

6 June 2017

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

as approved by the Council on 30 June 2017

1	Preface.....	2
2	Definitions	2
3	General principles	5
4	Raw data and associated metadata	6
4.1	Access to raw data and associated metadata.....	6
4.2	Curation of raw data and associated metadata.....	7
4.3	Access to raw data and metadata.....	7
5	Processed data and results.....	9
5.1	Ownership of results	9
5.2	Curation of processed data and results	9
5.3	Access to results.....	10
6	Warranty and liability regarding scientific data, metadata and results.....	10
7	Good practice for metadata captures and results storage	11
8	Publication information.....	12
9	Termination of custodianship or metadata catalogue.....	12

1 Preface

European XFEL GmbH aims to implement a Scientific Data Policy starting from 1 July 2017 in order to be prepared for the first beam operation of the European XFEL Facility.

The main elements of the Scientific Data Policy comprise:

- Data ownership
- Data curation
- Data archiving
- Open access to data

This Scientific Data Policy closely follows the recommendations of the PaN-data Europe Strategic Working Group 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).

2 Definitions

For the purposes of this Scientific Data Policy:

- 2.1. 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.
- 2.2. The term **scientific data** shall mean data collected from peer-reviewed and in-house experiments performed on European XFEL instruments. This definition includes (but is not limited to) data that are created automatically or manually by facility specific software and/or facility staff expertise in order to facilitate subsequent analysis of the experiment data.

- 2.3. The term **raw data** shall mean a category of the scientific data that is recorded during experiments, cannot be derived from other persistent data, and is registered in the metadata catalogue.
- 2.4. The term **processed data** shall mean a category of the scientific data that is derived from raw data.
- 2.5. The term **calibrated data** shall mean a subcategory of the processed data that is obtained from the raw data by applying detector-specific corrections.
- 2.6. The term **calibration data** shall mean a subcategory of the processed data that describes detector correction factors.
- 2.7. The term **alignment data** shall mean a subcategory of the processed data that is obtained from dedicated alignment measurements, providing the position of detectors and their relevant components.
- 2.8. The term **results** pertains to a subset of processed data and other outcomes arising from the analysis of raw data, excluding publications based on such analysis and intellectual property rights out of such analysis or resulting from the registration of a right based on results.
- 2.9. The term **metadata** shall mean information collected in relation to the scientific data, including (but not limited to) the information regarding the context of the experiment, the experiment team, experiment conditions, and other logistical information.
- 2.10. The term **principal investigator (PI)** pertains to the person identified as PI on the experiment proposal. For experiments outside of the European XFEL GmbH's proposal system, the PI is the person who initiated the experiment proposal and obtained the management board approval of the European XFEL GmbH.
- 2.11. The **proposal team** includes the PI and any other registered user of the European XFEL Facility who is defined as a member of the proposal team in the European XFEL user portal (UPEX).
- 2.12. The term **experiment team** includes the PI and any other registered user of the European XFEL Facility to whom the PI designates the right

to access resultant raw data and associated metadata in the European XFEL GmbH's proposal system. For experiments outside of the European GmbH's proposal system, the experiment team is defined in the metadata catalogue.

- 2.13. The term **beamtime** shall mean time period allocated by European XFEL GmbH to the experiment team for performing scientific experiments on a European XFEL instrument. Every beamtime has a well-defined start date, end date, and experiment team.
- 2.14. The term **public research** refers to research done through peer review or access via in-house research beamtime.
- 2.15. The term **proprietary research** refers to research done through purchased (commercial) access.
- 2.16. The term **metadata catalogue** pertains to a computer database of metadata maintained and provided by the European XFEL GmbH containing links to data files, which can be accessed by a variety of methods, including but not limited to web-based browsers.
- 2.17. The term **custodian** in the context of this Scientific Data Policy refers to the European XFEL GmbH which will store, curate, and provide access to raw data, metadata, and results.
- 2.18. The term **long-term** means a minimum of five years and the European XFEL GmbH will strive for 10 years (**storage period**), starting with the end of the respective beamtime. The precise period will depend on the type and volume of data concerned and the economic consequences associated to long-term data storage. Thus, the European XFEL GmbH reserves the right to restrict the storage periods of data sets in consultation with the respective communities of high data-rate instruments.
- 2.19. The term **open access** shall mean the generally unrestricted (but not anonymous) and free-of-charge access and use for the academic community. With respect to the European XFEL GmbH, open access shall mean that data will be published under CC-BY

(Creative Commons BY,
<http://creativecommons.org/licenses/by/4.0/legalcode>).

