



Table of Contents

	Seite
List of Abbreviations	2
1 Objective.....	3
2 General Principles	3
3 Definitions	3
4 Research Data Management	4
5 Raw data and associated metadata	4
6 Results.....	5
7 Legal Requirements	5
8 References	6

Appendixes

Appendix 1: Checklist for a Horizon 2020 Data Management Plan

Appendix 2: Data Cite Metadata Schema v4.0



List of Abbreviations

EMFL	European Magnetic Field Laboratory
FAIR Data	Data that is <u>F</u> indable, <u>A</u> ccessible, <u>I</u> nteroperable and <u>R</u> eusable
DMP	Data Management Plan
CC BY	Creative Commons Attributive License
CC0	Creative Commons Universal License



1 Objective

The proper management of research data is imperative to ensure that scientific findings are findable, accessible, interoperable and reusable (FAIR). The national and international research organizations support the long-term safeguarding of and the open access to research data from publicly funded research, as laid down in the "Berlin Declaration on Open Access" of 2003 [1] and the "Guidelines on Data Management in Horizon 2020" [2]. The Alliance of German Science Organizations, the Deutsche Forschungsgemeinschaft and the Helmholtz Association took up this idea in their guidelines for the Handling of Research Data.

Commenté [JW1]: This sentence can be adjusted in case other organizations (CNRS, RU) have similar guidelines.

The EMFL approves the principle of open access to research data. The EMFL facilities support its staff and guests to fulfill the requirements of funders and of the scientific community. To this end, it provides the necessary infrastructure for data management and regulates the access to data within the present policy. Open access to research data should be ensured wherever possible following the path of the citable data publication.

Commenté [JW2]: We may have to define EMFL facility. Each facility (or laboratory) should take care of the data independently. EMFL as head organization cannot handle all data of the four sites centrally.

2 General Principles

2.1 This policy sets the rules for the management of scientific data collected in the framework of publicly funded research at the EMFL research facilities. This includes the ownership, the storage, the curation of and the access to data.

2.2 The responsibility for the data management at a research facility must be clearly defined. The responsibility and the management process have to be documented as part of a Data Management Plan. Templates for such plans and support will be provided by the EMFL facilities.

2.3 For the data from proprietary research, users must set up a separate agreement with the management of each individual EMFL facility on how they wish their data to be managed before the start of any experiment.

2.4 Acceptance of this Data Policy and the "Terms and Conditions for the User Access" [3] is a condition of access to the EMFL research facilities.

Commenté [JW3]: That phrase might be erased or modified. HZDR has these Terms and Conditions, but EMFL as a whole does not have it. But maybe RU and/or CNRS/LNCMI has some, too.

3 Definitions

For the purposes of this policy:

3.1 The term *research data* refers to all data, which arise during the research process. This includes raw data, results, other data (e.g. validated data or simulation data) and the associated metadata.

3.2 The term *raw data* pertains to data collected from experiments performed on research infrastructures and to other primary research data. This definition includes data that are created automatically or manually by facility specific software and/or facility staff expertise in order to facilitate subsequent analysis of the experimental data.

3.3 The term *results* pertain to data that are outcomes arising from the analysis of raw data or intellectual property. This does not include publications.



3.4 The term *metadata* describes information pertaining to other data, including (but not limited to) the context of the experiment, the user, experimental conditions and other logistical information.

3.5 The term *public research* refers to research done during publicly funded experiments granted by the EMFL facilities usually after peer review of a proposal. Furthermore, all other research at the EMFL facilities is public research for the purposes of this policy.

Commenté [JW4]: Is that ok? Or is that specific for HZDR?

3.6 The term *proprietary research* refers to research done through purchased (commercial) access to the EMFL facilities.

3.7 The term *user* includes the proposer who has been granted access to the EMFL research facilities and any other person designated by him to perform the experiment (experimental team). Further persons may be added upon request of any member of the experimental team, if no other member objects.

3.8 The term *open access* or *openly accessible* means that data is made freely available to everyone.

4 Research Data Management

4.1 The sustainable utilization of research data requires a quality management. This covers the entire lifecycle of the data ranging from the data collection, processing and storage to a controlled deletion. The documentation of the processes and their application in the context of specific projects is usually carried out as part of a Data Management Plan (DMP). A checklist for creating a DMP is annexed (Appendix 1).

4.2 During data collection metadata should be recorded which allow conclusions about the context and the quality of the data collected. If possible, open and free data formats should be used. It is recommended to use the "DataCite Metadata Schema for the Publication and Citation of Research Data" (Appendix 2) [4].

