# Guide to handling research data in DFG proposals

This document is provided by the central FDM service team of Ruhr-Universität Bochum (RUB) and is intended for RUB members. Version as of 2022-10-10.

## General notes:

* The DFG expects applications for research funding to include answers to the questions from the following checklist: https://www.dfg.de/download/pdf/foerderung/grundlagen\_dfg\_foerderung/forschungsdaten/forschungsdaten\_checkliste\_en.pdf
* A clear, concise and transparent outline helps in the evaluation of the application and is a plus point in the evaluation of the application.
* Please also note the subject-specific recommendations: <https://www.dfg.de/en/research_funding/principles_dfg_funding/research_data/recommendations/index.html>
* The checklist includes ideas and aspects that should be considered in data management at the beginning and/or during the project.
* Not all of it applies to your data!
* When applying for funding, it is not necessary to have worked out a comprehensive solution for all points. However, you should at least have thought about them and express your ideas in the application.
* This checklist supports you in focusing on the data that will be re-used or created throughout the project and in considering them from different perspectives. The goal is to document the research process as a basis for potential reuse options for the data.

## Data description

**Original questions of the DFG checklist on data description**

* In what way does your project create new data?
* Will any existing data be reused?
* Which types of data, in the sense of data formats (e.g. image data, text data or measurement data) are created in your project and how are they processed?
* To what extent do they arise or what data volume is to be expected?

Explanation:

In this section, describe the data you reused in your project and the new data you created. Make clear whether you are reusing data or creating data yourself. Describe as comprehensively as possible what methods you use, and what type of data is created or will be used. If data are used that are not generated as part of the project, indicate the context from which the data originate, preferably as a citable source (e.g., via DOI). Otherwise, it is recommended to indicate that no suitable data are available according to your own research.

Additional notes:

Be clear that a systematic approach is being taken towards the data. Also describe how you document the individual steps. It can also be helpful to describe a system for naming file folders and files. In doing so, address the data types/data formats and their further use as well as the expected size of the data as a whole.

In addition, indicate which data volumes will only be needed in the short term for your work and which will have to be stored for the long term.

## Documentation and data quality

**Original questions from the DFG checklist on documentation and data quality**

* What approaches are being taken to describe the data in a comprehensible way (e.g., use of existing metadata or documentation standards or ontologies)?
* What measures are taken to ensure high quality of the data?
* Are quality controls in place, and if so, in what way?
* What digital methods and tools (e.g., software) are required to use the data?

Explanation:

In this section, describe how you proceed in order to describe the files created or re-used in the project, so that the collection and use can be traced. Provide information on the calibration of measuring instruments, test and reference measurements, plausibilities, etc. Describe the mechanisms used to ensure high quality data. Random data or misinterpretations should be prevented. Document which software was used for the evaluation. In addition to documentation, address how you will ensure the quality of the data and quality controls. In doing so, also indicate which methods or tools can be used to utilize the data.

Additional notes:

The data description must always include information on the origin of the data, measuring instruments used, day, time, location, persons performing the measurements, etc. This information should be documented in a measurement or data log. For standard software, it is sufficient to specify the name and version. If you use self-developed software for the evaluation, it should be made available for further use at the end of the project together with the data and the corresponding documentation. The question about the methods or tools (e.g. software) can also be answered in section 1.

Consider the following questions:

*Do you use existing metadata standards for descriptions (e.g., based on specifications in repositories)?*

*If applicable, do you need to document sensitive data in anonymized form? And if so, how?*

*What quality control mechanisms are used?*

## Storage and technical backup during the course of the project

**Original questions from the DFG checklist on storage and technical backup during the course of the project**

* In which way will the data be stored and backed up during the project duration?
* How is the security of sensitive data ensured during the lifespan of the project (access and usage management)?

Explanation:

In this section, describe where the data will be stored during the project and who will have access to it. Describe the storage devices and workflows used.

Additional notes:

Backups must be provided in any case. If necessary, differentiate between the various types of data. If sensitive data is involved in the project, please provide specific information on how to manage access and usage for this data.

## Legal obligations and framework

**Original questions from the DFG checklist on legal obligations and framework conditions**

* What legal particularities exist in connection with the handling of research data in your project?
* Are there any anticipated impacts or restrictions on subsequent publication or accessibility?
* In what way are aspects of use and copyright as well as ownership issues considered?
* Do important scientific codes of conduct or professional standards exist that should be considered?

