Introduction

This guidance document is intended to provide discipline-specific recommendations for applicants and reviewers, supplementing existing guidelines of general applicability, such as the Guidelines on the Handling of Research Data and the Code of Conduct "Guidelines for Safeguarding Good Research Practice".

Research data in mathematics are as diverse as the various mathematical disciplines themselves. Relevant research data may arise in every subdiscipline of mathematics, and their handling may therefore be an important aspect of any mathematical research project. In interdisciplinary research projects, care must be taken to ensure that research data are handled in accordance with the requirements of all disciplines involved.

Describing how research data will be handled during the proposed research project is not an end in itself. Rather, it is intended to encourage the deliberate and appropriate handling of research data in mathematical research projects. Ultimately, the purpose of this requirement is to improve research by ensuring that research data are managed in accordance with current best practice.

The Review Board welcomes further efforts to standardise data collection and storage across all mathematical subdisciplines. When preparing a DFG proposal, applicants are encouraged to follow the research data management practices established as best practice by the relevant international research community, where such practices already exist. For example, the existing NFDI consortium MaRDI provides information on discipline-specific best practices for handling research data through illustrative use cases and tutorials (https://www.mardi4nfdi.de/resources/publications).

Specifically, the Review Board recommends that applicants take account of the FAIR principles for research data (findability, accessibility, interoperability and reusability) when submitting proposals. The DFG's Checklist for Handling Research Data identifies the following six key areas, on which applicants should provide information that is as specific and binding as possible in their DFG proposals with regard to the planned research project:

1. Data description
2. Documentation and data quality
3. Storage and technical safeguarding during the project
4. Legal obligations and framework conditions
5. Data sharing and long-term accessibility
6. Responsibilities and resources

The Review Board points out that the proposal should already make clear whether the research data management plan will entail costs, for example for making research data accessible and reusable or for transferring data to a public repository. Where such costs are not covered by an adequate level of basic institutional resources, applicants may request funding for them as project-specific costs.

1. Data description

The nature and volume of the data to be generated or used in the project, as well as the software used to generate and/or process them, should be considered before the proposal is submitted. This will also help assess the effort involved in data management and any associated project-specific costs. The proposal should present these considerations

coherently and provide at least a summary of the types and volumes of data and software involved.

Where particular data formats have already become established in the international research community, the Review Board recommends using them. Wherever possible, software should be developed as packages for established systems recognised by the community, preferably open-source systems.

2. Documentation and data quality

The proposal should briefly explain the main steps involved in preparing, quality-assuring and documenting the research data and, where applicable, associated software (see also the Code of Conduct "Guidelines for Safeguarding Good Research Practice", Guideline 12). In the Review Board's view, documenting research data to an appropriate standard is the most fundamental step in sound research data management.

In the Review Board's view, unambiguous identification of data is essential for their documentation, attribution and subsequent use.

Organised filing, for example by project or publication, can also contribute to transparent data documentation. In the Review Board's view, responsibility for quality control rests with the participating project staff and the applicants (project leaders). Comprehensive induction in research data management is essential for project staff. Applicants (project leaders) must ensure that appropriate training takes place. The proposal may explain whether this will be provided, for example, through introductory events, training courses or induction by experienced members of the collaboration. In coordinated programmes, such training could be organised by the spokesperson or coordinator, for example through integrated doctoral training programmes or central projects. This enables doctoral researchers, postdoctoral researchers and student research assistants to be trained in careful and sustainable data documentation at an early career stage.

In the Review Board's view, the data quality required for a particular project generally falls within the expertise and responsibility of the applicants, as does the subsequent selection of data for archiving and long-term accessibility.

3. Storage and technical safeguarding during the project

The proposal should explain how research data, and any associated software, will be stored (backups, write protection, access rights, filing system, etc.). As outlined under "2. Documentation and data quality", an organised filing system and orderly, traceable storage are recommended, for example by project or publication. Existing requirements of the relevant research institution should be mentioned. In general, research data from all DFG-funded projects must be archived for ten years (Code of Conduct "Guidelines for Safeguarding Good Research Practice", Guideline 17). This applies to both analogue and digital data.

4. Legal obligations and framework conditions

Legal obligations and framework conditions should be reviewed before the project begins and briefly described in the proposal. These obligations cover the collection, documentation, storage, archiving and reuse of data generated or used. Where third parties hold rights to the data, any resulting restrictions should be documented.

5. Data sharing and long-term accessibility

The proposal should make clear which data are considered worthy of external archiving and why. Long-term accessibility requires not only archiving but also, wherever possible, reliable access (see also the Code of Conduct “Guidelines for Safeguarding Good Research Practice”, Guideline 13). In the Review Board’s view, all relevant and reproducible data underlying publications should be archived for the long term, in addition to data already accessible through publications. This may also include software if it was developed specifically for the project and/or is required for the reuse of archived data.

Project leaders should select these data judiciously, on their own responsibility, and provide reasons for their selection. Digital data should be archived in certified repositories or data centres. Applicants are advised to consider the project-specific requirements for a repository before the project begins, identify suitable repositories on that basis and, where appropriate, contact their operators. The discipline-specific services offered by NFDI consortia, whose use is strongly welcomed by the Review Board, deserve particular mention here. An overview of existing infrastructures is provided by the RIsources portal for research infrastructures and the re3data registry of research data repositories (https://www.re3data.org/). For mathematics, suitable options for storing software and research data include Zenodo, an EU-funded repository hosted at CERN that also offers extensive assurances of long-term preservation (https://zenodo.org/).

Early contact with service providers is recommended so that the time and financial resources required can be taken into account in the research project in good time. Archivable, standardised data formats should be used to facilitate subsequent reuse. In the Review Board’s view, technical compatibility between different datasets and software solutions presents another major challenge for data management. If the project develops practical solutions for converting data formats, it is desirable to make these solutions available for use in order to further advance the general standardisation of data formats.

6. Responsibilities and resources

The proposal should clearly set out who is responsible and accountable for the individual steps of research data management during the project. The effort involved should also be taken into account, particularly when requesting project-specific funding. In the Review Board’s view, responsibility for sound data management itself, as well as for final quality control, rests with the project leaders.