

Research Data Management Guideline (RDM Guideline)

Content

§ 1	Data Management Plan	2
§ 2	Data backup	3
§ 3	Publication	4
§ 4	Archiving	4
§ 5	File naming	5
§ 6	Research software	5
§ 7	Experimental data	7
§ 8	Responsibilities	8

Preamble

Motivation

Research data management (RDM) establishes a framework for organizing research data, preventing data loss, and facilitating access to research findings. Furthermore, RDM, along with comprehensive documentation of research outcomes, enhances their durability, scientific reliability, and supports their reuse and reproducibility. In essence, RDM greatly enhances the quality and efficiency of research, fosters collaboration among researchers, and significantly boosts the discoverability and citation frequency of research findings.

Scope

The Research Data Management Guideline (RDM Guideline) applies to all research-related activities and is mandatory for participants in the SPP100+ program, hereinafter referred to as researchers. Likewise, this guideline also extends to associated members whose research directly aligns with the objectives of the SPP100+ program. The enforcement of this guideline will begin on May 23, 2023. In the context of this guideline, research data encompasses all data that are either generated during the research process (such as through measurements or simulations, including source code) or preexisting data that are reused and serve as a fundamental foundation for the research outcomes.

Validity of higher-level guidelines

The RDM Guideline requires compliance with the "Verwendungsrichtlinien - Bedingungen für Förderverträge mit der Deutschen Forschungsgemeinschaft e.V. (DFG) über Graduiertenkollegs" (see https://www.dfg.de/formulare/2_22/, translation "Usage Guidelines–

Conditions for Funding Agreements with the German Research Foundation e.V. on Research Training Groups"), the "Leitlinien zur Sicherung guter wissenschaftlicher Praxis" (translation "Guidelines for Safeguarding Good Research Practice") and all other applicable DFG guidelines. In case of doubt, the regulations laid down in the currently applicable DFG guidelines take precedence over these RDM Guideline. Accordingly, the " Guidelines for Safeguarding Good Scientific Practice, Avoiding Scientific Misconduct and Dealing with Violations " Provided by the Technical university of Dresden" (see <https://tu-dresden.de/forschung-transfer/wissenschaftlicher-nachwuchs/ressourcen/dateien/Good-scientific-practice-TUD.pdf?lang=en>) must also be observed.

§ 1 Data Management Plan

A Data Management Plan (DMP) is an extensive and structured document that serves as a template created within the coordination project for managing research data to be used by each sub-project within SPP100+ program. It outlines the procedures and strategies for handling and organizing data throughout a research project. The DMP is typically stored and archived together with the research data itself. It specifically focuses on the data that is processed or generated during the research project and should be created at the project's initiation. This ensures a consistent and coherent approach to Research Data Management (RDM) practices from the beginning to the end of the project. In situations where certain information is initially unavailable, it is acceptable to provide estimations or propose tentative approaches. However, it is important to clearly indicate such information as provisional or preliminary. The DMP can be reviewed and modified as necessary during the project's course, with a recommended minimum review frequency of every six months. Any modifications made to the DMP should be thoroughly documented and communicated accordingly.

The DMP should be concise and easily understandable while still providing comprehensive information. To achieve this, the DMP should include, at a minimum, the following details:

- Current processing status and the individual responsible for managing the project.
- Project information: project name, project team member(s), affiliated institution, project duration, and a brief description of the project.
- Details about the generated and reused research data: This includes information on data formats, types, and volumes, the method of data generation, tools utilized (instruments, hardware, software, etc.), sources of reused data, data organization, and documentation of the research process.
- Data storage and backups: Specify the location of the research data, the backup strategy (number and location of backups, intervals, etc.), and any access restrictions in place.
- Ethical and legal considerations: Address ethical and legal aspects related to handling, archiving, and publishing sensitive data, including copyright, data protection, security measures for handling sensitive data, and access to data by third parties.
- Archiving: Provide a brief description of the archiving process, including the formats, volume, and location of the archived research data. Discuss measures taken to ensure long-term data availability, additional information on metadata to provide context for the data, legal and infrastructural requirements for archiving, and associated costs.



- **Publication:** Specify any published research data, including the publication location, degree of compliance with FAIR principles (Findable, Accessible, Interoperable, and Reusable), assigned licenses, access restrictions for published data, and other means by which data are made available to third parties.
- **Resources and responsibilities:** Define the responsibilities and tasks within the project, including the specific role in Research Data Management (RDM). Clarify access rights to data, provide contact information for data-related inquiries (preferably beyond the project duration), and identify the necessary resources (costs, personnel, materials, etc.) required for successful DMP implementation.

§ 2 Data backup

To ensure the protection of research data, it is of great importance to create regular backups of valuable data. Valuable data refers to both raw and processed data that serve as the foundation for intended or already published publications, or data intended to be shared with the scientific community through appropriate platforms. In this regard, it is recommended to create a backup of all valuable research data on the SPP100+ research data management system (SPP100+ Server) at least once a year, in conjunction with the annual report. However, it is advisable to perform backups at more frequent intervals. The access permissions on the SPP100+ Server are managed by the designated person responsible as outlined in § 8. It can be beneficial to transfer the project structure utilized on individual hard disks to the SPP100+ Server while reducing as possible the folder contents. Additionally, the annual reports should include a section detailing the uploaded data and an evaluation of its significance.

