Policy for Research Data Management at *liv*MatS

Introduction

Data are the backbone of science and thus one of the pillars upon which social, economic and technological progress rests. As researchers, we are aware of the immense value of data, and with it come important moral, ethical, and practical questions about the proper use of data.

Research Data Management (RDM) policies help answer these critical questions by providing a guide to best practices for handling, storing, maintaining, and archiving data. Following these practices benefits researchers by promoting discovery and efficiency, ensuring validity, and facilitating collaboration.

Beyond the personal benefits, rigorous RDM practices also build trust with our public stakeholders. As publicly funded institutions, we rely on a transparent and symbiotic relationship with the public - and as such, livMatS policies align with the prevailing call to action to make data **FAIR**: **F**indable, **A**ccessible, **I**nteroperable, and **R**eusable [1].

To help navigate the landscape of data management, we hereby introduce the *liv*MatS Research Data Management (RDM) Policy.

Our commitment

The Cluster of Excellence *liv*MatS aims to make data more FAIR - and in a broader sense, we are committed to taking an active role in the movement towards Open Science. To this end, we strive to handle data in a manner consistent with the *commitment to active data and software management in large research alliances* [2]. To support our researchers in their efforts to make their data FAIR, *liv*MatS offers appropriate training and accompanying tools.

Purpose

The goal of this policy is to align our RDM practices and principles with an Open Science framework - including Open Access - to enable more equitable and transparent research, create synergistic collaborations, and further promote the value of research in society. This policy also encourages the publication of open data, while recognizing that the release of data derived from publicly funded research requires consideration of the necessary boundaries of openness. The *liv*MatS RDM Policy supports researchers in complying with legal requirements and new conditions for funding and scientific publication related to personal data and commercial considerations. *liv*MatS strives for consistency in research practices related to data management principles that support effective data sharing, including Open Access, and the need for data to be discoverable, accessible, reusable, interoperable, and meet certain quality standards.

This policy concretizes the policy for handling research data at the University of Freiburg (0.5.0, German [3]). It defines a minimum set of binding principles for data management to which *liv*MatS and all affiliated researchers should adhere.

Resources for technical guidance and tools to help a project comply with this policy can be found at www.livmats.de [4].

Principles for handling research data

The following principles serve as a guide on the way to FAIR data, i.e., to improve the following aspects.

- **Findability**: data and their documentation (metadata) should be easy to find for both humans and machines.
- **Accessibility**: once users find the data they want, they need to know how to access them, possibly including authentication and authorization.
- **Interoperability**: data typically need to be integrated with other data. In addition, data must interoperate with applications or workflows for analysis, storage, and processing.
- **Reusability**: data and their documentation (metadata) should be well described so that they can be replicated and/or combined in different environments.

Applying the following principles requires little effort and increases the quality and reproducibility of one's work, promotes visibility of scientific contributions, and facilitates research collaboration. Electronic laboratory notebooks (ELNs) or other appropriate platforms for managing research data, in conjunction with appropriate data repositories, will readily satisfy many of these principles without the need for manual intervention.

Data storage and access

The following principles on storage and access offer guidance on how to store data and who should have access:

- Store research data in a correct, complete, unadulterated and reliable manner to preserve their integrity.

- Bundle data in datasets of suitable granularity. This can mean one dataset per project or one dataset per measurement, depending on what the data creator, the data creator's group and their community deem appropriate. This promotes data reusability.

- Store data in a suitable repository or archiving system. This supports findability and accessibility.

- Make all data available *liv*MatS-wide on a suitable access-restricted repository provided there are no sensitivity aspects forbidding to do so. Primary data must enter such repository unaltered. This internal sharing contributes to findability and accessibility on the group-wide or *liv*MatS-wide level even before publication.

- Do not grant unrestricted access to such internal repositories to external partners. This acknowledges the fact that openness must have limits.

Data documentation

Proper documentation give data a context and thereby make data understandable to others. This can be in the form of plain text, machine-readable metadata, or a combination of both.

- Document all research data at creation. This will help increasing the data's findability and reusability.

- Where possible, use metadata standards and ontologies that are already well established in the specific discipline. This works towards better interoperability.

- Always identify as the data creator by Open Researcher and Contributor ID (ORCID). This plays into findability and interoperability.

- Use other standardized resource identifiers following schemes recommended by DataCite [5] wherever applicable. This contributes towards interoperability as well.

- Clearly document the provenance of secondary (derived) data. This increases the data reproducibility and hence reusability.

