1st Conference on Research Data Infrastructure

Poster presentations II

https://doi.org/10.52825/CoRDI.v1i.378

Published: 07 Sept. 2023

FAIRmat guide to writing data management plans

A practical guide for the condensed-matter physics and materials-science communities

Ahmed E. Mansour[1] [https://orcid.org/0000-0002-3411-6808], Lucia Rotheray[1] [https://orcid.org/0009-0009-8969-1867], Kerstin Helbig[2] [https://orcid.org/0000-0002-2775-6751], Silvana Botti[3] [https://orcid.org/0000-0002-4920-2370], Heiko B. Weber[4] [https://orcid.org/0000-0002-6403-9022], Martin Aeschlimann[5] [https://orcid.org/0000-0003-3413-5029] and Claudia Draxl[1] [https://orcid.org/0000-0003-3523-6657]

[1] Physics Department and IRIS Adlershof, Humboldt-Universität zu Berlin, Germany

[2] Computer and Media Service, Humboldt-Universität zu Berlin, Germany

[3] Institut für Festkörpertheorie und -optik, Friedrich-Schiller-Universität Jena, Germany

[4] Department of Physics, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Germany

[5] Department of Physics and Research Center OPTIMAS, Technische Universität Kaiserslautern, Germany

Abstract: Research data management is becoming an increasingly important topic due to the growing amounts of types, formats, and volumes of data produced by scientific research. In addition, a growing demand to make data accessible and comprehensible requires standardizing, managing, and planning the data life-cycle. For this reason, many funding agencies now require a data management plan (DMP) as part of submitted research proposals. While some of them and other scientific bodies offer DMP templates, there is no one-size-fits-all solution, due to the heterogeneity of data generated by different scientific disciplines. Here, we present as an example FAIRmat's effort in enhancing data literacy on the topic of DMP aiming to guide physicists and materials scientists to writing DMPs that comply with the requirements of the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG).

Keywords: Data management plan, Data life-cycle, FAIR data, NOMAD, Materials Science

1. Introduction

Scientific research relies on the dissemination of research results. Traditionally, this process takes the form of conference contributions or publications. While such channels are currently accepted as the norm in the scientific community, challenges of reproducibility and accessibility arose along with concerns related to research integrity and compliance with ethical and legal requirements [1]. This has led to the advocacy of open access, the establishment of standards for reporting research methods and data availability statements to ensure reproducibility, and the enforcement of policies governing ethical and integrity issues [2–5]. As a result, there has been a paradigm shift in the research process from being linear – utilizing generated data towards a publication as the main outcome – to becoming a cyclic process – focusing on research data as key output as shown in Figure 1 [6,7]. Proper handling of research data requires a comprehensive management process, starting with planning prior to the project initiation, monitoring and controlling during the project, and defining the long-term fate of the data after the project closing. Such a management process is described by a data management plan (DMP).

Figure 1: Data life-cycle in research projects.

The German Research Foundation (DFG) now requires a detailed specification of the handling of research data to be included in research proposals [8]. This requirement is formulated as a catalogue of questions and is based on the criteria defined by the Science Europe association [9]. The DFG checklist is a high-level document that covers the diverse research disciplines [10]. It is crucial to adapt these requirements to each discipline, and it is up to the research communities to develop standards and best practices for preparing a domain-specific DMP.

2.The FAIRmat Guide

A core mission of FAIRmat is to support the scientific community to introduce and maintain high standards of reproducibility, research integrity, and compliance with ethical and legal requirements by adopting proper RDM practices based on the FAIR data principles [4]. A recent implementation of this mission is the FAIRmat guide to writing a research data management plan (see Figure 2), tailored for scientists in the fields of condensed-matter physics and materials science.

Figure 2: Front and back covers of FAIRmat's DMP guide.

The FAIRmat guide provides explanations of the elements of DMPs and practical tips for best practices in data management. It is prepared to meet the requirements for handling research data set by the DFG. The elements of the DMP described in the guide (as defined by the DFG requirements) are shown in Figure 3.

