DESY Photon Science Division Scientific Data Policy

Version 1.1

Status: 17.10.2025

Change history

Version	Author	Date	Summary
1.1		17.10.2025	Version approved by the DESY board of directors

Please send questions and comments/remarks to compliance@desy.de.

Imprint

Deutsches Elektronen-Synchrotron DESY
DESY Photon Science Division Scientific Data Policy

Version 1.1

For citation please use:
DOI: 10.3204/PUBDB-2025-05229

Editor:
Verlag Deutsches Elektronen-Synchrotron
Notkestraße 85
22607 Hamburg
Germany

Table of contents

DESY Photon Science Division Scientific Data Policy

Preface

The DESY Photon Science Division Scientific Data Policy regulates the responsibilities and rights of the involved parties with respect to scientific data from public research collected at user facilities operated by DESY Photon Science (DESY-FS) and stored on DESY compute and IT infrastructure. It does not apply to other DESY departments, nor does it apply to scientific data generated through proprietary research through purchased (commercial) access.
The scientific data policy defined in this document serves as the DESY Photon Science specific supplement to the DESY framework policy[1] ("Rahmenrichtlinie zum Forschungsdatenmanagement bei DESY").
This Scientific Data Policy pertains to curation, archiving, and access to scientific data from public research and aims at ensuring that the scientific data produced at the DESY-FS facilities will be Findable, Accessible, Interoperable, and Reusable in line with FAIR[2] data principles.
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[3] , the updated data policy recommendations developed within the H2020 project PaNOSC[4] and „Recommendations for guidelines for the Helmholtz Centers on handling of research data "[5].

General principles

2.1. This Scientific Data Policy pertains to the ownership of, the curation of and access to experimental data and metadata collected and/or stored at DESY by or on behalf of DESY Photon Science (DESY-FS). Where "data", as defined in section 3.4, includes raw, processed and auxiliary data which have been curated at DESY but originate elsewhere. The term "data" is used in this document according to this definition.

2.2. The Acceptance via DOOR of this policy is a condition for the award of publicly funded beamtime. In this context the use of DESY-FS user facilities is linked to the user's consent to this DESY Photon Science Division Scientific Data Policy. Persons who want to become users of the DESY Facility must ensure that their employer agrees with the person's registration in DOOR.

2.3. Users shall not attempt to access, exploit or distribute data including but not limited to raw data, processed data, auxiliary data, metadata or results unless they are entitled to do so under the terms of this policy.

2.4. Deliberate infringements of the policy may lead to denial of access to data and/or denial of future beamtime requests at DESY.

---

[1] Framework Guideline for Research Data Management at DESY, doi: 10.3204/PUBDB-2025-02454
[2] Wilkinson et al., Sci. Data 3, 160018 (2016), doi: 10.1038/sdata.2016.18
[3] Dimper et al., PaN-data Europe, Deliverable D2.1 (2011), doi: 10.5281/zenodo.373849
[4] Gotz et al., PaNOSC Deliverable D2.1 (2020), doi: 10.5281/zenodo.3862701
[5] Empfehlungen für Richtlinien der Helmholtz-Zentren zum Umgang mit Forschungsdaten. In der 109. Mitgliederversammlung der Helmholtz-Gemeinschaft am 13.-14. September 2017 beschlossen. (2019): Potsdam, 14 p. doi: 10.2312/os.helmholtz.002

2.5. Users are encouraged to ensure that all data are collected with accurate metadata to adhere to the FAIR principles. DESY will define a minimum subset of metadata that is mandatory to be collected during the experiment.
2.6. Users shall endeavor to include supplemental metadata and auxiliary data to augment the experiment data.
2.7. DESY cannot be held liable in case of unavailability or loss of data or data analysis software.
2.8. DESY may, for technical or financial reasons, place limitations on the volume or amount of data which can be retained.
2.9. As a matter of precaution and without prejudice to the question of ownership, all members of the experimental team grant DESY the unlimited, unrestricted, irrevocable, non-exclusive, sublicensable right to use the scientific data and metadata to the extent necessary to curate, store and make available datasets in accordance with this Scientific Data Policy, as well as to use said data during and after the embargo period to improve the processes and performance of DESY facilities and to support further developments that directly enhances the scientific user program of DESY.
2.10. The data protection legislation of the Federal Republic of Germany and the European Union, respectively, are applicable. Exclusive jurisdiction is given to the appropriate Courts of Hamburg.

