beamtime, unless otherwise agreed in the DMP. A standard set of services will be provided by the European XFEL GmbH to generate reduced data and validate their format. Optionally, users may use their own tools, provided that the reduced data complies with a format supported by the European XFEL GmbH, as defined in the document “European XFEL data and metadata formats”.
- 8.1.3. A subset of user-processed data not in a format supported by the European XFEL GmbH, as defined in the document “European XFEL data and metadata formats”, and selected by the PI, up to the volume defined by the European XFEL GmbH in the document “Quality of Data Services”, will be curated long term in the original format. At the end of

the embargo period, such data become read-only. The European XFEL GmbH does not guarantee the readability of such data in a format not supported by the European XFEL GmbH.

- 8.1.4. Temporary storage will be provided to enable data processing until the end of the embargo period. The exact amount of available storage space and storage period for beamtime-related processed data will be defined in the DMP.
- 8.1.5. Auxiliary data, up to the volume defined by the European XFEL GmbH in the document “Quality of Data Services” and selected by the PI and the European XFEL GmbH, will be curated long term in the original format. At the end of the embargo period, auxiliary data become read-only. The European XFEL GmbH does not guarantee the readability of auxiliary data in a format not supported by the European XFEL GmbH.
- 8.1.6. All calibration data and the associated metadata obtained in the context of public research through services provided by the European XFEL GmbH will be curated for the long term by the European XFEL GmbH.

8.2 Access to processed data, reduced data, and auxiliary data

- 8.2.1. During the embargo period, subject to Art. 3.10. of this Scientific Data Policy, and with the exception of calibration data and associated metadata, access to processed data, reduced data, auxiliary data, and associated metadata, which are curated by European XFEL GmbH, will be restricted to the experiment team. However,
 - a) high-level metadata, according to Art. 6.5. of this Scientific Data Policy, and
 - b) datasets that are made open public by the PI before the end of the embargo period, according to Art. 8.2.4. of this Scientific Data Policy, and

c) calibration data and associated metadata, according to Art. 8.2.2. of this Scientific Data Policy

are already publicly accessible before the end of the embargo period.

8.2.2. Calibration data and associated metadata will be made available as open access without applying embargo period except otherwise agreed with the European XFEL GmbH.

8.2.3. Reduced data and associated metadata that are curated by the European XFEL GmbH will be made available as open access after the embargo period.

8.2.4. Within the embargo period, the PI has the right to create open datasets from the content of reduced data, auxiliary data, and user-processed data curated by the European XFEL GmbH. These datasets must be defined by the PI, must be registered in the European XFEL metadata catalogue with assignment of a persistent identifier, and must be made read-only at the time of opening.

8.3 Intellectual property rights of processed data

8.3.1. Intellectual property rights of all processed data are determined by the contractual obligations of the person(s) performing the analysis, respectively, the applicable law.

8.3.2. In addition, the clauses on intellectual property of the Terms and Conditions for the non-proprietary user access to the European XFEL Facility apply.

9 Warranty and liability regarding scientific data

9.1. The European XFEL GmbH provides all services specified in this Scientific Data Policy on a best-effort basis.

- 9.2. The European XFEL GmbH will, at its own discretion, use reasonable efforts to ensure accurate storing and curating as well as uninterrupted access in accordance with the acknowledged IT standard. However, failures caused by technical or human error cannot be ruled out regarding any data processing. The European XFEL GmbH can also not warrant absolutely accurate storing and curating. Also, access might be temporarily limited or impossible, especially due to necessary maintenance or overhaul services or failure of third-party service providers. Depending on the data category, the storage place, and the amount of scientific data requested, requests for access to scientific data may, in rare cases, require immense financial, time, and human resources of the European XFEL GmbH. The European XFEL GmbH shall not be liable if making the scientific data accessible takes a longer period of time or if a request is rejected since it exceeds European XFEL GmbH's resources to an unacceptable extent.
- 9.3. The European XFEL GmbH shall not be liable in case of lost, inaccurate, or defective scientific data, as well as for access being limited or unavailable, other than in cases where European XFEL GmbH, a representative, agent, or employee of European XFEL GmbH, acted with gross negligence or intentionally.

10 Good practices

- 10.1. Subject to Art. 3.7. of this Scientific Data Policy, the European XFEL GmbH and the experiment team should adhere to the FAIR principle in all remaining aspects of dealing with scientific data.
- 10.2. The experiment team is encouraged to ensure that auxiliary data, including metadata, are as complete as possible, as this will enhance the possibilities for other parties to search for, retrieve, and interpret the data during the storage period.
- 10.3. The European XFEL GmbH provides means for storage of such auxiliary data that are not automatically captured by an instrument in order to facilitate recording the fullest possible description of the raw data.

Proposal and experiment teams are encouraged to provide auxiliary data in common, non-proprietary formats in order to enhance interoperability.

- 10.4. The experiment team is strongly encouraged to provide a complete log of the protocol carried out and what happened during the experiment. The log must be entered in the electronic logbook provided by the European XFEL GmbH. Selected information will be made available on an open access basis.
- 10.5. Any person who carries out analyses of scientific data is encouraged to link the results of these analyses to the raw datasets using the metadata catalogue and is furthermore encouraged to make such results available on an open access basis.
- 10.6. Researchers are strongly encouraged to follow best practices adopted by many journals concerning citing the software used and developed for the data analysis.
- 10.7. For each publication using scientific data collected during experiments at the European XFEL Facility, authors are strongly encouraged to make available the analysis procedure description, scripts, software, and software environments that completely describe the process of data analysis from the raw data and metadata to the published results, and that allow others to reproduce that analysis.
- 10.8. Authors are encouraged to deposit these files at the storage system of the European XFEL GmbH as auxiliary data associated with the dataset at the time of the submission of the manuscript and to make them available as open access after the publication date.
- 10.9. Where a software tool cannot be made available, e.g. for licensing reasons, the analysis procedure description should explain which tool and version has been used and how the analysis could be repeated if that tool was available.

11 Termination of custodianship or metadata catalogue

If the European XFEL GmbH decides to not continue to act as custodian and/or to maintain and provide the metadata catalogue, the European XFEL GmbH will inform the PIs concerned in a timely manner and provide them with effective means to make a copy of the scientific data, provided that the European XFEL GmbH is able to contact the PI at that time.

12 Effective date

This Scientific Data Policy shall enter into force on 1 January 2025.