

Recommendations for Handling Research Data at the Bauhaus-Universität Weimar

University Directorate ruling of 13 October 2021

Recommendations drafted by Kevin Lang of the Research Data Management Office, in conjunction with the Research Data Management work group

Contents

1	Preamble	3
2	Project planning	4
3	Project implementation	5
3.1	File organisation	5
3.2	Rights and obligations	5
3.3	Documentation.....	5
3.4	Versioning and backups.....	6
3.5	Virtual research environments.....	6
3.6	Using external IT services	7
4	Project completion	8
4.1	Publication in accordance with FAIR data principles	8
4.2	Research data repositories.....	8
4.3	Selection and deletion.....	9
4.4	Licensing	9
4.5	Long-term availability.....	9
5	Central service facilities.....	10
5.1	University Library	10
5.2	Research Operations Office.....	10
5.3	Computer centre (SCC).....	10
5.4	Legal and other advice	10
5.5	Bauhaus Research School.....	10

1 Preamble

The Bauhaus-Universität Weimar is committed to ensuring that research data is handled responsibly. In doing so, they strive for fundamentally free access to and the long-term protection of research data. The data collection process should be comprehensible and reproducible. This requires the Bauhaus-Universität Weimar's research activities to be transparent.

These recommendations for action supplement and flesh out the principles laid out in the »[Guidelines for handling research data](#)« that were concluded by the University Directorate on 12.02.2020, and indicate the various services provided and being continually further developed by the Bauhaus-Universität Weimar's central service facilities. This document is designed to serve as an aid for university members and indicates the various compulsory requirements of third-party funding providers. For general questions and further information, please note the following support services and structures:

[Research Data Management Office](#)

The Research Data Management Office is based in the Bauhaus-Universität Weimar University Library. All employees and members of the Bauhaus-Universität Weimar can contact this office should any questions about research data arise.

[Thuringian Competence Network for Research Data Management \(TKFDM\)](#)

The TKFDM was born of the »[Thuringian Strategy for the Digitalisation of Colleges](#)« and offers its services to all university research institutions in Thuringia. The network offers advice on various topics relating to research data management, organises regular events and networking sessions, provides information materials and handouts, and offers training on general and specific topics upon request.

[Forschungsdaten.info information platform](#)

The Germany-wide platform [forschungsdaten.info](#) is an information portal maintained by a nationwide RDM team and supported by state initiatives such as the TKFDM. It covers various topics relating to research data, tackles various scientific disciplines, and offers its own services, such as an instance of the Research Data Management Organiser (RDMO) that can be used to create data management plans.

Other national organisations that deal with the handling of research data, provide associated materials, and organise events include the [German Research Foundation](#) (DFG), the [Research Data Alliance](#) (RDA) and [RDA Germany](#), the [German Council for Scientific Information Infrastructures](#) (RfII), [Bausteine Forschungsdatenmanagement](#), the [German National Research and Education Network](#) (DFN) and the [Digital Curation Centre](#) (DCC).

2 Project planning

If data is collected during or forms the basis of research, early planning is recommended. Many funding organisations therefore already require submitting a data management plan (DMP). This sets out the requirements and the technical and organisational measures to ensure sustainable handling of research data during the project period and also after the project is complete.

The DMP should be viewed as a »living document« that can be updated both during the project phase and after completion of the project. It serves as a reference point and can be used to enable those involved to understand the project. The possible components of a DMP are listed below.

Overview: Includes metadata about the project, such as the title, objectives, funding providers, partners, project managers and/or run time.

Inventory: Description of the origin and quality of existing data and how it is currently or is planned to be incorporated into the project.

Work sequence: Description or estimate of what data types and volumes will result from the project, the formats in which they will be stored, and what quality assurance procedures will be followed.

Transfer: Description of all criteria, time points and methods for data transfer. Information is provided regarding data selection, the metadata used, and the validation and archiving of research data.

Distribution: Description of the selection and implementation of data exchange and publication. This touches upon interoperability with external data services and explains which publishing platforms have been selected and why.

Obligations: Description of the requirements of funding and archiving institutions, property rights and copyright, intended conditions of access and use, ensuring data protection and backup, and responsible individuals.

