

Research Data Policy of the Helmholtz Center Munich

1. Introduction

Helmholtz Zentrum München, Deutsches Forschungszentrum für Gesundheit und Umwelt GmbH (hereafter Helmholtz Munich) recognizes the fundamental importance of research data and encourages all scientists to apply the highest standards when handling these data within their projects (for the definition of "research data" see 8.1. 3). This implies that accurate and easily retrievable research data must be an integral part of any research project at the center so that it reflects the scientific culture of our research activities recognized in society and in the scientific community. A generally binding research data policy is not only necessary for the verification and validation of research results, but also forms the basic prerequisite for open, transparent and networked science (Open Science Policy).

Due to the dynamics of the use of new analytical instruments, research data from current and also from already completed projects can generate additional value for future scientific projects. However, this can only be realized and protected if the data obtained is already protected against loss during the course of the project and if it remains reusable for future projects thanks to its format, the link with corresponding metadata and the assignment of persistent identifiers (DOI). Securing the research data obtained at the center for subsequent use is an essential prerequisite for future innovative (meta)analyses, e.g. to gain new knowledge on the complex interactions between environmental factors and human health.

2. Aim of this Research Data Policy at Helmholtz Munich

- The aim of securing the research data obtained at the center so that it can be reused for future projects is being pursued in the knowledge that the center's material and financial resources are provided by the public sector. Therefore, the center feels obliged to use the research results obtained with the greatest possible potential for broad application in society.
- With the implementation of a generally applicable research data guideline, including appropriate regulations for standardization, maximum synergy effects should be achieved by creating the basis for linking data sets from different projects.
- In particular, the research data guidelines also ensure that the research data already obtained at the center can be reused for future scientific questions (1, 2).
- A generally applicable research data guideline for the Helmholtz Zentrum München facilitates cooperation with research institutions in Germany and abroad and increases the visibility of the center in the international science community through the principles of Open Science (2, 3).

3. Scope of this Research Data Policy

The following guideline for handling research data applies to all employees of the Helmholtz Zentrum München, regardless of how the respective research project is financed (basic funding or funding by grants). It is also used - within the framework of the respective contractual agreements - for research data that was / will be obtained in the context of cooperation with external partners.

All employees of the center must follow this guideline when handling research data, regardless of the legal basis of their employment (as permanent researchers, employees on temporary projects, visiting scientists, scholarship holders, etc.).

The Center's current guidelines on data protection and information security, as well as the guidelines on the storage of and access rights to research data, take precedence over this guideline. In particular, research data as carriers of intellectual property may require protection (IP guideline of the center). Your exploitation rights by the center or third-party donors have priority over their publication (see 8.1.6) and must not be jeopardized by an open data culture.

Exceptions to the research data guidelines can only be granted in justified cases. They always require verification/approval by the INNO and SPR departments.

4. Research Data Management

Research data management covers the entire life cycle of research data, from the creation of a data management plan (DMP) to the collection of primary data, their validation, aggregation to permanent archiving and their publication.

Research data must be recorded in accordance with the relevant standards of the respective research discipline, stating the tools and methods used to generate and process them. These metadata records contain descriptions of the data context in the long term and to publish them if necessary. Your archiving should be retrievable, readable, exchangeable and reusable according to the four premises of the FAIR data standard (Findable, Accessible, Interoperable and Re-usable) (4). The link to their metadata should be made using the established and standardized terminology of the relevant discipline, taking into account the possibility of interdisciplinary reusability and interoperability (5).

The handling of research data within the framework of defined projects follows the DMP, which is created before the project begins (6). The planned type of research data, its acquisition, storage and access options to the data are to be specified in the DMP and the respective responsible parties are to be named. In the case of technical or methodological changes in the data flow during the course of a project, the DMP must be adapted. In the case of third-party funded projects, the costs incurred for the implementation of the DMP must be taken into account as eligible expenses when applying for funding. The project funding department of the SPR department advises applicants, e.g. B. on the basis of data management plans already established in other projects.

5. Responsibilities

5.1 Every researcher is responsible for:

A. the administration and management of research data (sets) in accordance with the principles and requirements of these guidelines and in accordance with the rules for ensuring good scientific practice of the Helmholtz Zentrum München (7),

B. the collection, documentation, archiving and protection against loss of research data including research-related records (metadata). For this, the corresponding DMP is to be used, which specifies standard

operating procedures that define the collection, administration, integrity, confidentiality, storage and publication of the data obtained.

