UFZ-Directive | IR-5/18 | 17.06.2024

Principles for the responsible handling of research data at the Helmholtz Centre for Environmental Research – UFZ

Please note: *This is a translation of the German version of the above-mentioned directive and is intended solely as a convenience to the non-German-speaking employees. Any deviations from the original German directive are not binding and have no legal effect for compliance or enforcement purposes.*

Enforced by	
Executive Management	
Prof. Dr. Rolf Altenburger	Dr. Sabine König
in Executive Management meeting no. 380 on 17 June 2024	

Document history

Date	Version	
13.09.2018	1	Adoption of "Principles for the responsible handling of research data at the Helmholtz Centre for Environmental Research GmbH - UFZ"
17.06.2024	2	Supplementing the principles with the rules for the responsible handling of research data at UFZ

Content:

1   Principles for the responsible handling of research data at the UFZ

The UFZ is one of the world's leading research centers in the field of integrated environmental research. It supports politics, business and civil society in better understanding the consequences of human activities for the environment and in developing options for social decision-making processes. The UFZ uses its extensive experience in integrated environmental research and its innovative scientific infrastructures to address complex environmental problems. It supports various scientific communities on a national and international level and is committed to the social goal of reliably collecting and processing quality-assured research data on the state of and changes to our environment in accordance with recognized standards, and to preserving and making them available to the national and international community in the long term.

Research data are collected in all scientific disciplines and therefore in all departments of the UFZ. They range from classical laboratory, field and remote sensing data to research data from cohort studies or qualitative and quantitative social research.

The present principles are based on the recommendations of the Helmholtz Association for guidelines on the management of research data[1], on the Guidelines of the European Commission on Data Management according to the FAIR Principles[2] and the guidelines of the DFG (German Research Foundation/Deutsche Forschungsgemeinschaft) for the handling of research data[3].

1. The UFZ is committed to the principles of open access to knowledge, research results and technologies. Legal regulations, in particular concerning the protection of personal data, as well as scientific interests and contractual agreements with cooperation partners and, where applicable, exploitation interests combined with at least temporary confidentiality of research data must be taken into account. In individual cases, the UFZ may specify appropriate embargo periods for exclusive first use. The researchers of the UFZ observe the principles and rules of good scientific practice.[4]

2. The UFZ recognizes the need to provide the necessary infrastructural and organizational framework conditions for research data management (specialist personnel, qualification and infrastructures) on a permanent basis. The more detailed design requires further conception.

3. The UFZ recognizes the FAIR principles[5] as the superordinate principles for the sustainable provision of research data collected at the UFZ (see **Appendix 1**).

4. The UFZ supports national and international declarations on the handling of research data.

---

[1] https://www.helmholtz.de/fileadmin/user_upload/01_forschung/Open_Access/DE_AKOS_TG-Forschungsdatenleitlinie_Positionspapier.pdf [11.10.2017]
[2] EC (2016) Guidelines on FAIR Data Management in Horizon 2020:
http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf [11.10.2017]
[3] Deutsche Forschungsgemeinschaft - DFG (2019) Leitlinien zur Sicherung guter wissenschaftlicher Praxis. Kodex. https://doi.org/10.5281/zenodo.6472827
[4] Rules for Safeguarding Good Scientific Practice at the UFZ. UFZ-Regulation IR-17/12, https://www.ufz.de/index.php?en=46115 [2023-09-30]
[5] Wilkinson et al. (2016): *The FAIR Guiding Principles for scientific data management and stewardship*. Scientific Data, 3, 160018, https://doi.org/10.1038/sdata.2016.18

2   Scope of the Rules for the Responsible Handling of Research Data at the UFZ

(1)   The Rules for the Responsible Handling of Research Data at the UFZ (hereafter referred to as the "RDM Rules") are aimed at all people working at the UFZ (including scientific guests) who collect or have collected data in the context of research activities and/or work with research data. They form a binding framework for research data management (RDM) at the UFZ.