Definitions

3.1. The term **auxiliary data** refers to additional information with respect to the experiment, e.g. regarding the sample (images, provenance and preparation), electronic logbook records, commentary measurements or the software used for data processing. It may include data generated inside or outside DESY facilities.
3.2. The term **beamtime** refers to the period of time when the experimental team has access to the facility resources to conduct an experiment.
3.3. The term **custodian** in the context of this Scientific Data Policy refers to DESY, which will store, curate, and manage access to scientific data in accordance with this Scientific Data Policy.
3.4. The term **data** pertains to raw data (see 3.20), metadata (see 3.12), processed data (see 3.15), and auxiliary data (see 3.1). This explicitly does not include scientific publications.
3.5. The term **dataset** shall mean a collection of scientific data that aims to be self-describing, e.g. by including all information necessary to reproduce the results.
3.6. The term **DESY-FS user facilities** refers to scientific instrumentation and related IT infrastructure operated by or on behalf of DESY Photon Science including Helmholtz LK-II facilities.
3.7. The term **DOOR** stands for an online portal: DESY Online Office for Research with Photons.
3.8. The term **embargo period** refers to the initial period after a beamtime and before the data is made publicly accessible during which access to the data is restricted.
3.9. The term **experiment report** refers to a brief summary of each experiment session carried out at DESY Photon Science light sources that is to be submitted by the user after each beamtime session.
3.10. The term **experimental team** includes the PI (see 3.14) and any other person to whom the PI designates the right to access resultant raw data and associated metadata.
3.11. The term **long-term** refers to the time period after the embargo period expired.
3.12. The term **metadata** pertains to information collected in relation to the raw-, processed- and auxiliary data to ensure that research data can be found and used efficiently by providing the context (e.g. experimental technique, conditions or logistical information) related to this data.

3.13. The term **online catalogue** pertains to a computer database of metadata containing links to raw data files, that can be accessed by a variety of methods, including (but not limited to) web-based browsers.
3.14. The term **principal investigator** (PI) refers in the context of the DESY Photon Science User Program to an investigator that is assigned with the responsibility for administrative preparation of the beamtime and coordination with the participating investigators (experimental team). Following this definition, the PI reports – in the scope of the beamtime / proposal – to the project leader that is defined below.
3.15. The term **processed data** refers to the data obtained by processing raw data by the DESY provided data processing pipelines.
3.16. The term **project leader** refers to the leading scientist who takes the general responsibility for the project and participating personnel coming to DESY for the experiment, including safety and legal aspects. Therefore, the project leader is expected to be an experienced researcher like a "senior scientist" and should be employed at the institution proposing the experiment. [NOTE: This definition may coincide with the use of the term "principal investigator" in other contexts outside of DOOR / the DESY Photon Science User Program.]
3.17. The term **proprietary research** refers to research done through purchased (commercial) access.
3.18. The term **public research** refers to research done through peer review or access via in-house research beamtime through internal guidelines.
3.19. The term **persistent identifier (PID)** denotes a unique identifier that ensures permanent access for a digital object by providing access to it independently of its physical location or current ownership[6].
3.20. The term **raw data** pertains to data collected on DESY's instruments from peer-reviewed and in-house experiments beamtimes. This definition includes data that are created automatically or manually by facility specific software and/or facility staff expertise in order to facilitate subsequent analysis of the experimental data. Also included are accelerator data relevant for the analysis. Furthermore, the term **raw data** also refers to data created by other in-house research activities.
3.21. The term **release request** means that the curation and metadata addition for a dataset has been marked as completed. This is the prerequisite to make the dataset publicly available and/or to issue a permanent persistent identifier for the dataset (e.g. minting a dataDOI, or equivalent)
3.22. The term **result** pertains to data, intellectual property, and outcomes arising from the analysis of raw data. This does not include publications. Analysis of data involves a creative step with creative input from a researcher, which serves to distinguish results from automated data processing.
3.23. The term **open access** shall mean free availability for use by anyone without fees or patent restrictions and, in cases where there is a protection by copyright, as long as attribution is given to the creator[7]. DESY will release open access data preferable under the CC0[8] license or, if copyright protection applies and when sensible[9], alternatively under CC-BY[10] license. Additionally, machine accessible high-level metadata is released under CC0 license to enable automated harvesting of datasets.

