

# Guidelines for the publication and archiving of research data at the TU Braunschweig

TU Braunschweig offers its researchers the opportunity to publish and archive their research data on the [publication server](#) of the TU Braunschweig. Several points need to be considered, which should be described in more detail in the following.

## 1) Preparing your research data

### Short and concise

- Multiple files should be stored in hierarchically structured folders with meaningful naming.
- Folder names should also be comprehensible for third parties, i.e., systematic and content-related.
- The stored files should enable independent traceability and reusability.
- In the run-up to the publication, a conscientious data selection should be carried out and only data that are scientifically relevant and archived should be selected.
- The files should be available in standardized, widely used file formats for free usable programs (e.g. XML, CSV, TXT, TIFF).
- If data is available in a proprietary format (e.g. MatLab), it should also be stored in an open format if possible.
- The data must not be read-only.

### Folder structuring

A hierarchical structure is suitable for obtaining a clear structuring of research data that is as comprehensible as possible. How this works best depends on the data set and must be defined individually each time. With heterogeneous structure of the data (many different file types) one can create folder types, e.g.:

- Materials → Data, documentation, publications
- Data → Texts, photos, models, tables

If the data is more homogeneous, it can also be subdivided by time, place or method, e.g.:

- Project\_XY → Study\_Area\_01 → Measurements

It is important that you can find the desired file quickly and intuitively.

### File naming

When naming files, a naming convention should be defined right from the start, to which one sticks consistently throughout the entire project, especially when several parties are involved in a project. The naming should provide information about the content of the data and help identify them unambiguously. For this purpose, the creation date, a short description, the location, project name, sample number, analysis method or version number are particularly appropriate.

In addition, the following rules should be observed:

1. Filenames should be short but explanatory.
2. No special characters, spaces, or punctuation marks should be used.

3. Capital letters and underscores should be used instead.
4. Capital letters influence the sorting, so attention should be paid to consistent naming.
5. The best date format is YYYYMMDD.
6. If numbers are given, they should always be two-digit or even three-digit, so that the 0 serves as a placeholder (e.g. Sample01 and not Sample1).
7. The individual elements of the file name should be arranged in such a way that the file is found as quickly as possible, name beginnings such as "test" or draft" should be avoided.
8. If different versions of a file exist, a V with numbering should be used to keep order between the versions (V01, V02, etc.).
9. Repeating information from folder names in filenames should be avoided.

### Useful elements for file naming

Meaningful elements	Notes and example
Name of author or initials	LG
Content/technical description	Keywords
Date of creation especially for version control	YYYYMMDD
Processing status	Original, design, part
Version	V02

### Examples of file names

Well	Bad
20200813Agenda	13Aug2020Agenda
2020UBWorkshopFDM	The_2020_Workshop_with_the_topic_Research_Data Management_of_the_UB!
20200804_ProjectA_Protokoll_LG	Protocol Laura

Detailed information: [http://www.data.cam.ac.uk/files/gdl\\_tilsdocnaming\\_v1\\_20090612.pdf](http://www.data.cam.ac.uk/files/gdl_tilsdocnaming_v1_20090612.pdf)

The naming convention of the files and folder structure should be explained in the attached README file of the record, if abbreviations are used.

### Data selection

In order to decide which data are suitable for archiving/publication, they can be divided into the following categories.

**Category 1: Very valuable research data**

Research data that

- are unique or very difficult to restore, and
- are of interest to a large community and/or for a wide range of subsequent uses.

➔ Definitely archive / publish!

**Category 2: Limited valuable research data**

Research data that either

- can be reproduced with reasonable effort or
- are only interesting for a short time, a small target group or only for the proof of research results.

➔ Archive/publish if effort and cost do not argue against it.

**Category 3: Research data of little value**

Research data that is

- created only for testing purposes,
- not readable or not sufficiently described.

➔ Data are not suitable for archiving/publication.

**Recommended file formats**

Not every file format is suitable for long-term archiving. To ensure that research data can be read and used even after several years, open, non-proprietary file formats are recommended.