C. Compliance with the general guidelines of the center, in the case of third-party funded projects, also the guidelines of the sponsors, and in the case of cooperation projects with external (research) partners, the relevant contractual regulations. The special requirements of the specific individual case must be specified in the DMP.

5.2 Each project leader (PL) is responsible for:

A. The creation and implementation of the data management plan (DMP) for the respective project. In addition to technical questions regarding the collection, administration, integrity, confidentiality, storage and publication of the data, compliance with the general guidelines of the center and the sponsors must also be taken into account in order to - wherever possible -

B. to ensure the reusability and exchangeability of the research data after the end of the project. In particular, this includes the definition of data reuse rights with the assignment of appropriate licenses, as well as the clarification of data storage and archiving with persistent DOIs after the project has been completed.

C. the possible specified destruction of research data and research-related records for data that falls within the scope of the European data protection BER as personal data (patient data, test person data (cohorts)). This also applies in particular to the revocation of consent by data subjects.

D. The provision of suitable resources (technical / organizational) to enable FAIR-compatible data management for all employees within the project.

5.3. Helmholtz Munich / Research Data Management (RDM) / HA DigIT

A. For research data management, the center provides support through RDM and other participating units (INNO, RA, ZB, SPR project funding) in the form of regular training, targeted advice (including questions about IP, ethics, etc.). Software is available for the standardized creation of DMPs.

The use of "electronic laboratory notebooks" (ELN) is supported to create valid and high-quality research (meta)data and minimize data loss.

B. The technical requirements for long-term archiving of research data and the associated metadata (research data repositories) are provided by the center.

6. Publication of scientific findings including Research Data

Research data should be made publicly accessible (publications, patents, etc.) immediately after they have been used for the first time by researchers at the center, unless there are significant reasons to the contrary. These can be, for example, legal or contractual regulations or the use of data from other research projects with a non-permissive license, the protection of intellectual property or the intended use of the data as part of technology transfer. The scope of the intended initial use must be specified in the respective DMP. In the case of contracts with external cooperation partners and guest scientists, care must be taken to ensure that no contractual commitments are entered into that would stand in the way of the publication of results involving the center. Advice on this may be provided by SPR project funding and/or the LA department.

When it comes to making the research data accessible after initial use (for instance in the form of a peer reviewed research paper), the focus is on traceability and reusability of the data in accordance with the FAIR principles (see above). For this purpose, the research data repositories of the center or external repositories are to be used. The Helmholtz Zentrum München provides tools for this and advises the project leaders on how to use them.

If the research data is made publicly accessible, this should be done under a CC-SA license if possible. As far as is feasible, data that has not yet been published but is worthy of publication should subsequently be made freely accessible and - linked to a DOI - published in the form of citable data publications. The DOI is assigned by the central library of Helmholtz Munich.

7. Annex

7.1 Definitions

7.1.1 Data Management Plan (DMP)

In order to ensure sustainable data management, a data management plan (DMP) must be drawn up before the start of a research project. It is intended to ensure systematic and permanent handling of the research data generated as part of a research project. In addition, he goes into the copyright, usage, access and first exploitation rights as well as storage during and after the end of the research project.

7.1.2 Researcher/Scientist

Scientists are all employees of the research center who are active in research and have a university degree, including bachelor's and master's students and guests with the appropriate qualifications.

7.1.3 Research Data (RD)

Research data is all data that is created during the research process, is used for it or is the result of it. Depending on the specific research question and using various methods, they are generated or collected, observed, simulated or derived, then processed, validated, further processed, analyzed, published and finally archived. Accordingly, research data occurs in every research discipline and depending on the stage of its life cycle in different media types and formats, aggregation and quality levels. Research data that form the basis of a publication are part of the scientific output of the center's scientists.

7.1.4 Metadata

Metadata describe research data. On the one hand, they include authorship(s), contact details of the originator (organization), time of creation, licenses and keywords. They define the data formats used and contain the context that led to the data. In addition, metadata can include device settings, environmental conditions (e.g. temperatures, pressures), comments and measurement uncertainties. Metadata are essential for the subsequent use of research data. In addition to use for further research work, this subsequent use also includes verification of research results by third parties.

7.1.5 Research Data Management (RDM)

The term research data management refers to the handling of research data from planning, generation and processing to publication, long-term archiving and, if necessary, deletion in compliance with the rules for safeguarding good scientific practice and the European GDPR. The management of research data therefore affects the entire life cycle of research data.