(2)   In the context of research collaborations, the RDM Rules shall apply unless the other parties involved stipulate other equivalent or higher standards. The UFZ researchers involved are responsible for compliance with these principles. If research data guidelines are developed within the framework of research projects or are prescribed by third-party funding bodies and they have equal or higher standards, they take precedence over these principles.

(3)   The RDM Rules apply to digital research data and digital representations of analogue data.

(4)   The rules of the RDM were drawn up taking into account the Rules for Ensuring Good Scientific Practice at the UFZ (IR-17/12) and replace the Principles for the Responsible Handling of Research Data at the Helmholtz Centre for Environmental Research GmbH - UFZ (IR-5/18) (hereinafter referred to as "RDM principles").

3   Governance of research data management at the UFZ

(1)   The Research Data Management (RDM) unit was established in 2019 with the aim of facilitating the responsible handling of research data at the UFZ through professional advisory and support services and qualifications, as well as by further developing existing data infrastructures across the entire data lifecycle at the UFZ. To this end, the executive management and the research units gave the RDM unit strategic and personnel responsibility for the design of a professional data management system based on existing and future requirements. The unit was given responsibility for implementing the RDM Principles (IR-5/18) and the RDM Rules, a duty which includes specifying the responsibilities and fields of action as well as the necessary steps for implementation. The RDM unit represents the UFZ in relation to research data management questions in the DataHub of the Helmholtz Association's Earth and Environment research field as well as in other national and international initiatives and consortiums.

(2)   The RDM board was set up to ensure the targeted development of a modern and sustainable infrastructure. It serves as an interface for strategic developments in the research units and their specialist scientific communities, prioritizes developments and advises the executive management and the research units on data-related strategic issues. Each research unit has two representatives on the RDM board who are familiar with their research unit's strategic goals, program-oriented research (PoF) and the associated data management aspects.

(3)   Each department in the UFZ appoints a data representative, who receives ongoing training through suitable RDM courses. This person supports the respective department with respect

to data management matters, serves as a contact person for the RDM board representatives of the respective research unit, and records and communicates needs related to data management to the RDM unit.

4 Responsibility for the responsible handling of research data

(1) All UFZ employees who collect and/or process research data must handle them responsibly.

(2) The executive management, the research unit leaders and the department heads create the conditions that are necessary for the sustainable handling of research data throughout the entire data lifecycle.

(3) Department heads are responsible for informing staff (including students, scholarship holders) about the RDM Principles and the RDM Rules and for checking that staff comply with the requirements.

(4) When their staff leave, project managers and/or department managers are responsible for ensuring that the original data remains at the place of origin, for taking precautions regarding the appropriate transfer of primary data, and for clarifying access rights.

5 Qualification and development of personnel

(1) The UFZ provides its employees with professional advisory and support services and qualifications on RDM[6]. The services are tailored to the specific requirements of the specialist disciplines and infrastructures. This is done in cooperation and coordination with the educational opportunities offered by the Department of Human Resources and Corporate Culture (PACE) and with the offers available across the entire Helmholtz Association (e.g. those provided by HIDA[7] and HIFIS[8]). Training is available in areas such as:

- Fundamental aspects and practices of RDM
- Digitalization in research and handling research data infrastructures

(2) Employees should be able to take advantage of RDM training, qualifications and self-learning opportunities in order to adapt their working methods, workflows and technologies to the processes of digitalization.

(3) Students, doctoral candidates and interns must be informed by their supervisors at the UFZ about the appropriate handling of research data. They must be taught technical competencies and standards. They are encouraged to take advantage of relevant educational offers about RDM at the beginning of their time at the UFZ.

---

[6] RDM Guidelines UFZ, https://rdm.pages.ufz.de/guidelines/
[7] Helmholtz Information & Data Science Academy (HIDA), https://www.helmholtz-hida.de/en/
[8] Helmholtz Federated IT Services (HIFIS), https://hifis.net/

6   Scientific recognition

(1)   In accordance with the FAIR principles, the UFZ recognizes the generation, processing and provision of research data as a central and indispensable part of the research process. Quality-assured research data are part of the UFZ's scientific output and their production and publication are valued accordingly. This is done by including data publications in scientific reporting and annual discussions, among other things.

