



Guidelines and recommendations on research data management at the Friedrich Schiller University Jena

20 december 2016



Preface

At the Friedrich Schiller University Jena (FSU) research data are regarded an essential resource for gaining scientific insights. The FSU strives to support its members with suitable research data management services. These services are provided by the central service facilities and will be continuously evaluated and enhanced.

The present guidelines and recommendations supplement and specify the principles formulated in the "Policy on the handling of research data at the Friedrich Schiller University Jena". The Research Data Management Helpdesk serves as a central point of contact for all matters relating to research data. The [website of the helpdesk](#)¹ provides the most recent information.

I. Planning phase: before the research project

- While planning a research activity in which data are collected or where data form the basis of your research, it is highly recommended to examine the requirements and possibilities for efficient and sustainable research data management from early on. The Research Data Management Helpdesk is there to assist you in determining your individual requirements, find out about services and funding opportunities, and to develop a suitable strategy for your project.
- The [Service center for research and transfer](#)² provides general advice on proposals for third-party funding. Coordinated programs of the DFG (e.g. SFBs, research groups) are supported by the office of the [Vice-President for Research](#)³.

¹ <https://www.uni-jena.de/Forschungsdatenmanagement>

² <http://www.sft.uni-jena.de/>

³ https://www.uni-jena.de/Vizepraesident_Forschung.html



- It is recommended, and also required by some funding organisations, to create and submit a **data management plan** (DMP). A DMP contains the requirements as well as the technical and organisational measures ensuring a sustainable handling of research data during, and after, the project life-time. Throughout a project the DMP serves as a reference and guideline. It may evolve over time if conditions or requirements change (keyword: living document).
- Most funding organisations expect not only the **long-term safekeeping** of research data but also its **publication**. A suitable strategy should be determined in the proposal phase. Costs for research data management (e.g. labour, storage, publication), which cannot be covered by the university's basic facilities, should be requested in the proposal. The university computer centre assists you in determining the necessary IT infrastructure and services.
- While planning a research project, in particular, when publishing results, legal constraints apply and must be taken into account. Certain research data in the social sciences, life sciences and medicine, for example, are subject to strict conditions, such as privacy protection laws or require prior assessment by an ethics committee. Additionally, the protection of copyrights and legitimate interests of third parties must be ensured. It is recommended to clarify legal questions in advance while planning a research project. The [legal department](#)⁴ is there to advise you on these matters.

II. Implementation phase: during the research project

- Storing and processing, as well as collaborating on research data, should adhere to state-of-the-art IT principles. This includes, first and foremost, securing data in terms of accessibility, integrity (unaltered) and authenticity. For example, procedures for data backups and archiving need to be in place and secure data exchange platforms and versioning tools should be used.
- In terms of research data the [university computer centre \(URZ\)](#)⁵ provides (among other things) the following services. The services are offered partly free of charge as basic infrastructure or can be provided at net cost.

⁴ <https://www.uni-jena.de/Rechtsamt.html>

⁵ <https://www.uni-jena.de/URZ>



- use of network file systems (including data backup)
- mass storage (archive storage) including data mirroring
- renting virtual servers (server hosting)
- accommodation of real servers (server housing)
- block storage services for servers (virtual disks via a dedicated storage network)
- data exchange services
- versioning services
- long-term archiving services (in cooperation with ThULB)

Detailed service descriptions are available on the [URZ website](#).

- The reusability of research data is largely dependent on the availability and quality of accompanying **metadata**. Metadata is data about data and provides the context in which the data were created. As a rule of thumb, metadata should answer the classic six questions: Who? What? Why? How? When? Where? Metadata is a prerequisite enabling potential 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, MARC, or MODS, should be used.
- In the implementation phase of a project, datasets may evolve over several stages (e.g. by selection, aggregation, integration). It is good practice to label, document and keep the different **versions** at least for the duration of the project. Especially in the case of text-based data, the use of versioning tools (e.g. git, SVN) helps with the management of different versions.
- In collaborative projects or projects with large amounts of data, the use of **dedicated software** and **infrastructure** for data management is advisable. Operating these infrastructures usually requires additional resources, but they provide the advantage of a streamlined central management for research data. Finding and sharing research data is thereby facilitated, but should be governed by a **data policy** the project consortium agreed upon.



III. Final phase: after completion of the research project

- According to good scientific practice, research data must be archived and made available by the end of a project. In particular, public access has become a necessity in recent years in order to facilitate the traceability of research results as well as subsequent reuse of the data. It is recommended to publish research data following the principle "as open as possible, as restricted as necessary".
- There are various ways to publish research data. In addition to domain-agnostic repositories, a large number of discipline-specific repositories exist. To find a suitable repository for your research data, the portal re3data.org⁶ provides extensive search and filter capabilities. In addition, there are special data journals for publishing research data. In some disciplines it is common practice 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 selecting a publication method, the following criteria should be taken into account: long-term availability (at least 10 years), allocation of persistent identifiers (e.g. DOI, URN), licensing and usage terms of the data, visibility and cost.
- As a member of the FSU, you may publish and archive research data with the [Digital Library of Thuringia \(DBT\)](http://www.db-thueringen.de)⁷.
- If you published research data with an external repository, you are requested to register this publication with the Digital Library Thuringia, too.
- Individual websites (e.g. of projects, departments, employees) are generally not suitable for the publication of research data. The long-term availability of such websites is often not guaranteed and non-ambiguous identification (keyword: PID) and cross-referencing is limited.
- When selecting the data to be published, the DFG (2015) recommends, "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 that served as the foundation of a scientific article should be made public.

⁶ <http://www.re3data.org>

⁷ <https://www.db-thueringen.de>



- As with scientific articles, research data should also receive a unique, **persistent identifier** (PID). A PID (e.g. DOI, URN) is typically provided by the repository or data journal at the time of publication.
- In order to mandate usage and exploitation rights of your research data, data should always be published with a corresponding **license**. The selected license should allow at least open access for scientific purposes. Specific requirements of the funding organisation or repositories may apply and need to be considered. Suitable data licenses are available from [Open Data Commons](https://opendatacommons.org/licenses/)⁸ or the [Creative Commons](https://creativecommons.org/licenses/)⁹.
- In general, research data should be published as soon as possible. However, there may be reasons to publish research data with an **embargo period**. In this case, the data is submitted to the repository for archiving, but it is only accessible after the end of the embargo period. The researchers who are responsible for the data determine the length of the embargo period, unless otherwise requested by the policies and guidelines of funding organisations or the repositories. Embargo periods should not exceed 1-2 years after the end of the project.

Vorstehende Handlungsempfehlung wurde am 20. Dezember 2016 durch den Senat der Friedrich-Schiller-Universität Jena beschlossen.

Jena, den 20. März 2017

A handwritten signature in blue ink, appearing to read 'W. Rosenthal'.

Prof. Dr. Walter Rosenthal

Präsident

⁸ <http://opendatacommons.org/licenses/>

⁹ <https://creativecommons.org/licenses/>