	<b>Recommended formats</b>	<b>Unsuitable for archiving</b>	<b>Conversion possibilities</b>
<b>Text formats</b>	PDF/A, TXT (UTF8 encoded), XML	DOC, PPT	- Word & PowerPoint to PDF/A-1b - LaTeX & TeX to PDF/A
<b>Tables/ Spreadsheets</b>	CSV/ TVS	XLS	Tables to ASCII text files (CVS)
<b>Statistical Environments</b>	SDD, R files, Matlab files in HDF format		- Save Matlab files as v7.3 MAT (HDF5 based standard) - Save important tables in the workspace separately as CSV
<b>Raster graphics</b>	TIFF, PNG, JPEG2000	EPS	
<b>Vector graphics</b>	SVG		
<b>Multimedia</b>	WAV, AVI, MOV, MPEG-4	WMV	
<b>Geodata</b>	GML, KML, ESRI Shapefile, Georeferenced TIFF		
<b>Numerical data</b>	NetCDF		

Detailed information:

<https://www.ianus-fdz.de/it-empfehlungen/dateiformate>

<https://uni-tuebingen.de/forschung/forschungsinfrastruktur/escience-center/forschungsdaten-archivfaehigkeit/#c1168492>

## 2) Preparing your documentation

### Short and concise

- The documentation is intended to provide a comprehensive description of the published data so that the data can be reproduced and reused.
- The dataset should be accompanied by a README file (TXT or PDF/A) with descriptive metadata so that interested scientists can better understand and reuse the dataset.
- The README shall include at least the following information:
  - Title of the data publication
  - Authors, creators
  - Institution and project
  - Year or period of origin
  - Abstract/ Description of the dataset
  - Method/ Data collection
  - Reference to other publications describing/evaluating the dataset
  - License
- Metadata information is also required for submission to a repository. The more precise these are, the better the publication can be found and understood later on.
- Additional material such as laboratory book entries, software code and other accompanying materials should also be provided if needed for understanding.

### Structuring of the README files

For very heterogeneous data sets (e.g. a mixture of geodata, Matlab scripts and tables with measurement data), (1) all information can be described in one README file, referring to individual data specifications or (2) a "main" README can be created, that describes the relationship of the entire data set and additionally specified README files on sub records or individual files can be used. For several README files, care should be taken that the naming of the README file provides information which record/file they describe. Attention should be paid to uniform formatting of the README files.

### README template

We have created a template for README files. If you are interested in using them, please contact us!

## 3) Digression on research software

### Short and concise

- Software should be versioned (for example with the help of GitVersion).
- Software (here program in executable state) should always be provided with an associated source code.
- The purpose of the software should be formulated briefly but in a meaningful way, for example in an attached README file.

- The code should be modularized to get a better overview.
- Each function should show what its intention is (for example, by self-explanatory names or comments).
- The software should be accompanied by a test data set, which can be used to quickly check whether the program is running properly. Even better would be to develop the software directly test-driven.

### **Publication of research software**

Software plays an increasingly important role in research. It not only helps in the analysis of data, but is often also crucial for its production. Often you need the underlying software to understand research data correctly and, above all, to be able to reproduce it. In addition, self-developed research software can represent an independent research result. Therefore, research software such as research data should be published for reusability.

### **Metadata for research software**

Research software needs its own metadata for description and comprehensibility. An example of a metadata standard is [CodeMeta](#).

### **Helpful tools to make research software executable**

ExecShare: <http://www.execandshare.org/CompanionSite/>

jupyter notebooks: <http://jupyter.org/>

Docker container: <https://www.docker.com/>

### **Archiving of research software**

The long-term preservation of software is difficult. On the one hand, it must be described in a comprehensible manner and on the other hand, it must be ensured that it is kept executable. This is a problem, because software is often highly dependent on the computing environment in or for which it was developed, i.e., the operating system, hardware or software libraries. These change very quickly with advancing technology, so that developed software can no longer be read or executed. There is no final solution to this problem. One approach is the use of emulation, in which, in addition to the actual software, the executing computer and operating system are also simulated.

