

Handouts for research data management:

Planning and implementation of research projects

Conclusion phase

Storage and publication

As an institution, TH Wildau is committed to the “DFG Guidelines on the handling of research data” (GWP) and refers to the GWP in its Leitlinie zum Umgang mit Forschungsdaten an der Technischen Hochschule Wildau.¹ In Guideline 13 on the creation of public access to research results and Guideline 17 on archiving, the GWP refers to the handling of the publication of research data. The most important points include:

- Research results and data, as well as associated central materials and any research software used, should be stored in an appropriate manner
- Orientation should follow the standards of a subject area
- Data should be stored for at least 10 years
- Publication should be as open as possible
- If certain data cannot be retained, this must be justified (e.g. GDPR)
- The FAIR principles must be observed during publication

Data selection

In general, it is important to select data according to whether it is suitable for publication, i.e., whether the data is relevant for third parties, and which data should be archived. These guidelines provide information for selecting data:

- <https://www.tu-braunschweig.de/forschung/forschungsdaten-transparenz/forschungsdaten/grundkurs-forschungsdatenmanagement/datenorganisation-dokumentation/datenselektion>
- <https://www.dcc.ac.uk/sites/default/files/documents/publications/Five%20Steps%20to%20decide%20what%20data%20to%20keep.pdf>

¹ TH Wildau. Leitlinie zum Umgang mit Forschungsdaten an der Technischen Hochschule Wildau. Amtliche Mitteilungen 08/2025 vom 27.02.2025. https://www.th-wildau.de/files/2_Dokumente/Amtliche_Mitteilungen/08_2025_FDM_Leitlinie_THWildau.pdf

PID

A persistent identifier (PID) is a unique and permanent (= persistent) digital reference that consists of a defined combination of digits and/or alphanumeric characters. It can refer to a digital object (e.g. a data record), but also to persons and organizations. PIDs are also being developed for software, research instruments, data management plans, repositories or scientific conferences. In contrast to a URL, a PID does not refer to a location, but to the object itself, so that it can still be found even if the location changes.

A common standard for digital objects is DOI (Digital Object Identifier), another is URN (Uniform Resource Name). ORCID (Open Researcher Contributor Identifier) has been established for individuals and ROR (Research Organization Registry) for organizations. Another standard is GND (Gemeinsame Normdatei), which can be used to represent entities, i.e., persons, corporate bodies, conferences, geographies, subject headings and works.

Further information:

- <https://forschungsdaten.info/themen/veroeffentlichen-und-archivieren/persistente-identifikatoren/>
- <https://projects.tib.eu/pid-service/persistent-identifiers/persistent-identifiers-pids/>

Databases

Definition: Data can be collected and managed in a database. They are also suitable for storing large amounts of data in a structured and long-term manner. So-called database management systems (DMBS) enable the organization of data and the establishment of connections; complex queries are also possible. Databases are designed so that several users can access the data at the same time.

(Research) databases can enjoy copyright protection: "If research data is collected in a database, this can justify its own copyright protection and should therefore be taken into account by the project (see Section 4 of the Copyright Act). Under German law, databases are subject to specific protection, which grants the creators of the database the sole right to distribute and reproduce them (see Section 87b of the Copyright Act)." (Source: [forschungsdaten.info](https://forschungsdaten.info/themen/rechte-und-pflichten/urheberrecht/), <https://forschungsdaten.info/themen/rechte-und-pflichten/urheberrecht/>)

Repositories

Definition: Repositories are managed storage locations for storing digital objects, which are usually publicly accessible. A distinction must be made between institutional, discipline-specific and cross-disciplinary repositories.

Offer from (enter the university here) RADAR-BB / Recommendation for cross-disciplinary repository: As part of the "IN-FDM-BB" project funded by the BMBF and MWFK, RADAR-BB is being developed as a Brandenburg-wide repository for research data for which suitable subject-specific repositories do not exist.

Other cross-disciplinary repositories that have proven themselves in the FDM community:

- [Zenodo](#)
- [Figshare](#)
- [Dryad](#)

Overview: The service "Registry of Research Data Repositories" provides an international overview of repositories: <https://www.re3data.org/>

Repositories located in Germany can be found on the portal for research infrastructures: "RIsources": <https://risources.dfg.de/>

Find a repository – selection criteria

The quality features listed below help in selecting a suitable repository. They are intended to ensure that the research data is findable, reusable, citable and available in the long term:

- The data sets are given a persistent identifier such as DOI (Digital Object Identifier) or URN (Uniform Resource Name).
- It is possible to describe the data set in detail using metadata. Information on the origin/context of the data set can be provided in a text field or in a ReadMe file.
- The data sets are made public for Internet search engines by indexing them.
- The data set can be downloaded.
- The repository allows various access options to be set for the data set.
- Data sets can be provided with a license (e.g. Creative Commons).
- Ideally, a certificate for trustworthy archives is available, e.g. CoreTrustSeal or nestor seal for trustworthy digital long-term archives.

Data centers

Definition: "A data center is a central repository that handles the storage, management and distribution of research data and related information for a specific knowledge and/or memory institution." (Source: forschungsdaten.info, (<https://forschungsdaten.info/praxis-kompakt/glossar/#c269828>))

Overview: RatSWD offers a comprehensive overview of existing data centers (<https://www.konsortswd.de/angebote/forschende/alle-datenzentren/>)

Data journals

Definition: Data journals are publications with the main purpose of publishing data sets. In addition to the research data, a detailed description is published to enable its reuse.