Resources: Description of the anticipated costs of compliance with the DMP, including staff, preparing metadata, and curating and archiving research data.

The Research Data Management Office is available to advise on preparing the DMP. There are also various documents and software solutions that could help you prepare this plan, including:

- [Templates](#) (from the Humboldt-Universität zu Berlin)
- [Research Data Management Organiser](#) (RDMO, funded by DFG, open-source, instance on forschungsdaten.info)
- [DMPonline](#) (developed by the Digital Curation Centre (DCC), open-source)
- [ARGOS](#) (developed by OpenAIRE-EUDAT, open-source)
- [Data Stewardship Wizard](#) (funded by various institutions)

3 Project implementation

During implementation of the project, various elements need to be taken into account when handling research data. This section highlights various focal points and support services that enable organised research data management during the project implementation phase.

3.1 File organisation

The project data should be organised in such a way that the individual concerned and any new employees can quickly navigate the structure. The data structure should correspond with the working area and file formats. It is helpful to establish guidelines within the work group for maintaining these structures.

The 5S model can be used for implementing a suitable file organisation structure, and consists of the following steps:

- 1) **Sort:** Find and delete unnecessary files or mark them for deletion
- 2) **Set in order:** Create useful file structures and develop naming conventions
- 3) **Shine:** Regularly review and adjust structures
- 4) **Standardise:** Identify standards and best practices and discuss them with the work group
- 5) **Sustain:** Make organisation a habit, update and pass on to new employees as required

So that both employees and computers are able to quickly identify the content of files and folders, these should have metadata (such as date, person or title) and meet technical requirements (such as no spaces or special characters). An example of a document filename that follows these guidelines might be:

20200501_presidium_switch_to_digital_teaching.pdf

3.2 Rights and obligations

Researchers are subject to various rights and obligations, resulting notably from legal requirements but also from contracts with third parties. These many include data protection or prior review by an ethics committee.

The European General Data Protection Regulation (GDPR), the German Federal Data Protection Act (BDSG) and the Thuringian Data Protection Act (ThürDSG) form the legal basis for the handling of personal data. The legal framework conditions for a research project also include documented agreements regarding usage rights for the research data and findings resulting from the project. The general objective is to minimise personal data within the project, so that data storage and publication are subject to fewer protective measures. It may be unfeasible for researchers to obtain declarations of consent in either collecting or using personal data, if the associated workload proves impossible or disproportionate.

3.3 Documentation

In order to adhere to the FAIR data principles (Findable, Accessible, Interoperable, Reusable), documentation must ensure that research data is comprehensible and can be reused. This also includes methodological, evaluation and analysis steps, the resulting software, and the formulation of hypotheses.

A project's research data is generally documented via metadata. This is data about the research data to note the context in which it was created. It may answer the classic 6 »W questions«: Who? What? Where? When? Why? How? This ensures that the metadata is machine-readable, subject-specific standards and standardised terminology may be used. However, universal metadata standards such [Schema.org](#), [Dublin Core](#), [MARC](#) or [MODS](#) also in particular apply. In addition, a collaborative wiki, readme files, or simple documents (e.g. in PDF format) may be used to describe project processes.

3.4 Versioning and backups

Research data must be protected against manipulation as effectively as possible. To do so, it may be helpful to implement both data backup and data versioning measures.

Research data should be stored and processed in line with the latest technological standards and should comply with the information security protection goals of confidentiality, integrity and availability. To protect data from unintended changes or accidents, data should be protected using a versioning system and a good backup strategy whenever possible. In addition to helping restore data in an emergency, this also provides a level of control over the progress of process steps. File versioning can be performed manually, or automatically by a system. If file versioning is performed by hand, it should be determined what factors are used to determine versions and how these are identified. For a backup strategy, the 3-2-1-0 backup rule can be used as a rule of thumb: 3 copies of files, 2 different storage media, 1 copy in a different location and 0 errors in file recovery. A cloud solution can be very helpful in implementing this strategy (→ 3.5. »Virtual research environments« and → 5.3 »Computer centre«).