The preparation of this practical guide began with detailed discussions with established scientists in the field, covering both computational and experimental solid-state physics to identify the needs, currently available standards, and prospects for an effective DMP.

An effective DMP starts with a detailed description of what needs to be managed, i.e., the data generated or reused in the research project. This description forms the first section of the DMP, and guides all other sections that follow. Next, the plan should establish a comprehensive documentation for the project data through a structured organization and adopting a comprehensible naming convention for the data files. A key component is the use of standard and community-accepted metadata and ensuring the availability of rich metadata.

Management of the data storage, accessibly, and dissemination is described for both the short term (during the project) and the long term (after the project has ended). Our guide explains data repositories, their types and requirements, and refers the reader to resources for selecting a data repository suitable for subject of research and data type, such as the Science Europe criteria and the re3data directory [9,11].

Data description	• Comprehensive description of data generated or reused in the project: data type, source, format and volume.
Documentation and data quality	• A record of all the workflows for generating and structuring data, associating metadata, and how data quality will be ensured.
Storage and technical archiving	• Methods and locations for secure and accessible storage of the data and detailed backup procedure.
Data exchange and long-term accessibility	• Plan for the data fate after the project ends. It includes description of both preservation and dissemination.
Legal obligations and conditions	• Identified legal issues relevant to the project data, such as: ownership, intellectual property, copyright and licensing laws.
Responsibilities and resources	• List of tasks, assignments for project members, and costs relevant to data management

Figure 3: List of DMP sections and contents that meet the DFG requirements.

3. Conclusions

Our guide assists the research community in preparing an effective DMP to be submitted in accordance with the requirements of research proposals. This fills a gap in available resources on research data management for scientists. In addition to helping researchers to meet the minimum requirements for funding, our goal is to ensure the long-term success of research projects and improve the quality of the research process. Continuous improvement and regular updates are planned to ensure that the guide fulfills the latest standards and requirements of the DFG and other funding bodies.

Competing interests

The authors declare that they have no competing interests.

Funding

FAIRmat is a consortium of German National Research Data Infrastructure (NDFI) funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project 460197019.

References

[1] Deutsche Forschungsgemeinschaft (DFG), Guidelines for Safeguarding Good Research Practice. Code of Conduct, 2022.

[2] European Commission, *The EU Open Science Policy*, https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science_en#documents.

[3] M. Scheffler et al., *FAIR Data Enabling New Horizons for Materials Research*, Nature **604**, 635 (2022).

[4] M. D. Wilkinson et al., *The FAIR Guiding Principles for Scientific Data Management and Stewardship*, Sci Data **3**, 160018 (2016).

[5] L. M. Ghiringhelli, C. Carbogno, S. Levchenko, F. Mohamed, G. Huhs, M. Lüders, M. Oliveira, and M. Scheffler, *Towards Efficient Data Exchange and Sharing for Big-Data Driven Materials Science: Metadata and Data Formats*, NPJ Comput Mater **3**, 46 (2017).

[6] ELIXIR Research Data Management Kit (RDMkit), *Data Life Cycle: Sharing*, https://rdmkit.elixir-europe.org/sharing#what-should-be-considered-for-data-sharing.

[7] K. Briney, *Data Management for Researchers: Organize, Maintain and Share Your Data for Research Success* (Exeter: Pelagic Publishing, UK, 2015).

[8] The Deutsche Forschungsgemeinschaft (DFG), *Specification of Requirements Relating to the Handling of Research Data in Funding Proposals*, https://www.dfg.de/en/research_funding/announcements_proposals/2022/info_wissenschaft_22_25/.

[9] Science Europe, Practical Guide to the International Alignment of Research Data Management, 2018.

[10] The Deutsche Forschungsgemeinschaft (DFG), Handling of Research Data: Checklist for Planning and Description of Handling of Research Data in Research Projects, 2021.

[11] re3data.org, *Registry of Research Data Repositories*, https://doi.org/10.17616/R3D.