[6] EOSC Glossary (December 2020). Zenodo. doi: 10.5281/zenodo.4472643 and ISO 5127:2017
[7] Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities, 2003, https://openaccess.mpg.de/Berlin-Declaration
[8] https://creativecommons.org/publicdomain/zero/1.0/
[9] N.B. CC-BY for data severely restricts re-usage of data in huge data collections, since attribution will become very complex. On the other hand, even in the CC0 case, the obligations arising from "good scientific practice" require the citation of the sources.
[10] https://creativecommons.org/licenses/by/4.0

Data access and curation

4.1. Access to data and metadata

a. Any raw, processed, and auxiliary data stored by DESY excluding data arising from proprietary research, will be open access after an initial embargo period of 3 years after the end of the experiment. During the embargo period access is restricted to the experimental team, represented by the PI. This initial embargo period is based on considerations such as the typical length of a PhD and the typical time needed for data analysis and scientific publishing.
b. The embargo period begins at the end of the experiment session i.e. at the end of the scheduled beamtime on the beamline.
c. Data can be made openly accessible earlier than 3 years if the project leader informs DESY as the custodian to do so.
d. Any project leader or PI that wishes data to remain restricted access for a period longer than three years will have this possibility by submitting a request via DOOR, specifying the reasons for the proposed prolongation.
e. DESY is the custodian of all data stored at DESY that is covered by this Scientific Data Policy. This does not apply to scientific data generated through proprietary research performed through purchased (commercial) access.
f. DESY is committed to the DFG guidelines for storing data. DESY strives to store any raw, processed, and auxiliary data for ten years. This will depend on the type and volume of data concerned and the economic consequences associated with long-term data storage. Thus, DESY reserves the right to restrict the storage period or total storage volume of datasets in consultation with the respective beamline responsible for high data rate instruments. Metadata is foreseen to be stored indefinitely.
g. The experiment report submitted by the user after each beamtime session will be included as part of the auxiliary data, subject to the same embargo rules (i.e. three-year embargo period) as the raw data.
h. Raw data and metadata will be locatable via a searchable online catalogue.
i. Authorized DESY staff (e.g. instrument scientists, computing group members) may access to any curated data or metadata for facility related purposes. DESY will undertake that confidentiality of such data is preserved during the embargo period.
j. The project leader has the possibility to transfer parts or the totality of her/his/their rights during the embargo period to another DOOR-registered person via informal email to door@desy.de .

4.2. Curation of raw data and associated metadata.

a. All raw data and metadata will be curated in formats such as JSON for high-level metadata and HDF5 following the NeXus convention[11]. Where necessary the means of reading the data will be made available by DESY.
b. It is planned to capture metadata automatically by instruments. The corresponding metadata will be curated either within the raw data files, within an associated online catalogue, or within both.
c. Only data that is associated with the necessary corresponding metadata will be archived.
d. Upon a release request data will be read-only for the remaining duration of their life time. In order to be scientifically citable data must be released and have globally unique identifiers assigned. Further a permanent persistent identifier for the

[11] https://www.nexusformat.org/

dataset (e.g. a dataDOI, or equivalent) can be issued[12]. Correction and addition will only be possible in exceptional cases and have to be logged and documented.

