TECHNISCHE UNIVERSITÄT BERLIN

Guidelines

on the implementation of the Research Data Policy of the Technische Universität Berlin

July 2023

SZF
Servicezentrum Forschungsdatenmanagement

On 23 October, 2019, the Academic Senate of TU Berlin adopted the Research Data Policy of the Technische Universität Berlin. It has been updated due to the decision of the executive board on 15 March 2023. With this policy, TU Berlin wants to provide its current and future researchers with an orientation for handling research data. The *Guidelines on the implementation of the Research Data Policy of the Technische Universität Berlin* provide practical advice for the Research Data Management (RDM) in the different phases of a research project. The Academic Senate agreed to the Guidelines on 23. October, 2019. They are maintained and continuously updated by the Research Data Management Service Center.

Contents

Abbreviations	
DFG	German Research Foundation
DMP	Data management plan
Dept. V	Research and Technology Transfer Department
FAIR	Findable, Accessible, Interoperable, Reusable
RDM	Research data management
UB	University Library
SZF	Service Center Research Data Management
ZECM	Center for Campusmanagement

Central Point of Contact at TU Berlin

The **Service Center Research Data Management** is the central point of contact for all issues related to RDM at TU Berlin.

Within the Service Center Research Data Management (SZF), the University Library (UB), the Center for Campusmanagement (ZECM), and the Research and Technology Transfer Department (Dept. V) cooperate and bundle their competences to support the University's researchers in handling research data. SZF is managed and coordinated by the University Library. SZF operates the research data infrastructure at TU Berlin, which is integrated into the University's IT infrastructure. The central technical infrastructure services include *DepositOnce*, the repository for research data and publications of TU Berlin, and *TUB-DMP*, a web tool for creating data management plans (DMPs). A wide range of training and advisory services complements the technical services.

RDM and the research data infrastructure of TU Berlin are aligned with the FAIR principles, international guidelines for the optimal management of research data so that the data is findable, accessible, interoperable, and reusable.

The SZF website serves as a central platform for RDM at TU Berlin. Here, you will find comprehensive information on the RDM services at TU Berlin, including the respective contact persons.

Corresponding to the task sharing of SZF, there are **distributed responsibilities** along the research data life cycle (see figure).