Explanation:

In this section, describe the legal aspects of the data. Specifically address the possibility of subsequent publication or accessibility. Do important scientific codices or professional standards exist that should be taken into account? Who owns the rights of use (researchers? projects? employers?)?

Additional notes:

Legal specifics may exist in the case of:

- personal data,

- business data,

- data with copyright protection,

- patents,

- data whose publication may pose a threat to persons, institutions, groups or objects (e.g. precise location information in the case of endangered animal and plant species or valuable objects; naming of persons or groups in the case of religious or political endangerment),

- Data that have been provided to you for your research only under restrictive conditions.

In the case of industry cooperation, the question of the use and, if necessary, publication of the data must also be clarified. Agreements may need to be signed here.

You can obtain further information on data protection aspects here:

<https://www.ruhr-uni-bochum.de/en/datenschutz> and <https://dsb.ruhr-uni-bochum.de/en/>. Both sites are currently only available in German.

Please note that the ethics committee may have to approve the data collection: <https://www.ruhr-uni-bochum.de/ethik/index.html>.

Further information on legal and ethical frameworks can be found here: <https://wissenschaftliche-integritaet.de/en/code-of-conduct/legal-and-ethical-frameworks-usage-rights/>.

Information on scientific codices or subject-specific standards can be found here: <https://wissenschaftliche-integritaet.de/en/comments/further-links-relating-to-legal-and-ethical-frameworks/>.

## Data exchange and permanent accessibility of data

**Original questions of the DFG checklist on data exchange and permanent accessibility of data**

* Which data are particularly suitable for reuse in other contexts?
* Which criteria are used to select research data to make them available for reuse by others?
* Do you plan to archive your data in a suitable infrastructure?
* If so, how and where?
* Are there any embargo periods?
* When will the research data be made available for use by third parties?

Explanation:

In this section, describe whether there are plans to publish data during the project period or after the project has ended and according to which criteria the data will be selected. Please also describe whether and where the data will be archived and whether embargo periods must be taken into account in case third parties are allowed to access the data. If you do not plan to archive and provide any data at all, please give detailed reasons. This will always be considered a point of criticism by the reviewers and the DFG.

Please observe the FAIR Data principles: <https://www.go-fair.org/fair-principles/>.

Describe how you will implement or fulfill these principles throughout the project.

Additional notes:

Well prepared and documented data can be used for further research and verification of your research results. Preliminary results, results from test calculations or readily reproducible data, e.g. from model calculations, should not be designated for reuse. Check whether there is a secure, permanent, and approved storage facility (e.g., subject-specific repository) for your field. Not all requirements of the FAIR principles need to be met, and FAIR in no way means that all data must be publicly available.

## Responsibility and resources

**Original questions from the DFG checklist on responsibilities and resources**

* Who is responsible for adequate handling of the research data (description of roles and responsibilities within the project)?
* What resources (cost, time or otherwise) are required to implement adequate handling of research data in the project?
* Who will be responsible for curating the data after the project ends?

Explanation:
At this point, describe who is to assume which responsibilities within the project and after the end of the project with regard to data, and in which role
 Consider the following questions:

* Who is responsible for data generation and documentation?
* Who is responsible for preparing the data for publication/provision/reuse?
* Who is the contact person for the data after the end of the project?

In addition, provide a rough estimate of the resources (time, costs, etc.) required for this. Also consider the costs that may be incurred for processing and archiving the data.

Additional notes:

First of all, every individual researcher is responsible for his or her own data. In larger projects, however, it is also possible to assign one single person the responsibility for ensuring that the regulations for handling data are observed. In a proposal, it is a positive signal if such regulations and responsibilities are provided for. Take advantage of existing DFG funding opportunities; especially in large collaborative projects, it is also a good idea to create a position for IT and data management. Do not consider research data management as a minor issue that is done "somehow". The DFG and reviewers place a high priority on this topic. Data curation refers to the further maintenance of the data in terms of content, e.g., the addition or correction of data in a database, but also the permanent usability and interpretability, which can be achieved, for example, by migration to other data formats. As a rule, no or only few resources will be available for these tasks after completion of the project.