3.5 Virtual research environments

Virtual research environments (VREs) are dedicated working environments or portals that can be used to establish a consistent workflow, comply with standards and, if relevant exchange, files with colleagues outside of the university.

Creating a VRE may be a sensible solution for working with large volumes of data, different data types, or large numbers of colleagues, or for joint projects. Various criteria must be taken into account when selecting a VRE, including: local or external hosting, potential organisational structures, storage capacity, the ability to define user settings and access options, the ability to award licences, maximum duration, subject-specific features, data protection, administration, maintenance, and the costs incurred by the system. As the environment rarely specifies all of the work steps organised, these should be set by work group and project (see section 3.1 »File organisation«). Open-source research environments are also preferred to avoid being dependent on third-party providers, and solutions already offered by the Bauhaus-Universität Weimar should be considered. For example, collaborative work on office documents (such as texts, tables or presentations) may use the university's cloud app »Nextcloud«, or an instance of »GitLab« could be used to organise and process source code projects. In addition to electronic laboratory journals, there are various other software solutions and established portals that can help researchers with specific work processes in their subject area. The TKFDM has created an [overview](#) of the publicly funded solutions available in Germany.

3.6 Using external IT services

Under the joint IT usage regulations for the Bauhaus-Universität Weimar and the University of Music FRANZ LISZT Weimar (section 11 »Using external IT services«), external IT services should only be used in exceptional circumstances.

External IT services should only be considered if the relevant facility offers value-added content or another direct benefit (e.g. improved quality or simplification) as the Bauhaus-Universität Weimar cannot assume any responsibility for the integrity, availability, or confidential and legally compliant handling of data. It is therefore recommended that you only use the university's IT services or install applications on the facility's own servers where the data owners have full control over their data and how it is processed. This is generally the case for open-source solutions. However, if an external IT service that stores research data is being used, a usage contract (generally a service agreement) must be established with the service provider and, if relevant, a contract covering contract data processing and a list of the responsible party's processing activities should be prepared, if it relates to data that requires protection (→ 3.2 »Rights and obligations«). Students and other users must also be informed if external IT services are being used; this use must be voluntary and revocable.

4 Project completion

In the spirit of good scientific practice and open science (particularly open data, open source and open access), all research findings and their underlying research data must be made available to the public as soon as possible after the end of the project. The following principle applies: »As accessible as possible. As restricted as necessary«. This should help ensure that knowledge is disseminated as freely as possible and made both comprehensible and reproducible.

4.1 Publication in accordance with FAIR data principles

The FAIR data principles are a collection of guidelines to ensure that research data is Findable, Accessible, Interoperable and Reusable when published. They are already supported by many organisations and institutions, and set as required guidelines by numerous funding providers.

Various measures should be put in place to comply with the FAIR principles. For **findable**, the data must be given a clear identifier that refers to various pieces of metadata and allows data to be found using search engines. For **accessible**, access to the data set must be clearly defined via conditions, authentication steps and authorisation steps, and must be downloadable in various forms. A data set is considered **interoperable** if the file formats and metadata are in line with open, subject-specific standards. Finally, a data set is considered **reusable** if it is documented in a comprehensible way (→ 3.3 »Documentation«) and is licensed for reuse (→ 4.4 »Licensing«). It should be noted that not all measures can be fully implemented depending on the subject area, as subject-specific repositories and platforms (→ 4.2 »Research data repositories«) do not yet always have the necessary features in their portfolios. The FAIR principles should therefore also be viewed as the gold standard to be striven for. The Research Data Management Office is available to help with reviewing FAIR principles for your own data set.

4.2 Research data repositories

The research data and central materials on which the publication is based should be made accessible in recognised archives and repositories.

The first stage in publishing research data should be finding a suitable subject-specific repository that meets the specialist requirements (e.g. a map view for geological surveys). This can be found using the DFG-funded registry re3data.org. There are also interdisciplinary repositories available, such as [Zenodo](https://zenodo.org) or [Eudat B2Share](https://eudat.eu), which are funded by the EU and therefore also provide suitable platforms, particularly if the data sets in question are mixed or cannot be assigned to a specific subject. In general, repositories must meet the following quality requirements: a constant identifier (e.g. DOI) for the data set, constant identifiers (e.g. ORCID) for authors, as many different pieces of descriptive metadata as possible with links, download and export options for the data set, description and/or documentation options, access adjustments (e.g. upon request or with an embargo), licensing with standard licences, overview of the data set and/or thumbnail of individual files, data set versioning, registration and/or processing options for authors, and the ability to find data sets using common search engines.