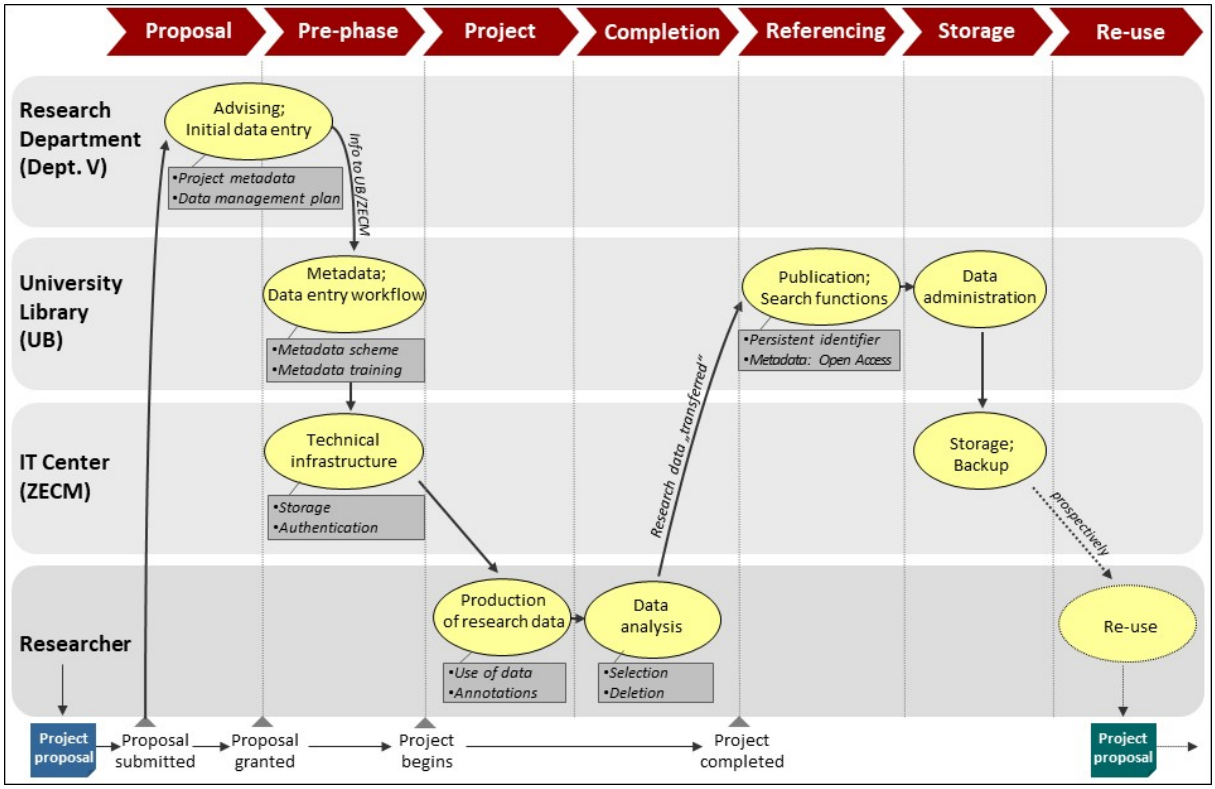

Fig. 1: Distributed responsibilities in RDM of TU Berlin

Guidelines on the Handling of Research Data

Early planning of your research data management not only has the advantage of creating a transparent framework for the uniform handling of data in a project. It is also important in terms of possible costs for RDM. You can apply for funding for staff (e.g. for the development and administration of research infrastructure) or even the infrastructure itself (e.g. INF project in Collaborative Research Centres).

In Advance: Ten Questions about RDM during the Proposal Phase

1. Are there any guidelines for the handling of research data (e.g., from the funding organization, the research institution, the project partner or the discipline)?
2. Have there been assigned responsibilities for research data management in the project?
3. Is there a data management plan for the project?
4. Will you use the IT infrastructure of TU Berlin for storage? Did you already contact ZECM on this?
5. Will you use third-party research data? Will you need funding to access this data?
6. Do you plan to use external repositories for archiving and/or publication of research data? If so, which ones?
7. Is there a publication strategy for research data and publications in your project (e.g., open access)?
8. Will you need funding for archiving and/or publishing the research data?
9. Will you apply for staff to manage the research data and/or to administrate the technical infrastructure?
10. Is it ensured that the staff is trained in the handling of research data?

I. Planning and Proposal Phase (before the Research Project)

- In case of a **third-party funded project**, inform yourself in advance of any applicable funding guidelines regarding research data. Costs arising for RDM – e.g., human resources for data processing or for the development of project-internal workflows, publication costs or resources for long-term archiving of the data that extend beyond the University's basic infrastructure - can and should be part of the funds applied for. Use the advisory offers of Dept. V to learn about the different requirements of funding organizations and funding possibilities and to determine your individual needs.
  - → DFG and BMBF proposals: Research Promotion Section
  - → Research Training Groups proposals: Center for Junior Scholars
  - → EU proposals: EU Office
  - → Contract negotiations in research and development projects: Center for Intellectual Property
- In all research projects and also in the publication of the research results **legal frameworks** must be observed and preferably are to be settled in advance. Certain research data, for example in the social or life sciences, are subject to strict requirements, e.g., on **data protection** or **ethics**. Copyright and the protection of third-party interests must also be ensured.
  - → Advice on Copyright and Patent Law: Center for Intellectual Property
  - → Advice on Data Protection and Ethics in general: Data Protection Team
  - → Advice on ethical issues in the faculties:
    - Ethics Committee of Faculty III
    - Ethics Committee of Faculty IV
    - Ethics Committee of the Department of Psychology and Ergonomics (Faculty V)
    - Ethics Committee of Faculty VI
    - Ethics Committee of Faculty VII
- In order to receive adequate support for your new research project, it is recommended to announce the project during the proposal phase to Dept. V. This project announcement is achieved by using the electronic project notification (ePA).
  - → Information on the electronic project notification: Research Promotion Section