Furthermore, it is essential to upload the research data to the servers of the respective institutes. It is highly advisable for each institute to establish a backup strategy that includes daily backups. In cases where backup strategies are not yet in place within the institutes, it is recommended to contact the person responsible for Research Data Management (RDM) as specified in § 8. Collaboratively, suitable solutions can be explored to address this backup requirement.

§ 3 Publication

According to the ("Verwendungsrichtlinien – Bedingungen für Förderverträge mit der Deutschen Forschungsgemeinschaft e.V. (DFG) über Graduiertenkollegs), the following obligations and recommendation are outlined:

- Project findings resulting from the funding provided by the Research Association must be made accessible to the general public in an appropriate manner.
- Publications and public presentations must include a reference to the funding received within the Research Association. Only the specified spellings mentioned in the currently applicable. Usage Guidelines should be used. The project number for the Research Association is.
- Research results funded within the Research Association should ideally be published digitally and made freely accessible on the internet (open access).

The Research Data Repository of TU Dresden which serves as Open Access Repository and Archive (OpARA) is suggested as an electronic repository for digitally publishing research results and data. If an alternative channel is chosen for the primary publication of research findings, it is recommended to make a secondary publication through the TU Dresden Research Data Repository. To facilitate this, a request for secondary publication can be submitted to the TU Dresden University Library OpARA concurrently with the initial publication.

Moreover, it is advisable to publish the associated research data or, at the very least, their metadata through the TU Dresden Research Data Repository alongside the actual research findings. When publishing research data, it is important to provide all necessary information to ensure optimal reproducibility and traceability of the results. This requirement encompasses details regarding the usage of research software and the generation and utilization of experimental data. In §§ 6 and 7, the guidelines specify the essential information that should be provided for these two scenarios. It is also emphasized that this information must be documented in a text file located at the top level of the directory that must have the name Metadata.txt. These text files that contain the metadata information would be collected in automatic based process and further stored in a database.

The research results must be provided with a license that regulates the options for use of the data by third parties. When choosing the license, the principle of "as open as possible and as closed as necessary" should be followed. As well as the selection of a Creative Commons license or another individual license is enabled when publishing the individual data set for example at TU Dresden OpARA or any other digital repository.

§ 4 Archiving

The archival of research data is a compulsory requirement. The data set should be archived without delay, at the latest following the publication associated with it. Ideally, the archiving process should occur as soon as the data set is considered complete and finalized independently.

The documentation of data archiving is an essential component of the Data Management Plan (DMP), and it should be regularly updated to reflect any changes. § 1 of the guidelines outlines the necessary information that the DMP should include for archiving purposes. Additionally, if legal considerations need to be taken into account during the archiving process, they should be documented as well. It is essential to maintain accurate and up-to-date records in the DMP to ensure proper documentation of the archiving activities.

When archiving copyright-protected data, including information regarding the copyright status and any relevant copyright restrictions is highly required. This information should be documented alongside the archived data to ensure compliance with copyright laws and regulations. It is important to provide clear and accurate information about the copyright status of the archived materials to facilitate proper usage and dissemination while respecting intellectual property rights.

Furthermore, the Data Management Plan (DMP) should include the specification of a contact person for the archived research data. Ideally, this contact person should be available throughout the entire archiving period. To maintain contact with the researchers who generated the data set, even after their affiliation with the research institution ends, it is recommended to involve the secretariats of the research institutions as a point of contact. The secretariats can internally manage the communication with the researchers. Alternatively, using an ORCID ID can be a useful way to identify authors and contributors of the archived data.

§ 5 File naming

The objective of adopting a consistent file naming convention is to facilitate efficient file retrieval and organization, not only for published and archived datasets but also for local datasets. It is advisable to employ a systematic naming approach for local datasets as well. As a guideline, the directory structure should not exceed ten folders in width and three folders in depth. Both folder and file names should be concise and clear, avoiding any ambiguity. It is recommended to limit the length of designations to a maximum of 50 characters per name.

File naming should follow the following pattern, using underscores instead of spaces:

- Date (YYYYMMDD) + short description (project dependent) + author (abbreviation)
→ Example: 20210406_FileNaming_RDM_FL.doc

Moreover, it is worth mentioning that:

- Do not use umlauts (ä, ö, ü)
- Consciously use upper-and-lower case letters
- Select language according to target group (default English)
- Label different versions as V01, V02, etc.
- Add the version number of the research software (if necessary)

At the first folder level, it is advisable to create a README file that provides explanations for the file naming conventions and abbreviations used and how the folder is structured in width and depth. This will ensure consistency and facilitate understanding of the file structure and naming conventions within the dataset.