### **Further literature**

Matthias Katerbow, Georg Feulner et al. (2018): Handreichung zum Umgang mit Forschungssoftware, <http://doi.org/10.5281/zenodo.1172970>

## **4) Clarify legal aspects and choose a license**

### **Short and concise**

- All legal (copyright/ exploitation/ usage rights, data protection, service agreements) and ethical provisions must be complied with.
- The choice of a suitable license (e.g. Creative Commons) is also essential for the desired type of re-use.

- The following applies: As open as possible, as closed as necessary.
- Observe the [contract of use](#) of the research repository of the TU Braunschweig.

## Copyright

Intellectual creations and services are protected by copyright. If a copyright exists, the use, distribution and processing of the work is only possible with the permission of the author. Research data is rarely subject to copyright, but there may be exceptions. Often, however, published research data are provided with licenses that clarify a possible subsequent use. Moreover, databases are an exception and, according to German law, they enjoy specific protection, whereby the authors have the sole right to distribute and reproduce (§ 87b UrhG). In general, copyright protection ends 70 years after the death of the author. According to this, the works are considered public domain.

Descriptive metadata of research data are generally not protected by copyright, as they are usually relatively short, purely descriptive representations.

In addition to a copyright, research data can also be subject to patent law in certain cases if the research data describes technical teaching. In such cases, only the patent proprietor may use the patented data.

## Contractual requirements

Open Access Guideline of the TU Braunschweig: [https://www.tu-braunschweig.de/fileadmin/Redaktionsgruppen/Einrichtungen/UB/PDF/OA/Leitlinie\\_zu\\_Open\\_Access\\_an\\_der\\_TU\\_Braunschweig.pdf](https://www.tu-braunschweig.de/fileadmin/Redaktionsgruppen/Einrichtungen/UB/PDF/OA/Leitlinie_zu_Open_Access_an_der_TU_Braunschweig.pdf)

Guideline for the handling of research data of the TU Braunschweig: <https://www.tu-braunschweig.de/fileadmin/Redaktionsgruppen/Einrichtungen/UB/PDF/forschungsdatenleitlinie.pdf>

It is generally necessary to pay attention to what is recorded in the employment contract regarding the rights of use (usually belonging to the employer) of generated research data and whether, for example, there is a trade secret.

Sponsors of research projects can also have claims, for example, when it comes to the rights to created software in the course of the research project. In case of doubt, it is advisable to inquire prior to publication in order to obtain consent for distribution if necessary.

## Data protection

When handling personal data, various aspects of data protection law must be taken into account. Data protection restrictions can be avoided by anonymisation. Anonymization means the alteration of personal data in such a way that the individual data concerning personal or factual circumstances no longer exists or only with a disproportionate amount of time, costs and labor power can be assigned to a specific or identifiable natural person (§ 3.6, BDSG). There is also the possibility of pseudonymisation. This describes the processing of personal data in such a way that the data can no longer be attributed to a specific subject without the use of additional information; if this additional information is kept separately and is subject to technical and organizational measures that ensure that the data cannot be assigned to a data subject (§ 46 Abs.5, BDSG neu).

If a personal reference exists, research data may only be collected and further processed with the signature of a declaration of consent of the subjects.

Samples for declarations of consent can be found here:

<https://www.econstor.eu/dspace/bitstream/10419/97181/1/785263330.pdf>

The area of privacy policy is a wide field. The TU Braunschweig has a data protection officer who is available for any questions about data protection (<https://www.tu-braunschweig.de/datenschutz>).

## **Publication agreements with the repository/ archive**

User agreement of the publication server of the TU Braunschweig: <https://publikationsserver.tu-braunschweig.de/content/publish/contract.xml>

Repositories establish user agreements for the publication/ archiving of research data. This should be read well in advance and it should be ensured that all specified requirements are met.

## **Agreement of research data re-use**

### Access authorization

It is recommended to make research data available as openly as possible. However, it is also possible to place an embargo on the data (delay access for a certain period of time) or restrict who can access the data. Sensitive research data, e.g., containing personal information that cannot be fully anonymized, would be a case for such a restricted access.

The TU Braunschweig publication server offers the following opportunities:

1. Publication (unrestricted access).

In this variant, there are no restrictions on access to the data. Everyone can view and download the data. By choosing a suitable usage license, it is possible to determine how the data may be subsequently used.