(2)   To ensure that data publications are included in scientific reporting, all researchers are obliged to report their data publications to the UFZ's publication index[9].

(3)   In addition to established metrics for text publications, the UFZ also calculates metrics for research data and research software. This is done based on the guidelines and recommendations of the Helmholtz Association.

(4)   A properly citable data publication promotes the reproducibility of the studies that are based on it. Research data publications should therefore be made citable by means of a persistent identifier (PID), e.g. DOI or handle.

(5)   The researchers shall indicate the use of data by citation and comply with the recognized Data Citation Principles[10].

7   Relevance of data

Research data may have different degrees of relevance with regard to their strategic importance for the research units and research programs at the UFZ. This has consequences for the sustainable storage of the data and for the development and operation of the necessary research data infrastructures. Strategically relevant data is prioritized in these processes.

8   Use of research data infrastructures

(1)   The term 'research data infrastructures' comprises the combination of hardware, software and personnel support that make it possible to generate or collect, analyze, simulate and finally publish and/or archive data using various methods

(2)   The UFZ implements and maintains a research data infrastructure with a basic range of functions and thus ensures adequate storage, long-term availability, data security and the technical availability of digital research data in accordance with the FAIR principles. Additional usage-specific or project-specific requirements can be implemented via the RDM board when the necessary resources are available; these requirements should already be considered during the project application process.

---

[9] Publishing at UFZ, https://www.ufz.de/index.php?de=50166 (page only available in German)
[10] Martone M. (ed.) (2014): Data Citation Synthesis Group: Joint Declaration of Data Citation Principles. San Diego CA: FORCE11. https://doi.org/10.25490/a97f-egyk

(3) The UFZ provides the corresponding high-performance infrastructures by the RDM unit and the IT department (WKDV) and further develops them with due regard for proportionality

(4) Data generating and processing research infrastructures (e. g. large scientific equipment, expansion investments) are to be connected to the data infrastructures provided by the RDM unit. The provision and operation of basic services and storage infrastructures are carried out by the WKDV. If national and international research data infrastructures recognized in the specific disciplines are available as alternatives to the basic services provided by the UFZ, their use must be examined and realized, taking proportionality into account.

(5) If national and international research data infrastructures are available for the processing, storage and archiving of data and metadata that require a connection to the infrastructures provided by the UFZ in order to be used, a connection to these infrastructures must be considered and implemented in accordance with proportionality.

9 Planning the handling of research data

(1) The researchers and the RDM unit shall work together to create a data management plan that takes into account the relevant requirements of the funding bodies and partners. All the research data management steps – from data collection to publication and making the research data accessible – shall be defined, documented and implemented according to this plan. The depth of content should be based on international recommendations[11]. Ethical, legal and utilization aspects must be taken into account and the rules of good scientific practice must be complied with. The required content can be specified at the level of departments or working groups (e.g. in relation to discipline-specific metadata standards established in the communities).

(2) The data management plan should be understood as a living document that can be adapted and added to throughout the course of a project. It should be archived in the UFZ archive or another suitable location, together with the relevant data records, upon completion of the project at the latest. During the preparation of content management plans, the RDM unit and the data officers provide advice on content-related and organizational issues when creating data management plans.

(3) The RDM unit provides advice on issues related to the consideration and acquisition of funds for data management in project proposals.

10 Data access and open data

(1) The UFZ supports and promotes the long-term secure storage of and generally open access to research data from publicly funded research. Researchers, supervisors and project managers make joint decisions as early as possible about the public provision of the research data.

---

[11] Science Europe (2021): Practical Guide to the International Alignment of Research Data Management – Extended Edition. https://doi.org/10.5281/zenodo.4915862

(2) Wherever possible, research data (including relevant metadata) should be made publicly available with the associated scientific publication as early as possible. The following must be taken into account: intellectual property rights, data protection rights, legal regulations, in particular the protection of personal data, ethical aspects, scientific interests, contractual agreements with cooperation partners and, where applicable, utilization interests.