- Assign all data with unique identifiers such as universally unique identifiers (UUIDs). This usually happens automatically in the background if working with an ELN or other RDM solutions and supports unique identification and findability even before data publication.

Data publication

Same as when publishing a journal article, publishing data means sharing a product of academic research with the world under certain intellectual property terms.

- Together with a (textual) publication, publish all underlying (primary) data in a timely fashion. This is an important aspect of reproducibility.

- All data should be reviewed by a *liv*MatS PI or RI ahead of publication. Once the review is complete, make a record of the responsible reviewer including their name and ORCID. This record is kept within the published data's documentation. This quality control mechanism acknowledges the fact that not all data are worth publishing. It assures data integrity and increases reproducibility and reusability.

- *liv*MatS strongly encourages assigning permissive Open Access licenses to published research data. This is critical for enabling reusability.

- Assign digital object identifiers (DOIs) as persistent identifiers to data at publication. This ensures globally unique identification and findability over time.

Limits of openness

- *liv*MatS acknowledges that openness has limits, for example to protect personal data, commercial interests or for security concerns. Beyond said concerns, *liv*MatS recognizes the reluctance of data creators to share data assets before these have been fully exploited.

- Ongoing research should not be compromised by premature data release. If deemed necessary, *liv*MatS hence tolerates a period of exclusive use of up to two years between data creation and unrestricted publication.

- In preference to employing an exclusive use period, release data immediately with an embargo. A user requesting access to embargoed data must enter into a data use agreement with the data provider. The embargo period should not apply to those data that support published findings and which are necessary for validation.

- Before any publication, evaluate data for the possibility of establishing intellectual property. *liv*MatS encourages consulting with the *liv*MatS Intellectual Property Officer.

- Not all data are worth sharing beyond the scope of the concerned project. Keeping a concise up-to-date data management plan for every project, as described below, helps to identify which significant and potentially valuable data should be shared — for example to colleagues within livMatS or for publication.

Data retention

Clear rules on data retention and deletion guarantee the long-term availability of data.

- The minimum archival duration for research data and records is 10 years after either the assignment of a persistent identifier or at the time of publication of related work following project completion, whichever comes later.

- For many domains,longer periods apply due to legal requirements. This is, for example, the case for the documentation of personnel handling hazardous substances and is regulated within the ordinance of hazardous substances [6].

- No data, neither published nor unpublished, can be deleted before the end of this obligatory retention period. Data may, however, be marked as invalid or retracted in a suitable manner.

Data management plans

Specific details regarding data management are offered in data management plans (DMPs). Each DMP is created for a corresponding project within *liv*MatS and is continuously updated throughout the project. *liv*MatS provides template DMPs for recommended research data management systems [4]. DMPs are generally not immutable, but rather are intended to document data management over the course of each project. Certain aspects regarding the implementation of this RDM policy, however, are non-negotiable and clearly marked as such in the template DMPs.

Responsibilities, rights, duties

The responsible management of research data involves multiple stakeholders, namely: *liv*MatS researchers, the Excellence Cluster *liv*MatS, and the University of Freiburg. Each stakeholder has a different role and responsibilities and should comply with codes for the responsible conduct of research (Ordnung der Albert-Ludwigs-Universität zu Sicherung der Redlichkeit in der Wissenschaft [7]).

Researchers are responsible for:

a. Management of research data and datasets in adherence with the principles and requirements expressed in this policy;

b. Collection, documentation, archival, access to and storage of, proper destruction of research data and research-related records. The DMP includes the specific definition of protocols and responsibilities within a joint research project. Researchers will produce a DMP for every research project and commit themselves to continued updates as a means of documentation in the course of the project;

c. Compliance with the general requirements of the funders and the research institution; special requirements in specific projects should be described in the DMP;

d. Planning to enable the continued use of data even after project completion. This includes defining post-project usage rights, with the assignation of appropriate licenses, as well as the clarification of data storage and archiving in the case of discontinued involvement at *liv*MatS;

e. Backup and compliance with all organizational, regulatory, institutional and other contractual and legal requirements, with regard to both research data and the administration of research records.