In addition, the term includes the subject-specific documentation of the processes in the context of their collection. DMP facilitate the documentation of these processes and the description of the data.

7.1.6 Protection of Intellectual Property and the priority right for RD exploitation

The need for protection and the exploitation rights of intellectual property (IP) are defined in the employment contract between the researcher and the center as worker or host and are described in detail in the current IP guideline of the center. Copyright and exploitation rights to all forms of intellectual property can be expanded or specified by further agreements (e.g. third-party funding contracts).

In principle, Helmholtz Munich has the right to decide how research data can be published and shared as a carrier of intellectual property. The treatment of all research data according to the FAIR principles represents the basis for publications and data exchange, but does not enforce the latter. The need for protection of the intellectual property resulting from the research of the center has priority over the exchange of data within the framework of a future open science policy.

The first use of research data is usually one or more text publication(s) in a journal based on this data. The scope and duration of the initial use can be determined by the scientists in the DMP. In particular, the end of the initial exploitation phase should be determined in a context-dependent and justified manner.

7.1.7 Data Publication

Data sets can be published in two ways:

A. by simply making them available via a web interface, preferably at a local data repository of Helmholtz Munich, or via a citable, dedicated data publication (linked with a DOI). Such a data publication can include the open distribution of the data and metadata, but does not have to. Instead it can also show a simple "landing page", which broadly describe the context and content of the data set and via which interested parties can submit a request to the respective decentralized data manager (see above under "Responsibilities and responsibilities") for data release.

A central technical tool for the publication of research data is a data repository, a server service on which data can be stored by its creators. This data is given a globally unique identifier (e.g. a DOI) and can be searched and, if necessary, downloaded. A restriction of access by the data producer(s) by means of authorization settings is possible depending on the respective repository.

7.1.8 Licences

The license of a publication is a contract between the publishing and re-using person or institution, which regulates the conditions of re-use. Many license "types" are already pre-formulated and are available on the Internet, so reference is made to them here (8). A permissive license restricts subsequent use only minimally. Well-known examples of permissive licenses are CC-0-SA and CC-BY of the Creative Commons.

→ CC-0: <https://creativecommons.org/publicdomain/zero/1.0/deed.de>

→ CC-BY: <https://creativecommons.org/licenses/by/4.0/deed.de>

7.2 Legal Status of Research Data

Ownership of data does not exist in Germany. However, copyright guarantees everyone the protection of their intellectual creations. Whether research data is subject to the protection of copyright law depends on whether the requirements for the level of intellectual creation or the requirements of database law are met. Both requirements are not regularly met for research data collected in routine measurements. Depending on the type of collection, research data can have a level of intellectual creativity that is worthy of protection. In order to ensure legal certainty, it should always be assumed that it is worthy of protection, that is rights of use and exploitation are usually to be clarified contractually with external partners. Careful licensing of published data must also be ensured.

Regardless of the right to the data obtained, the question of the right of exploitation must be seen. The researcher or the research institution can assert a (temporary limited) right of first use of the data obtained as part of their research activities in the form of publications or patents.

References

- (1)Strategiepapier des Forum Gesundheitsforschung des BMBF „Nutzbarmachung digitaler Daten für KI - und datengetriebene Gesundheitsforschung“, Berlin 2020
- (2)Berliner Erklärung über den offenen Zugang zu wissenschaftlichem Wissen, Helmholtz Open Science 2003, : <https://openaccess.mpg.de/Berliner-Erklaerung>

- (3) Positionspapier des Deutschen Wissenschaftsrates „Digitalisierung und Datennutzung für Gesundheitsforschung und Versorgung“, Köln, Juli 2022
- (4) Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3:160018. doi: 10.1038/sdata.2016.18 (2016).
- (5) LEARN Toolkit of Best Practice for Research Data Management. <http://learn-rdm.eu/en/research-data-management-toolkit-now-available/>
- (6) Empfehlungen für Richtlinien der Helmholtz-Zentren zum Umgang mit Forschungsdaten. In der 109. Mitgliederversammlung der Helmholtz-Gemeinschaft am 13.-14. September 2017 <https://doi.org/10.2312/os.helmholtz.002>
- (7) https://hint.helmholtz-munich.de/fileadmin/Intranet/Recht_und_Compliance/Interne_Regelungen/Dokumente_DE/Regeln_zur_Sicherung_guter_wissenschaftlicher_Praxis_06-10-2015.pdf
- (8) <https://creativecommons.org/licenses/>

Munich, 01-01-2023