(3) When research data is published and made available, they must be accompanied by license information. When selecting the infrastructures for the provision and publication of research data, the researchers shall review the access and licensing conditions. (→ Legal aspects)

(4) Basic metadata, which facilitate the retrieval of research data, should be available without access restrictions (in the public domain), as long as this action does not raise any confidentiality or secrecy concerns.

(5) If possible, the persons, text publications, institutions, software and related data recorded in the metadata should be referenced using suitable persistent identifiers (PID) (e.g. ORCID[12], ROR[13], DOI[14]).

11 Quality assurance

(1) Quality assurance should cover the entire life cycle of the data and thus ranges from data collection, workflows and methods of processing, data storage and backup, up to the controlled deletion of the data.

(2) A core element of quality assurance is the collection of metadata according to case-specific standards. The metadata contain information about the quality of the data as well as the quality assurance measures that have been applied to them (e.g. methods, software). It should be possible throughout the entire data lifecycle to determine who created and is responsible for the data, as well as the methods used. The methods and software used to process the data, as well as the persons who created and who are responsible for the data, should, if possible, be referenced with PIDs. Wherever possible, the metadata that is necessary for meeting the FAIR requirements should be collected.

(3) The research data must be managed correctly and completely in an unadulterated and reliable manner, and the integrity of the research data must be ensured. The UFZ provides the appropriate infrastructures (e.g. redundant data storage) (→ Use of research data infrastructures).

12 Long-term availability and reuse

(1) Research data should also be stored and secured in the long term in a suitable, reliable way for subsequent use, and published if necessary. For the provision and appropriate reuse of

---

[12] ORCID – Open Researcher and Contributor ID. https://orcid.org/
[13] ROR – Research Organisation Registry ID. https://ror.org/
[14] DOI – DataCite Digital Object Identifier. https://datacite.org/

research data, it is necessary for technical and descriptive metadata to be recorded during the data collection process in addition to a description of the method of data collection (→ Quality assurance).

(2) The RDM unit provides an archiving solution for the long-term storage of research data in order to meet the requirements specified in the *Rules for Ensuring Good Scientific Practice at the UFZ (IR-17/12, as amended from time to time)*. External specialist repositories and data centers can also be used for long-term storage, as long as the measures stipulated here are taken into account.

(3) Researchers must also take into account the information regarding the backup, storage and use of primary data and protocols contained in the *Rules for Ensuring Good Scientific Practice at the UFZ (IR-17/12, as amended from time to time)*.

(4) Regarding the format of the data, preference should be given to free and **standard formats that are established in the respective discipline**, in order to facilitate interoperability and ensure long-term readability. It is also necessary to ensure long-term access to software for reading and processing the data - especially in the case of data stored in copyright-protected (proprietary) data formats - to guarantee reusability.

13 Legal aspects

(1) The ownership of data is regulated in the employment contracts or guest contracts between the researchers and the UFZ.

(2) Each data record must be provided with licence information (terms of use). The use of suitable open, standardized licenses should be prioritized. It is recommended to use the Creative Commons licences[15], and specifically the CC0 licence for metadata and the CC-BY licence for data.

(3) Instructions on the processing of personal data are dealt with in the *Rules for Ensuring Good Scientific Practice at the UFZ, Part A, Section 6 "Legal Framework, Research Ethics and Rights of Use" (IR-17/12, as amended)* and must be taken into account by the researchers.

14 Embedding in national and international initiatives and consortiums

(1) The UFZ cooperates with other scientific institutions and participates in national and international initiatives and consortia to develop common standards and structures in the field of RDM and Open Science. To ensure international connectivity and help shape the digital data ecosystem, the UFZ participates in the program-oriented research of the Helmholtz Association (PoF), the Helmholtz Incubator Platforms[16], the DataHub[17] of the Earth and

---

[15] Creative Commons licences. https://creativecommons.org/licenses/?lang=en
[16] Helmholtz Incubator. https://www.helmholtz.de/en/research/challenges/information-data-science/helmholtz-incubator/
[17] Helmholtz DataHub. https://datahub.erde-und-umwelt.de/en/

Environment research field, several domain-specific consortia of the German National Research Data Infrastructure (NFDI)[18] and the European Open Science Cloud (EOSC)[19].