*liv*MatS is responsible for:

a. Empowerment of organizational units, providing appropriate means and resources for research support operations and employee education, enabling researchers to exercise their responsibilities as outlined above;

b. Support of data handling practices beginning with the project proposal. This is possible through the drafting and provision of template DMPs, monitoring, training, education and support, codes of conduct, and other relevant guidelines;

c. Careful selection and provision of services and infrastructure via the University of Freiburg's central services and third-party providers. Such central services and third-party providers may refer to the University of Freiburg's central computing services, its research data management group, the library services as well as external providers. *liv*MatS closely monitors the upkeep of such services and infrastructure and assures their compliance with obligations to third-party funders or other legal entities.

University of Freiburg's central services and third-party providers are entrusted with:

a. Developing and providing mechanisms and services for the storage, safekeeping, registration and deposition of research data in support of current and future access to research data during and after the completion of research projects;

b. Providing access to services and infrastructures for the storage, safekeeping, and archiving of research data and records.

Intellectual property rights

Intellectual property rights (IPR) are defined in the work contract between a researcher and their employer. IPRs might also be defined through further agreements (e.g. grant or consortial agreements). In cases where the IPR belong to the institution that employs the researcher, the institution has the right to choose how to publish and share the data.

For data that form the basis of protectable intellectual property, one is generally obligated to file an invention disclosure pursuant to Sections 5, 42 of the German Employee Invention Act (Arbeitnehmererfindungsgesetz) .

Jurisdiction

This policy for the management of research data applies to all researchers active at *liv*MatS and was approved by the *liv*MatS board on 2022-09-30.

Validity

This policy will be reviewed and updated as required by the Research Data Management Steering Board of *liv*MatS every year. The *liv*MatS board is responsible for appointing the Research Data Management Steering Board.

Annex: Definitions

1. **Research** is any creative and systematically performed work with the goal of furthering knowledge, including discoveries regarding people, culture and society, in addition to the use of such knowledge for new applications.

2. **Researchers** refers to all research-active members of an institution including employees and doctoral candidates. Persons not directly affiliated with an institution, but who make use of or are physically present at the institution for research purposes, are also included in the term. Visiting researchers or collaborators may also be expected to comply with the policy.

3. **Research data** refers to all information (independent of form or presentation) needed to support or validate the development, results, observations or findings of a research project, including contextual information. Research data include all materials which are created in the course of academic work, including: digitization, records, source research, experiments, measurements, surveys and interviews. This includes software and code. Research data can take on several forms: during the lifespan of a research project, data can exist as gradations of raw data, processed data (including negative and inconclusive

results), shared data, published data and Open Access published data, and with varying levels of access, including open data, restricted data and closed data.

4. A **Data Management Plan (DMP)** is a structured guideline which depicts the entire lifeline of data and its technical implementation. DMPs can be updated if needed in the course of a project. Data management plans must assure that research data are traceable, available, authentic, citable, properly stored and that they adhere to clearly defined legal parameters and appropriate safety measures governing subsequent use.

References

[1]   M. D. Wilkinson et al., "The FAIR Guiding Principles for scientific data management and stewardship," Scientific Data, vol. 3, no. 1, 2016, doi: 10.1038/sdata.2016.18.

[2]   C. Brecher et al., "Commitment zu aktivem Daten- und -softwaremanagement in großen Forschungsverbünden: Commitment to active data and software management in large research alliances," Bausteine Forschungsdatenmanagement, no. 1, Art. no. 1, Mar. 2022, doi: 10.17192/bfdm.2022.1.8412.

[3]   Policy zum Umgang mit Forschungsdaten an der Universität Freiburg, review 0.5.0

[4]   Data Management Plan (DMP) Templates, https://www.livmats.uni-freiburg.de/en/members/research-data-management

[5]   Table 3, ID 2.4, 2.4a, 2.4b, DataCite Metadata Working Group, "DataCite Metadata Schema Documentation for the Publication and Citation of Research Data and Other Research Outputs v4.4,", 2021, doi: 10.14454/3W3Z-SA82.

[6]   BAuA - Arbeiten mit Gefahrstoffen - Gefahrstoffverordnung (GefStoffV) (Diese Fassung der Verordnung trat am 01.10.2021 in Kraft) - Bundesanstalt für Arbeitsschutz und Arbeitsmedizin. https://www.baua.de/DE/Themen/Arbeitsgestaltung-im-Betrieb/Gefahrstoffe/Arbeiten-mit-Gefahrstoffen/pdf/Gefahrstoffverordnung.pdf (accessed May 18, 2022).

[7]   Ordnung der Albert-Ludwigs-Universität zur Sicherung der Redlichkeit in der Wissenschaft, https://uni-freiburg.de/forschung/redlichkeit-in-der-wissenschaft/