- **Research data management**

  For any project in which data is collected or forms the foundation of your research it is strongly recommended that you make an early examination of the requirements and possibilities of efficient and sustainable research data management. Already during the proposal phase, a **strategy for the sustainable archiving and availability** of the research data should be determined. Also, the legal status of the data and suitable security measures for data use during and after the research project are to be defined.

  These definitions should be included in a **DMP**. A DMP documents the process how the research data is generated and how it is stored appropriately so that in later years it can be interpreted and verified and remains available, authentic, citable, and reusable. To optimize RDM and as a basis for institutional assistance, a DMP should be created before the start of the research project and updated over the course of the project (*living document*).

  Almost all funding organizations require a DMP to be submitted as part of the project proposal. As support for your DMP, you can use the web tool TUB-DMP. It contains templates with relevant questions, which you can answer in a step-by-step workflow as they pertain to your project.

  - → Advice on RDM and TUB-DMP: SZF

II. Implementation Phase (during the Research Project)

- State of the art processes are to be used for the **storage and processing** of research data as well as for collaboration based on this data. This particularly includes compliance with data security as regards the availability, integrity, and authenticity of data. This requires, for instance, the use of data backup and secure data exchange platforms.
- In the implementation phase of a project, datasets may evolve several stages (e.g., by selection, aggregation, integration). It is a good practice to label, document and keep the different **versions** at least for the duration of the project. Especially in case of text-based data, the use of versioning tools, as commonly used in software development (e.g., GitLab, SVN), helps with the management of different versions.
- ZECM provides the following services for research work at the University; further information can be found on the ZECM website.
    - Use of network file systems (incl. data backup)
    - Archive storage services on tape drives
    - Provision of virtual root servers (server hosting)
    - Accommodation of real servers (server housing)
    - Block storage services for servers (virtual hard disks over a dedicated storage network)
    - Data exchange services
    - Versioning services

  These services are either included as basic equipment and are free of charge or offered at cost price. The same applies to the collaborative working tools provided by ZECM.
- Different scientific disciplines and their research domains apply different methods in handling research data. This makes comprehensive recommendations for the concrete procedures hard to define. Therefore, it is generally recommended to become acquainted in advance with the established **data formats, software, and standards** that are used in your scientific community for the documentation and annotation of research data (e.g., ontologies, controlled vocabulary, or metadata schemes). Using open, nonproprietary file formats supports the access to and long-term availability of research data.
- Describing research data with **metadata** is fundamental for the reusability of research data. Metadata is data about data and describes the context in which the data was created. As a rule of thumb, metadata should answer the classic six questions: Who? What? Why? How? When? Where? Metadata are a prerequisite for enabling potential subsequent users to find data and assess its suitability for the intended use. Ideally, the description is structured and machine-readable. For this purpose, metadata standards and standardized terminologies exist in most disciplines. If these do not exist, generic standards, such as Dublin Core, should be used to describe the data. They are developed and promoted by worldwide initiatives and help to make research results better findable and interoperable.
- In collaborative projects or projects with large amounts of data, the use of dedicated **work environments** and portals for data management is advisable. Operating these infrastructures usually requires additional resources, but they provide the advantage of a uniform and central management for research data. Finding and sharing research data is thereby facilitated, but should be governed within the project consortium by a project-specific **data policy**.