(2) The RDM unit helps UFZ researchers transfer their data products to the infrastructures of national and international initiatives, as well as to use these infrastructures and integrate them into their own work.

15 Definitions

a) Data lifecycle

The data lifecycle is a model that describes the cyclical nature of working with data of all kinds, including information, in their various processing and use stages within the process of scientific value creation. The main stages of this lifecycle are data generation (e.g. measurements), data preparation, data evaluation/analysis, storage and long-term archiving, as well as the provision of data through publication (e.g. in databases and repositories, as journal publications, on online platforms) through to the reuse of data in other or new research contexts that may also arise from teaching.[20]

b) Data management plan (DMP)

A data management plan (DMP) is an important tool for structuring the handling of your own research data. DMPs can be used as checklists as well as for the ongoing documentation of the data collection and the long-term storage or publication of the data.[21]

c) Data quality

The term data quality refers to the quality and reliability of data objects themselves. For data objects to be scientifically usable, the data they contain must be obtained in accordance with documented standards (e.g. measuring instruments, calculation algorithms) and recognized methods, and these standards must be transparent and ensure sustainability. Whenever possible, any uncertainty in the data should be appropriately quantified.

The evaluation of data quality depends on the requirements to be defined, such as the accuracy of the measured values, and this depends on the respective research question and thus on the intended use. In addition, data quality is determined according to whether data sets and collections contain sufficient information about how they were generated, processed and analyzed. This information should be in the form of (preferably standardized) technical and specialized metadata. This is an important prerequisite for any subsequent use of the data and the reproducibility of the research results.[22]

---

[18] NFDI – German National Research Data Infrastructure. https://www.nfdi.de/?lang=en
[19] EOSC – European Open Science Cloud. https://eosc-portal.eu/
[20] RfII – German Council for Scientific Information Infrastructures (2016): Definitions. RfII Reports No. 1. http://nbn-resolving.de/urn:nbn:de:101:1-201607146410 (document only available in German)
[21] Helmholtz Metadata Collaboration (2021): Glossary. https://helmholtz-metadaten.de/en/glossary
[22] RfII – German Council for Scientific Information Infrastructures (2016): Definitions. RfII Reports No. 1. http://nbn-resolving.de/urn:nbn:de:101:1-201607146410 (document only available in German)

d) Research data

Research data includes, but is not limited to, measurement data, laboratory data, audiovisual information, text, survey data or observational data, methodological testing procedures and questionnaires. Since research data in some areas is based on the analysis of objects (e.g. samples of tissue, material, rock, water and soil, as well as test specimens, installations, artefacts and works of art), these objects must be handled equally as carefully. It is likewise necessary to consider adequate approaches for their subsequent professional use, whenever reasonable and possible.[23]

e) Research data management (RDM)

Research data management refers to the process of transforming, selecting and storing research data with the aim of keeping it accessible, reusable and verifiable in the long term, independent of who generated the data. Structured measures can be taken at all stages of the data lifecycle to maintain the scientific validity of research data, to preserve its accessibility by third parties for the purposes of evaluation and analysis, and to safeguard the data trail.[24]

f) Research data infrastructure

The term research data infrastructure comprises the combination of hardware, software and personnel support that make it possible to generate or collect, analyze, simulate and finally publish and/or archive data using various methods.

g) Metadata

Metadata or meta information are structured data that contain information about the characteristics of other data.[25] Metadata refers to all additional information that is necessary or useful for the interpretation of the actual data, e.g. research data, and that enables (automatic) processing of the research data using technical systems.[26]

h) Metadata standards

Metadata standards have been created for specific purposes in order to ensure interoperability, i.e. the linking and joint processing of metadata. They are used to describe similar data in a uniform manner, both in terms of content and structure. A metadata standard can often be converted to another metadata standard by means of 'mapping'.[27]

i) Open Data

Open data refers to data that can be used, disseminated and reused by anyone for any purpose. Restrictions on use are only permitted in order to safeguard the origin and openness of the knowledge, for example by naming the author or using a share-alike clause[28].