e. Data and metadata will be migrated or copied to archival facilities for long-term curation. Retrieval of data from archival storage may induce significant waiting time before data can be accessed given the limited throughput capacity of the archival facility.
f. Experiments and datasets will be issued a worldwide and time-independent unique identifier. On request a persistent DOI based on the identifier can be minted via DESY. DESY strives to preserve such data permanently. Once a DOI is issued any publication based on open access data should quote this identifier (and related publications if available & required).
g. High level metadata such as Title, Project Leader, Principal Investigator, Experimental Team, Beamline, Technique may be made public as soon as the experiment has been carried out. In case a persistent identifier is minted for the corresponding data, this information will be available via the landing page on the web.
h. The proposal abstract and summaries of the experiment i.e. the experimental reports may be made public after expiration of the embargo period or upon request of a persistent identifier through the PI or project leader. Prior to publication PI or project leader may update the abstract if necessary to better describe the dataset. For high data volume experiments, lossy compression may be used to store data, if required because of raw data volume. Where feasible automated processes to minimize data volume may be put in place for high data rate instruments.
i. DESY encourages users to condense data acquired during their experiments (raw and processed), e.g. by keeping only scientifically useful datasets at the end of the experiment or by supplying suitable metadata that labels data according to scientific relevance, in order to limit the stored data volume through ex post deletion of non-relevant data. Additionally, users are encouraged to delete data that is evidently without scientific value. However, to prevent fraud, DESY as the custodian reserves the right to track and log the deletion of raw data in a transparent and traceable manner.

4.3. Curation of processed and axillary data

a. DESY will provide curation of processed data on a best effort basis, and acts as custodian of results in the long term.
b. When appropriate, depending on the experimental technique and in consultation with the respective beamline responsible, DESY will store only processed data and not the raw data such as individual detector frames.
c. Access to the processed and axillary data stored at DESY is governed in the same manner as access to raw data.

Results

5.1. Ownership of all results derived from the analysis of the raw data, including intellectual property therein, is determined by the contractual obligations of the person(s) performing the analysis. In particular, the clauses on intellectual property of the Terms and Conditions for the non-proprietary user access to DESYs Facility apply.

5.2. DESY will provide curation of results on a best effort basis, and acts as custodian of results in the long term.

[12] All released datasets will automatically have globally unique identifiers assigned. Issuing a permanent persistent identifier is done on request.

5.3. Access to the results of analysis performed on raw data and metadata is restricted to the experimental team, which includes the person or persons performing the analysis. If the raw data being analyzed is still restricted or under embargo, access to the analysis results can only be granted by the project leader or PI on request.

Good practice for metadata captures and reuse of results

6.1. The experimental team is encouraged to ensure that experiment metadata are as complete as possible, as this will enhance the possibilities to search for, retrieve and interpret the data in the long term.

6.2. DESY plans to provide means for the capture of metadata items that are not automatically captured by an instrument, in order to facilitate recording the fullest possible description of the raw data.

6.3. DESY will encourage researchers who aim to carry out analyses of data which are openly accessible should, where possible, contact the project leader or PI to inform her/him/them and suggest a collaboration.

6.4. The researcher who carries out analyses of data which are openly accessible is encouraged to link the results of these analyses to the raw data / metadata using the facilities provided by the online catalogue. Furthermore, they are encouraged to make such results openly accessible.

6.5. The project leader and PI are encouraged to collaborate with researchers who aim to analyze open access data.

Liability regarding scientific data

7.1. DESY provides all services specified in this Scientific Data Policy on a best-effort basis.

7.2. DESY 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. DESY 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 DESY. DESY 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 DESY's resources to an unacceptable extent.

7.3. DESY 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 DESY, a representative, agent, or employee of DESY, acted with gross negligence or intentionally.

Termination of custodianship or metadata catalogue or deletion of data

8.1 If DESY decides to not continue to act as custodian and/or to maintain and provide the metadata catalogue or needs to delete archived data, DESY will inform the project leader and PI concerned in a timely manner and provide them with effective means to make a copy of the scientific data, provided that DESY is able to contact the project leader or PI at that time.

8.2 In the event that project leader, PI or their assignee becomes uncontactable the director of DESY-FS is responsible for the decision.