4.3 Research data need to be long-term stored in suitable and trustworthy research-data infrastructures and be published if possible. For research data which constitute the basis of publications, safeguarding for 10 years corresponds to "good scientific practice" [5]. Appropriate data repositories are provided by EMFL or external facilities.

Commenté [JW5]: This reference is specific for HZDR. We might just erase it and leave the text as is.

5 Raw data and associated metadata

5.1 The user waives all copyright and related rights together with all associated claims and causes of action with respect to the raw data and associated metadata according to the Creative Commons CC0 Dedication (ref. 7.2) [6].

5.2 Raw data will be curated in well-defined formats, for which the means of reading the data will be made available by the EMFL facilities.

5.3 Associated metadata will be curated either within the raw data files, within an associated on-line catalogue, or within both.

5.4 Raw data and associated metadata will be stored by the EMFL facilities for at least ten years.



5.5 Access to raw data and the associated metadata is restricted to the user for a period of five years after the end of the experiment. Thereafter, it will become openly accessible with the EMFL facilities acting as custodian. Any user who wishes to maintain the restricted access to his data for a longer period will be required to make a special case to the EMFL-facility management. Data can always be made openly accessible earlier on simple request of any member of the experimental team, if no other member objects.

5.6 In deviation from 5.5, appropriate facility staff (e.g. instrument scientists, computing group members) has access to any EMFL-facility curated data or metadata for facility-related purposes. EMFL will undertake that they will preserve the confidentiality of such data.

5.7 Any member of the user has the right to create and distribute copies of his/her raw data.

5.8. In the Data Management Plan regulations are to be made, if and how data shall be deleted at the end of their life cycle and how this will be documented.

6 Results

6.1 Copyright and related rights with respect to results are not affected by this policy.

6.2 EMFL will provide means for the user to upload results and associated metadata to the facility and enable him/her to associate these results with raw data. The storage period for results is determined by the storage of associated raw data. The upload of results and associated metadata may be subject to volume restrictions.

6.3 The user is requested to ensure that associated metadata are as complete as possible, as this will enhance the possibilities for him to search for, retrieve and interpret his data in the future. For the long-term re-usability of data it is essential to use open or standard formats or to ensure that software to read / manipulate this data is made available.

6.4 Each EMFL facility is in charge of the curation of results stored in EMFL repositories.

6.5 The EMFL facilities undertake to provide means for the complementation of such metadata items that are not automatically captured by an instrument, in order to facilitate recording the fullest possible description of the raw data.

6.6 Access to results is restricted to the user. They may be made openly accessible on request of the principal investigator of the team. Paragraph 5.6 applies accordingly.

6.7 Publications related to experiments carried out at the EMFL facilities shall acknowledge the support of EMFL, including the facilities used, supporting staff or any other assistance.

7 Legal Requirements

7.1 The proposer must ensure in the design and preparation of the experiment that the raw data comply with the legal data protection regulations applicable in the respective country of the facility's site and do not contain any personal or otherwise particularly sensitive data. The EMFL is not responsible for the compliance of the user data with data protection legislation.



7.2 Research data may but do not need to be subject to copyright protection. For a transfer of rights for use or exploitation of data to third parties it has to be ensured, that the research data remain accessible at least for scientific purposes. It is recommended to assign an open license for non-exclusive right of use (CC BY 4.0) for copyrighted data [7]. For research data without any copyright it is suggested to use the public domain license (CC0 1.0) [6]. In any case the owner of the copyright should be specified in the DMP and the metadata.

7.3 EMFL makes no warranty or representation as to the accuracy, timeliness, completeness or correctness of data or software stored in EMFL repositories.

7.4 The access to research data constituting the basis of citable publications has to be ensured. This enables the exchange and reuse of this data with collaboration partners. In addition, verification of the scientific results is enabled in this way.

8 References

- [1] Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities, 2003, URL http://openaccess.mpg.de/67605/berlin_declaration_engl.pdf
- [2] H2020 Programme, Guidelines on FAIR Data Management in Horizon 2020, 2016, URL http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf
- [3] B110: "Safeguarding Good Scientific Practice and Proceeding in Case of Scientific Malpractice", HZDR institute regulation
- [4] DataCite Metadata Schema for the Publication and Citation of Research Data, https://schema.labs.datacite.org/meta/kernel-4.0/doc/DataCite-MetadadataKernel_v4.0.pdf
- [5] Terms and Conditions for User Access to the Experimental Facilities, HZDR Institute Regulation B210, URL <http://www.hzdr.de/db/Cms?pOid=44493>
- [6] Creative Commons Attribution CC BY 4.0 International, URL <https://creativecommons.org/licenses/by/4.0/>
- [7] Creative Commons Public Domain Dedication CC0 1.0 Universal, URL <http://creativecommons.org/publicdomain/zero/1.0/>