    → Advice on technical infrastructure and IT services: ZECM

    → Advice on metadata and metadata standards: SZF

III. Final Phase (after the Research Project)

- According to good academic practice, by the end of the project research data is to be stored and, if possible, made accessible, if there are no contradictory contractual, ethical, or legal regulations. Many funding organizations now place particular importance on **accessibility**, in order to enable the verification of the research results and the reuse of the research data. In keeping with its Research Data Policy, TU Berlin supports open access to research data. When publishing research data, TU Berlin recommends to follow the principle, “Accessible, if possible, restricted if necessary”.
- The following basic principles regarding the **publication of research data** should be observed:
  - Individual websites (e.g., of projects, working groups, academic chairs, employees) are generally not a suitable location for the publication of research data. The long-term availability of such websites is often not ensured and the unique identification (keyword: persistent identifier) is usually not possible.
  - When selecting data to be published, the DFG recommends: "Research data should be made accessible at a stage of processing that allows it to be usefully reused by third parties (raw data or structured data)." In particular, data which forms the basis of a scientific article should be made accessible, if there are no contradicting data protection, legal or research ethics regulations.
  - As is the case for scientific articles, research data should be assigned a unique persistent identifier (PID) upon publication. In this way, research data can be found and cited independently of a publication. Well known examples are DOI (digital object identifier) or URN (uniform resource name).
  - To regulate the rights of use and utilization of research data, data should always be published with an appropriate license. The choice of a license should at least allow open access for scientific purposes. Any special requirements of the funding organization or of repositories are to be observed. Established free licenses in software are the GNU General Public License (GPL), MIT License or the Apache License. Creative Commons licenses are standard for texts, images, music and videos.
- As a member of TU Berlin, you and your cooperation partners can use the interdisciplinary repository of TU Berlin, DepositOnce, to publish your research results (research data and publications). Research results, meaning consolidated data and all information needed to reproduce these results (such as notes, time histories/recordings, calculations, etc.) are stored in DepositOnce. Pursuant to the Statute on the Safeguarding of Good Academic Practice of TU Berlin, research data is stored for at least 10 years.
  - All data in DepositOnce is provided with metadata (standard format Extended Dublin Core).
  - All datasets automatically receive a persistent Internet address (DOI).
  - Various free licenses can be assigned to the datasets.
  - Via the DOI, related research data and publications in DepositOnce can be linked to each other and then refer to each other.
- In accordance with the rules of good scientific practice, published research data can no longer be modified in DepositOnce. This maintains the data's citability and verifiability. DepositOnce utilizes a versioning in which new versions can be published while previous versions remain available. Every new version receives a new DOI; previous and current versions are automatically linked to one another and refer to each other.

  DepositOnce is committed to Open Access. The metadata are publicly accessible on the Internet and are broadly distributed and made searchable via standard interfaces (Google Scholar, etc.). An embargo can be placed on the research data itself.
- As part of various initiatives and projects in which TU Berlin researchers are also involved, there were built and are still being built discipline-specific research data infrastructures for many disciplines. Meanwhile there is a large number of **discipline-specific repositories**. These may offer advantages compared to DepositOnce, such as discipline-specific metadata schemes and specific search options. If you already use a repository in your scientific community, you should continue to do so. The same applies if it seems reasonable to publish the research data in a discipline-specific repository and such a repository exists for your discipline.

In some disciplines, it is common to publish data as a supplement to the respective article. However, this form of data publication has the disadvantage that the data can only be found via the article and does not form an independent, citable publication object.

When choosing a discipline-specific repository, the following criteria should be observed: long-term availability (at least 10 years), allocation of persistent identifiers (e.g., DOI, URN), licenses and usage rights of the data, reputation and visibility, costs. The portal re3data.org offers a helpful overview with comprehensive search and filter functions when you are searching for a suitable discipline-specific repository for your research data.

➢ Research data is to be **published as soon as possible**. If applicable reasons exist, an embargo can be placed on data in DepositOnce. In this case only the metadata are published; the data itself is stored in the repository and is only visible after expiry of the embargo. Interested persons can request the data via email during the embargo. The embargo is determined by the responsible researchers whereby the requirements and guidelines of research funding agencies and repositories must be observed. Embargo periods should not exceed a maximum of 5 years after the project end. An embargo must be justified, for example in a file in the repository which also includes the expiration date of the embargo.

→ Advice on the publication of research data, licenses, DepositOnce/repositories: SZF