Overview of interdisciplinary data journals:

- Data Science Journal: <https://datascience.codata.org/>
- Data in Brief: <https://www.sciencedirect.com/journal/data-in-brief>
- (MPDI) Data: <https://www.mdpi.com/journal/data>
- F1000Research: <https://f1000research.com/>
- Scientific data: <https://www.nature.com/sdata/>
- Patterns: <https://datascience.codata.org/>

Further information:

- https://www.forschungsdaten.org/index.php/Data_Journals
- <https://www.uni-wuerzburg.de/rdm/informationen/datenpublikation/>

Self-developed software

When publishing software, the FAIR Principles for Research Software must be observed in addition to the FAIR Principles. Furthermore, the handling of software along the research data lifecycle should be planned with a software data management plan (SMP).

- FAIR Principles for Research Software: <https://zenodo.org/records/6623556>
- Software management plan (SMP): <https://zenodo.org/records/1460504>

There are two different ways to publish software, each with its own advantages and disadvantages. Making software available via Github is widespread in the community. This has the advantage that repositories can be continually expanded. However, in some aspects this method does not comply with the FAIR principles. It is therefore advisable to publish software in a repository. This has the advantage that the software can be cited and published by assigning a persistent identifier (e.g. DOI). However, publication only represents one point in the software development process and the software should already be at a high level of maturity or be completed before it is published. A published data set should be well documented and contain the following elements:

- License-txt
- ReadMe for copyright
- Software as a folder with all necessary files
- Manual or technical report for software instructions

- When publishing software, particular attention must be paid to the awarding of an appropriate license. The criteria for selection are based on the intended use or subsequent use scenarios for the software. The following websites provide an overview of licenses: <https://ifross.github.io/ifrOSS/Lizenzcenter#open-source-lizenzen>
- <https://www.tldrlegal.com/>

Literatur suggestion:

- Biernacka, K. & Schulz, S. (2022). Forschungsdatenmanagement in der Informatik. Logos Verlag Berlin. <https://doi.org/10.30819/5490>

Data protection requirements and data access

When publishing data, the aim should be to achieve the most complete access to the data possible. The rule is: as open as possible, as closed as necessary. This principle applies particularly to data to which the GDPR applies. The access and usage conditions of the data must also be clarified here. All of these aspects can be clarified in the form of a DMP before publication.

There are a few things to consider when publishing research data that are related to personal data. The following points should be observed:

- The requirements of third-party funders and your own institution regarding data archiving must be checked and observed.
- Check whether data should be deleted.
- Data should be made available if possible.
- It should be clarified whether only limited subsequent use is possible, which results from the properties of the data (personal data).
- Then decide whether data should be published as a Public Use File (PUF) without restrictions or as a Scientific Use File (SUF) with restrictions.
- The location of publication must be clarified (repository or data center).
- Determine the time of publication.

Further Informationen:

- RatSWD [Rat für Sozial- und Wirtschaftsdaten] (2020): Handreichung Datenschutz, 2. vollständig überarbeitete Auflage, RatSWD Output 8 (6), <https://doi.org/10.17620/02671.50>
- Kienbaum, Janna; Fischer, Patryk & Paßmann, Sven (2023): Forschungsdatenmanagement bei personenbezogenen Daten - eine Handreichung, <https://doi.org/10.5281/zenodo.7428524>

Licenses

A license is a contractually agreed right of use. With it, the person holding the rights determines whether, how and under what conditions a work can be reused or used. There are commercial licenses that charge a fee for use and free licenses that allow free use. Free licenses allow anyone to use copyrighted content. However, there may

be restrictions on use or editing. Standardized licenses have been developed that reflect the frequently occurring conditions. The usual ones are:

- [Creative Commons \(CC\)](#), suitable for text, images and data
- [GNU General Public License \(GPL\)](#), suitable for software
- [Open Data Commons \(ODC\)](#), suitable for databases
- [Community Data License Agreement \(CDLA\)](#), suitable for data

The use of a free license is highly recommended because it makes the use of data much easier, as it eliminates the effort of obtaining permission from the person holding the rights in each individual case. A free license can be understood as a generalized permission with which the author informs the user that the work may be used in this way under these conditions. Licenses are usually issued by the repository. It is important to remember that once a license has been issued, it cannot be withdrawn! In addition, the requirements of your own institution and the sponsors with regard to the FAIR principles and open access guidelines must be observed. When reusing research data, the respective license requirements of this data must be observed. These are binding, which is why potential use must be clarified with the author in advance.

Further information:

- <https://forschungsdaten.info/themen/rechte-und-pflichten/forschungsdaten-veroeffentlichen/>
- https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/Infomaterialien/IN_Freie_Lizenzen_-_einfach_erklaert_Broschuere_2017_07_20.pdf

Contact

Counseling for research data management

Blanka Goßner

Forschungsdatenmanagement | Zentrum für Forschung und Transfer

Technische Hochschule Wildau | Hochschulring 1 | 15745 Wildau

Haus 13 | Raum 0.41

Telefon: +49 3375 508 322

Mail: blanka.gossner@th-wildau.de

Counseling on Application for third-party funded research projects

Zentrum für Forschung und Transfer, Mail: forschung@th-wildau.de