§ 6 Research software

To guarantee the long-term usability of research software projects that are utilized for generating, processing, or analyzing research data, it is important to take into account the following aspects:

- It is advisable to prioritize the use of open-source software when developing research software in order to minimize reliance on commercial products.
- The implementation of basic guidelines for good programming should be observed as far as possible. This includes maintaining a clear program structure and writing code that is well-commented and easy to understand.

- Tests play an essential role in research software by validating its behavior. These tests, integrated within the source code, enable the examination of different aspects of the software's correctness and performance. To guarantee its continued viability, it is important to execute these tests automatically and consistently over time.
- If research software is being developed or improved as part of SPP100+, it is necessary to store and version. The latest status of the software project should be regularly uploaded and updated on the Server.

Structure of the folder that contains the software

The structure and components of the repository may vary from project to project and are therefore at the discretion of the developer. In addition to the actual source code, however, the following components should always be present:

- A README file in the top directory level, which contains the most important information and metadata about the software project. Beside a general description of the project, this may be instructions for installing and using the software. This includes, for example, information about the operating system used, the version of the compiler and software dependencies.
- A license file in the top directory level, which determines the rights of use of the software project.
- A detailed and easy to find documentation of all components of the software project. Every user should in principle be able to fully understand and use the research software

Documentation of the software project

The documentation should prioritize the following aspects:

- The foundation of the software project lies in its source code. In addition to providing explanatory comments within the code for individual components, it is often valuable to include a concise documentation that offers descriptive references to the functionality scope and instructions for utilizing the software. Moreover, it is important to reference the methods employed and the sources utilized in the software development process.
- When utilizing research software, it is common to require input data in a specific format. The documentation should encompass the detailed structure of this input data, offering guidance on its creation, and provide a list of any necessary software for this task. Additionally, it can be advantageous to include sample blank input data and/or test input data for reference.
- It is equally important to thoroughly describe and document the output format of the software. This entails providing a suggested procedure for utilizing or further manipulating the output data, along with specifying any necessary software for these tasks.

Backup of generated data

- In any case, at least the input data with reference to the version number used should be backed up. In this context, further information on the procedure for creating the specific input data should also be provided, if necessary.
- If the output data are easily reproducible by repeated use of the software, they should be saved only in some cases. In this case, the calculation effort and the required storage space should be evaluated.

§ 7 Experimental data

These regulations apply to all research data resulting from experimental work. The guidelines are applicable to data that researchers consider valuable for publication or long-term archiving.

To ensure the reproducibility of experiments, the documentation of experimental data should encompass all essential information required to replicate the experiment. This documentation is applicable to all types of experimentally generated data and should be recorded in a README file located at the top directory level of the respective data structure. At the very least, this README file should contain the following information:

The documentation of experimental data should include the following basic information: the name of the person processing the data, the date of processing, and any relevant details. It should also provide a comprehensive description of the experiment, including its objectives. The documentation should specify the raw data collected during the experiment, along with the method employed to gather this raw data. Additionally, a summary of the results should be included to provide an overview of the outcomes obtained from the experiment.

Types of experimental data

When considering publication or archiving, it is crucial to differentiate between various types of experimental data. Specifically, a distinction should be made between the following categories:

- **Preliminary Experiments:** These experiments are conducted solely to determine parameters or establish initial insights for subsequent test series. The data obtained from preliminary experiments may not be suitable for direct publication or archiving but can provide valuable context for understanding the subsequent experiments.
- **Raw Data from a Series of Experiments:** This refers to the unprocessed, original data collected throughout a series of experiments. It captures the primary measurements, observations, or recordings made during the experimental process.
- **Processed Data from a Series of Experiments:** This category encompasses data that has undergone analysis, transformation, or refinement. It includes data derived from the raw data through calculations, statistical methods, or other processing techniques.

By distinguishing between these types of experimental data, researchers can ensure appropriate handling, documentation, and decision-making regarding their publication or archiving.

The following applies to data from preliminary experiments:

- Only the metadata for reproduction are saved, the actual measurement data do not have to be saved.
- Description of the recorded parameters and indication of the parameters that are determined as decisive for later test series.

For raw data:

- The raw data should be backed up as raw as possible. Here, only the file format should be converted to a file format suitable for archiving and publication, if required.

For processed data, the following applies:

- Researchers weigh how costly it is to reproduce the evaluation.
 - If the reproduction of the evaluation is too time-consuming, the processed data should also be saved in its entirety.
 - If the reproduction of the evaluation is not time-consuming, the documentation of the evaluation process is sufficient.
- The complete process of obtaining and processing raw data up to the final evaluation must be recorded.
- The license for further handling of the data should be kept as open as possible, but the processing of the data involves interpretation by the researchers, so that modification by third parties is not advisable.

§ 8 Responsibilities

During the interim period, the SPP100+ Coordination project team will serve as the primary point of contact for any inquiries related to Research Data Management (RDM) within SPP100+. They will also provide guidance and orientation to new members on the subject. In case the management positions are vacant, the board will be responsible for appointing an alternative contact person to assume the corresponding responsibilities. This ensures that there is always a designated individual available to address RDM-related matters within the SPP100+.