2. Embargo (delayed access)

An embargo can be placed on the publication of the data set. Until the end of the embargo, the dataset is not available and not accessible for others. Only after the self-selected embargo period ends, the dataset will be published. The use of an embargo can be useful if the publication of research data is to be delayed, e.g. until the publication of a related journal article or for commercial reasons when applying for a patent.

3. Publication (restricted access)

There are different types of access conditions. For example, the publication server allows that only members of the TU Braunschweig can view the research data. If this is not enough, restricted access can also be selected, where only the metadata of the dataset can be viewed, but not the dataset itself. If a researcher is interested, they can contact the author and he or she can decide individually if the researcher should have access to the dataset. This can help if legal or ethical reasons do not allow making the dataset freely available. In addition, it is also possible to clarify why the interested researcher wants to use the data.

#### 4. Archiving

Of course, it is also possible to archive research data according to good scientific practice for at least 10 years without making it public. In this case, only UB staff can view the dataset to manage it.

#### Licences for data

The Creative Commons (CC) licenses are recognized worldwide and, as of version 4.0, are also suitable for research data in an international context (<https://creativecommons.org/>). It can also be used in an international context to clarify in a simple way how a research data set can be used, so that other scientists are informed about a possible type of subsequent use.

License	Abbreviation	Description	Open access compliant
	CC0	No rights reserved, public domain	✓
	CC BY	Attribution - unrestricted processing and commercial use permitted	✓
	CC BY-SA	Attribution + equal licensing (SA = share alike) - unrestricted further processing and commercial use under equal licensing	✓
	CC BY-NC	Attribution + non-commercial - unrestricted re-use but not for commercial purposes is allowed	✗
	CC BY-NC-SA	Attribution + non-commercial + equal licensing - unrestricted redistribution for non-commercial purposes is permitted under the same licensing	✗
	CC BY-ND	Attribution + no derivatives - commercial use is permitted, but no processing	✗
	CC BY-NC-ND	Attribution + non-commercial + no derivatives - further processing and commercial use are not permitted	✗

In order to prevent the misuse of research data, it is also possible not to make the data set immediately accessible, but to first request contact with the author, so that he can allow a release in the case of legitimate scientific interest. In this case, only the descriptive metadata file is initially freely accessible.

**ATTENTION:** A given license cannot be effectively withdrawn!

#### Licenses for software

For created software there are own open source software licenses. Common licenses are:

- GPL: GNU General Public License
- LGPL: GNU Lesser General Public License
- BSD: Berkeley Software Distribution
- WITH: MIT license

A detailed list of open source licenses can be found here: <https://ifross.github.io/ifrOSS/Lizenzcenter>

[Help for choosing the right license](#)

<https://creativecommons.org/choose/?lang=en>

<https://choosealicense.com/>

### **Further links and literature**

Helpful flowchart "Publish research data" by forschungsdaten.info:

<https://www.forschungsdaten.info/themen/veroeffentlichen-und-archivieren/daten-publizieren/entscheidungshilfe-daten-veroeffentlichen/>

General information on the topic:

<https://www.forschungsdaten.info/themen/rechte-und-pflichten/>

[https://tu-dresden.de/gsw/jura/igetem/jfbimd13/ressourcen/dateien/dateien/DataJus/DataJus\\_Zusammenfassung\\_Gutachten\\_12-07-18.pdf?lang=de](https://tu-dresden.de/gsw/jura/igetem/jfbimd13/ressourcen/dateien/dateien/DataJus/DataJus_Zusammenfassung_Gutachten_12-07-18.pdf?lang=de)

## **5) Publish**

Are all the previous points considered and implemented? Then nothing stands in the way to archive/ publish the research data. The data can be submitted either via Shibboleth and the TU Braunschweig identifier on the [publication server](#) or one of the UB employees ([forschungsdaten@tu-braunschweig.de](mailto:forschungsdaten@tu-braunschweig.de)) can offer support (e.g. for larger data volumes).

After the submission of the data, it is checked by the UB employees and if everything is in order, the research data will be archived or provided with a DOI and published.