- 2.20. The term **European XFEL GmbH** shall mean European X-Ray Free-Electron Laser Facility GmbH.
- 2.21. The term **European XFEL Facility** shall mean the research facility of European XFEL GmbH located in Hamburg (Bahrenfeld, Osdorfer Born) and in Schenefeld, Schleswig-Holstein, Germany.
- 2.22. The term **embargo period** shall mean a period of three years, starting from the end of the beamtime.

3 General principles

- 3.1. The present Scientific Data Policy pertains to the ownership, curation, archiving and access to scientific data and metadata collected and/or stored at the European XFEL GmbH. 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. Acceptance of this Scientific Data Policy is a condition for the award of beamtime.
- 3.3. Users must not attempt to access, exploit or distribute scientific data or metadata unless they are entitled to do so under the terms of this Scientific Data Policy.
- 3.4. Culpable and repeated non-culpable infringements of the Scientific Data Policy may lead to denial of access to scientific data or metadata and/or denial of future beamtime at the European XFEL GmbH, as well as actions of the European XFEL GmbH in the court of law.
- 3.5. If and to the extent scientific data and metadata include personal data, the data protection legislation of the Federal Republic of Germany and the European Union, respectively, will be applicable. This Scientific Data Policy will be governed by and constructed in accordance with the law of

the Federal Republic of Germany. Exclusive place of jurisdiction is Hamburg, Germany.

- 3.6. Users of the European XFEL Facility must be personally registered at the user portal (UPEX); for details regarding the collection and processing of personal data in connection with UPEX, the UPEX Privacy Policy applies¹. 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 Facility² apply.

4 Raw data and associated metadata

European XFEL GmbH is the custodian of the raw data and the associated metadata during the storage period.

4.1 Access to raw data and associated metadata

- 4.1.1. All raw data and the associated metadata obtained as a result of public research will be made available as open access after an initial embargo period during which access is restricted to the experiment team, represented by the PI.
- 4.1.2. All raw data and the associated metadata obtained as a result of proprietary research will be owned exclusively by the client who purchased the access and is not covered by this Scientific Data Policy. Data from proprietary research will neither be made available as open access nor curated or stored at European XFEL GmbH. Data from proprietary research will be removed after the experiment from the European XFEL GmbH's storage, unless otherwise agreed with the

¹ The UPEX Privacy Policy is ready and available on the webpage of European XFEL GmbH.

² 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.

management board of the European XFEL GmbH before the start of the experiment.

4.2 Curation of raw data and associated metadata

- 4.2.1. All raw data and metadata will be curated in well-defined formats, for which the means of reading the data will be made available by the European XFEL GmbH.
- 4.2.2. Metadata that are automatically captured by instruments will be curated either within the raw data files, within an associated metadata catalogue, or within both.
- 4.2.3. Only data with metadata generated by European XFEL GmbH's software will be archived.
- 4.2.4. Raw data will be read-only for the duration of its storage.
- 4.2.5. Raw data and metadata will be migrated or copied to archival facilities for long-term curation.
- 4.2.6. It is planned that each data set will have a unique persistent identifier. Anyone providing data with the same identifier must make sure that the copy is identical to the data in the facility repository. Anyone publishing results based on open access data must quote the same identifier (and related publications if available and required).
- 4.2.7. High-level metadata, such as title, authors, instrument, will be made public as soon as the experiment has been carried out. This information will be available via the persistent identifier landing page on the web.

4.3 Access to raw data and metadata

- 4.3.1. Access to raw data and metadata is foreseen to be via a searchable metadata catalogue.
- 4.3.2. Access to the metadata catalogue of the European XFEL GmbH will be restricted to registered users. The European XFEL GmbH sets up the

online procedure to become a registered user of the metadata catalogue.

- 4.3.3. Any PI that wishes to extend the embargo period might submit a written request, specifying the reasons for the proposed prolongation, to the management board of European XFEL GmbH, which decides on the request at its own discretion. In exceptional circumstances, data can be made openly accessible during the embargo period if the PI informs the European XFEL GmbH to do so and subject to its own discretion.
- 4.3.4. It is the responsibility of the PI to ensure that the proposal, experiment numbers, and sample description are correctly entered through the control system or directly into the metadata for each raw data set.
- 4.3.5. Authorized European XFEL staff (e.g. instrument and detector scientists, computing group members) have access to any curated data or metadata for facility related purposes. European XFEL GmbH will undertake that the confidentiality of such data is preserved during the embargo period.
- 4.3.6. The metadata catalogue will enable linking scientific data to experiment proposals. Access to proposals will be provided only to the proposal team, the experiment team, the members of the proposal review panels, and appropriate facility staff, unless otherwise authorized by the PI.
- 4.3.7. The PI has the possibility to transfer the totality of her/his data-related rights and responsibilities during the embargo period to another person registered in the user portal (UPEX) as a user of the European XFEL Facility. The transfer must be documented in writing.
- 4.3.8. The PI has the possibility to delegate parts or the totality of her/his rights under this Scientific Data Policy during the embargo period to another person registered in the user portal (UPEX) as a user of the European XFEL Facility. The delegation must be documented in writing.