Appendix 1: Checklist for a Horizon 2020 Data Management Plan [2]

DMP component	Issues to be addressed
1. Data summary	<ul style="list-style-type: none"> <input type="checkbox"/> State the purpose of the data collection/generation <input type="checkbox"/> Explain the relation to the objectives of the project <input type="checkbox"/> Specify the types and formats of data generated/collected <input type="checkbox"/> Specify if existing data is being re-used (if any) <input type="checkbox"/> Specify the origin of the data <input type="checkbox"/> State the expected size of the data (if known) <input type="checkbox"/> Outline the data utility: to whom will it be useful
2. FAIR data 2.1. Making data findable, including provisions for metadata	<ul style="list-style-type: none"> <input type="checkbox"/> Outline the discoverability of data (metadata provision) <input type="checkbox"/> Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers? <input type="checkbox"/> Outline naming conventions used <input type="checkbox"/> Outline the approach towards search keyword <input type="checkbox"/> Outline the approach for clear versioning <input type="checkbox"/> Specify standards for metadata creation (if any). If there are no standards in your discipline describe what type of metadata will be created and how
2.2 Making data openly accessible	<ul style="list-style-type: none"> <input type="checkbox"/> Specify which data will be made openly available? If some data is kept closed provide rationale for doing so <input type="checkbox"/> Specify how the data will be made available <input type="checkbox"/> Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)? <input type="checkbox"/> Specify where the data and associated metadata, documentation and code are deposited <input type="checkbox"/> Specify how access will be provided in case there are any restrictions
2.3. Making data interoperable	<ul style="list-style-type: none"> <input type="checkbox"/> Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability. <input type="checkbox"/> Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?



2.4. Increase data re-use (through clarifying licences)	<ul style="list-style-type: none"><input type="checkbox"/> Specify how the data will be licensed to permit the widest reuse possible<input type="checkbox"/> Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed<input type="checkbox"/> Specify whether the data produced and/or used in the project is useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why<input type="checkbox"/> Describe data quality assurance processes<input type="checkbox"/> Specify the length of time for which the data will remain re-usable
3. Allocation of resources	<ul style="list-style-type: none"><input type="checkbox"/> Estimate the costs for making your data FAIR. Describe how you intend to cover these costs<input type="checkbox"/> Clearly identify responsibilities for data management in your project<input type="checkbox"/> Describe costs and potential value of long term preservation
4. Data security	<ul style="list-style-type: none"><input type="checkbox"/> Address data recovery as well as secure storage and transfer of sensitive data
5. Ethical aspects	<ul style="list-style-type: none"><input type="checkbox"/> To be covered in the context of the ethics review, ethics section of DoA and ethics deliverables. Include references and related technical aspects if not covered by the former
6. Other	<ul style="list-style-type: none"><input type="checkbox"/> Refer to other national/funder/sectorial/departmental procedures for data management that you are using (if any)

Appendix 2: Data Cite Metadata Schema v4.0 [8]

“There are three different levels of obligation for the metadata properties:

- **Mandatory (M)** properties *must* be provided,
- **Recommended (R)** properties are optional, but strongly recommended for interoperability and
- **Optional (O)** properties are optional and provide richer description.”

Table 1: DataCite Mandatory Properties

ID	Property	Obligation
1	Identifier (with mandatory type sub-property)	M
2	Creator (with optional name identifier and affiliation sub-properties)	M
3	Title (with optional type sub-properties)	M
4	Publisher	M
5	PublicationYear	M
10	ResourceType (with mandatory general type description sub-property)	M

Table 2: DataCite Recommended and Optional Properties

ID	Property	Obligation
6	Subject (with scheme sub-property)	R
7	Contributor (with type, name identifier, and affiliation sub-properties)	R
8	Date (with type sub-property)	R
9	Language	O
11	AlternateIdentifier (with type sub-property)	O
12	RelatedIdentifier (with type and relation type sub-properties)	R
13	Size	O
14	Format	O
15	Version	O
16	Rights	O
17	Description (with type sub-property)	R
18	GeoLocation (with point, box and polygon sub-properties)	R
19	FundingReference (with name, identifier, and award related sub-properties)	O

“Those clients who wish to enhance the prospects that their metadata will be found, cited and linked to original research are strongly encouraged to submit the Recommended as well as Mandatory set of properties.”