---

[23] German Research Foundation (DFG) (2021): Handling research data.
https://www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/forschungsdaten/forschungsdaten_checkliste_de.pdf (document only available in German)

[24] forschungsdaten.info (2022): Research data and research data management. Glossary.
https://forschungsdaten.info/praxis-kompakt/glossar/#c269836 (page only available in German)

[25] Helmholtz Metadata Collaboration (2021): Glossary. https://helmholtz-metadaten.de/en/glossary

[26] forschungsdaten.org (2015): Metadata. https://www.forschungsdaten.org/index.php/Metadaten (page only available in German)

[27] forschungsdaten.info (2022): Research data and research data management. Glossary.
https://forschungsdaten.info/praxis-kompakt/glossar/#c269914 (page only available in German)

[28] https://de.wikipedia.org/wiki/Open_Data

j) Open Science

The term Open Science bundles strategies and procedures, all of which aim to consistently use of the opportunities offered by digitalization to make all elements of the scientific process openly accessible and reusable via the internet. This is intended to create new opportunities for science, society and industry with regard to the use of scientific findings.[29]

k) Persistent identifiers (PID)

Persistent identifiers (PID) are globally unique strings used to identify resources. They are managed in a distributed infrastructure and that is capable of resolving them.[30]

l) Repository

A repository is a database or data archive used to store and publish digital research data. Its primary purpose is to store data for a very long (ideally unlimited) period of time, as well as to ensure they remain available, citable and reusable. In addition, repositories can use rights and license management systems to regulate the conditions under which research data can be accessed and used.

m) Standards

Standards and rules are used to standardize the content-related, formal, legal, organizational and technical aspects involved in the processing of data, information and knowledge. Standards consequently provide a type of quality assurance (quality in terms of reusability). The digitalization of all areas of science and the rapid evolution of digital technologies make standardization necessary, especially in relation to data and metadata, exchange formats, interfaces, data models, markup languages and vocabularies. Standardization and compliance with standards in data management are complex, yet scientifically and economically sensible investments in quality assurance and sustainability.[31]

16 Entry into force

The regulation comes into force on June 17, 2024 and replaces the regulation "Principles for the responsible handling of research data at the Helmholtz Centre for Environmental Research GmbH - UFZ" in the version dated September 13, 2018.

---

[29] Open Science AG (2014): Mission Statement. V 1.0.
[30] Helmholtz Metadata Collaboration (2021): Glossary. https://helmholtz-metadaten.de/en/glossary
[31] RfII – German Council for Scientific Information Infrastructures (2016): Definitions. RfII Reports No. 1. http://nbn-resolving.de/urn:nbn:de:101:1-201607146410 (document only available in German)

Appendix 1 – The FAIR Principle

The FAIR principle:

- stands for "Findable, Accessible, Interoperable and Re-usable" - findable, accessible, integrable and sustainably usable;
- serves to enable not only humans, but essentially also machines, to automatically search for and use research data and to reuse it sustainably;
- facilitates exploration, access, integration, reuse, citation and referencing of research data;
- should apply not only to research data, but also to the tools and workflows used to generate that data.

The use of "(meta)data" means that the principles should be applied not only to the data, but also to the metadata.

To be **F**indable:

F1. (meta)data are assigned a globally unique and persistent identifier

F2. data are described with rich metadata (defined by R1 below)

F3. metadata clearly and explicitly include the identifier of the data it describes

F4. (meta)data are registered or indexed in a searchable resource

To be **A**ccessible:

A1. (meta)data are retrievable by their identifier using a standardized communications protocol

A1.1 the protocol is open, free, and universally implementable

A1.2 the protocol allows for an authentication and authorization procedure, where necessary

A2. metadata are accessible, even when the data are no longer available

To be **I**nteroperable:

I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

I2. (meta)data use vocabularies that follow FAIR principles

I3. (meta)data include qualified references to other (meta)data

To be **R**e-usable:

R1. meta(data) are richly described with a plurality of accurate and relevant attributes

R1.1. (meta)data are released with a clear and accessible data usage license

R1.2. (meta)data are associated with detailed provenance

R1.3. (meta)data meet domain-relevant community standards