4.3.9. The PI has the possibility to create and distribute copies of the raw data. During the embargo period other experiment team members must obtain agreement from the PI to do so.

4.3.10. As a matter of precaution (and without prejudice to the question of ownership), all members of the experiment team grant the European XFEL GmbH the unlimited and unrestricted right to use the raw data and the metadata to the extent necessary to curate and make available the raw data and the metadata in accordance with this Scientific Data Policy.

5 Processed data and results

5.1 Ownership of results

5.1.1. Ownership, including any IP rights that might arise, of all results derived from the analysis of the raw data is determined by the contractual obligations of the person(s) performing the analysis respectively the applicable law.

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

5.2 Curation of processed data and results

5.2.1. Unless otherwise stated in this Scientific Data Policy, the processed data from the interim analysis steps and the associated metadata hereto will not be curated for long-term by European XFEL GmbH. In particular, calibrated data and the associated metadata hereto will not be curated for long-term, but the tools will be provided to regenerate calibrated data from raw data using facility curated calibration data. A temporary storage will be provided to enable data analysis. The exact amount of available storage space and storage period for beamtime related processed data will be published by European XFEL GmbH.

- 5.2.2. All calibration and alignment data and the associated metadata obtained in the context of public research through the European XFEL GmbH provided services will be curated for long-term by the European XFEL Facility.
- 5.2.3. If agreed between PI and European XFEL GmbH and to the extent results are entered into its system, the European XFEL GmbH will provide curation of results on a best effort basis, and act as custodian of results for the storage period.
- 5.2.4. European XFEL GmbH will not be responsible for the full curation of the results, e.g. to ensure that software to read/manipulate this data is available.
- 5.2.5. As a matter of precaution (and without prejudice to the question of ownership), all members of the experiment team grant the European XFEL GmbH the unlimited and unrestricted right to use the processed data and the metadata hereto to the extent necessary to curate and make available the processed data and the metadata hereto in accordance with this Scientific Data Policy.

5.3 Access to results

- 5.3.1. Calibration and alignment data and associated metadata will be made available as open access without applying embargo period, except as otherwise agreed with European XFEL GmbH.
- 5.3.2. Access to the results of analysis performed on raw data and metadata and curated by European XFEL GmbH is restricted to the experiment team, unless otherwise requested by PI.

6 Warranty and liability regarding scientific data, metadata and results

- 6.1. European XFEL GmbH will at its own discretion use reasonable efforts to ensure an accurate storing and curating as well as an uninterrupted

access in accordance with the acknowledged IT standard. However, failures caused by technical or human mistakes cannot be ruled out regarding any data processing. European XFEL GmbH cannot warrant an 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.

- 6.2. European XFEL GmbH shall not be liable in case of lost, inaccurate, or defective scientific data, metadata, or results as well as for access being limited or unavailable unless European XFEL GmbH, a representative, agent, or employee of European XFEL GmbH acted in a grossly negligent manner or intentionally.

7 Good practice for metadata captures and results storage

- 7.1. The experiment team is encouraged to ensure that metadata are as complete as possible, as this will enhance the possibilities for everyone to search for, retrieve, and interpret the data during the storage period.
- 7.2. European XFEL GmbH provides means for the capture of such metadata items that are not automatically captured by an instrument in order to facilitate recording the fullest possible description of the raw data.
- 7.3. PIs and researchers who carry out analyses of scientific data and metadata are encouraged to link the results of these analyses to the raw data/metadata using the metadata catalogue. Furthermore, they are encouraged to make such results available on an open access basis.

8 Publication information

- 8.1. Publications related to raw data and metadata from experiments carried out at European XFEL Facility must cite the persistent identifier of the beamtime data in their publications.
- 8.2. References for publications related to experiments carried out at the European XFEL Facility must be deposited in the European XFEL publications database within three months of the publication date, or during any new application for beam time, whichever is earlier.

9 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 respective raw data, metadata, calibration data, alignment data, and results, provided European XFEL GmbH is aware of the effective email address of the PI at that time.