4.3 Selection and deletion

At the end of the project at the latest, any files that are redundant, not required to prove research findings, or do not have any added value for further research must be deleted.

The data to be deleted should be selected by the project's research director and/or the responsible members of staff who collected and/or processed the data. Deleting data fields, data sets or defined groups of data must not compromise the integrity of the remaining data pool. Generally speaking, the granularity with which data is to be stored or deleted is primarily defined by protection requirements, workload, and further use of the data. There may be research data that must be deleted once a specific period has expired in order to comply with statutory regulations.

4.4 Licensing

The published research data must bear a licence so that its usage conditions can be clearly set. The licences selected should ideally be free and suited to the data type.

The most frequently used licence contracts are offered by Creative Commons (CC). Use of the data set can be restricted via various attributes: »BY« for author attribution, »NC« for no commercial use, »ND« for no publishing of derivative versions, and »SA« for distribution only under the same conditions. The CC0 (zero) licence as a special version with no attributes offers total freedom in the further use of data, but is difficult to apply under German copyright law and is therefore not recommended. Any CC licences used should be version 4 or later (e.g. CC BY 4.0), as these are the first to include research data protection. In addition to Creative Commons licence contracts, other licence templates are also available, such as Open Data Commons (ODC) for data collection, GNU General Public Licence (GPL) or Apache for software projects.

4.5 Long-term availability

The research data underlying research findings (generally raw data) should generally – depending on the subject area – be accessible for a period of ten years and traceable to the institution where it originated, or stored in site-wide repositories.

Research data can be considered for long-term storage or archiving if, as well as serving as evidence of research findings or activities, it is also expensive to collect or compile, is not reproducible, or presents critical points of the research. The data itself must use open, standardised, unencrypted and non-patented/non-proprietary file formats as far as possible. It should also be sensibly structured and documented, and any applications, scripts or methods that are essential for using the files should be made available. Subject-specific repositories (see 4.2 »Research data repositories«) very rarely offer long-term storage of research data alongside its actual publication. The SCC offers a long-term storage option for storing research data as an offline version on magnetic tape, which can also be used as WORM (Write Once Read Many) tape. Local projects and nationwide developments are seeking new solutions for the long-term archiving of research data. These are being handled by institutions such as the [Rfil](#) and by [NFDI initiatives](#).

5 Central service facilities

The Bauhaus-Universität Weimar's central service facilities offer various services to assist university employees and members with handling research data. These services are described below.

5.1 University Library

In addition to the general Research Data Management Office, the University Library also offers subject specialists from the faculties' relevant fields of study to provide support for students, employees and members. These contacts are knowledgeable about open access and research data repositories.

5.2 Research Operations Office

The Research Operations Office advises on suitable funding programmes and helps researchers apply for third-party research funding. This also includes supporting applications for large equipment. Information regarding current funding programmes is made available via a regular newsletter and events. The Research Operations Office also offers assistance with research technology transfer and with participation in trade fairs.

5.3 Computer centre (SCC)

The Service Centre for Computer Systems and Communication (SCC) offers the following services for free or at additional cost:

- Network infrastructure
- Identity management
- Data storage and backup
- Data exchange services
- Long-term storage
- Virtual servers
- Web service

It also provides advice and support for implementing the necessary level of security in the IT systems being used. The [data backup concept](#) can be viewed online.

5.4 Legal and other advice

The Legal Affairs office, the Data Protection Officer and the Ombudspersons for ensuring good scientific and artistic practice are on hand to help with legal and other questions relating to projects, research data and software solutions.

5.5 Bauhaus Research School

The Bauhaus Research School is a central research institution at the Bauhaus-Universität Weimar that promotes early stage scientific, artistic and design talent. It offers advice and support with doctoral degrees in Weimar. This also includes an annual curriculum with workshops, including on research work (e.g. good scientific practice).