

## Handling of research data

Checklist for planning and description of handling of research data in research projects.

### Note

The DFG provides detailed [information on handling research data in DFG projects](#). In particular, the [checklist](#) provided here must be followed when submitting a proposal.

The checklist has been supplemented with comments and explanations (blue text parts) by the Centre for Sustainable Research Data Management at the Universität Hamburg. Please contact us if you have any questions. [www.fdm.uni-hamburg.de/en](http://www.fdm.uni-hamburg.de/en) (April 2022)

This checklist should assist you to describe key aspects of the handling of research data in a structured manner and to identify the resources and competence required for implementation. Please use section 2.4 of your application to comment on the topics mentioned below.

Research data includes measurement data, laboratory values, audiovisual information, texts, survey or observation data, methodological test procedures and questionnaires. Compilations, software and simulations can equally represent a central result of scientific research and are therefore also included under the term research data. Research data in some subject areas is based on the analysis of objects (such as tissue, material, rock, water and soil samples, test specimens, installations, artefacts and art objects), so its handling must be just as careful and consideration must be given to a technically adequate option for subsequent reuse whenever meaningful and possible. Should subsequent reuse of the resulting research data be closely associated with objects, then please also elaborate on this by providing all relevant information.

Please consider the existing standards in your discipline, any current subject-specific recommendations and any existing infrastructure services (such as data repositories, archives or collections). An overview of existing structures is available in the RIsources research infrastructure portal ([https://risources.dfg.de/home\\_en.html](https://risources.dfg.de/home_en.html)) and the re3data registry of research data repositories (<http://re3data.org>).

Further information on this topic and subject-specific recommendations are available via: [www.dfg.de/research\\_data/checklist](http://www.dfg.de/research_data/checklist)

## Checklist Regarding the Handling of Research Data

The DFG expects applications for funding to answer the questions from the following checklist. A clear, structured and transparent presentation helps in the assessment of the proposal and is a plus in the evaluation of the proposal. If a visit by a review committee takes place during the application process, it is highly likely that the topic of research data will also be addressed.

You can create a data management plan with the tool Research Data Management Organizer (RDMO) of the UHH <https://www.fdm.uni-hamburg.de/service/rdm.html> and attach it to your proposal.

It is essential that you clarify questions of data protection and ethics with the relevant authorities before submitting your application.

Do you plan to develop a database and/or web application in your project that will be published as a result of your project? For a positive evaluation of such projects, the DFG expects reliable statements on long-term operation beyond the project duration. The long-term operation of such an application requires ongoing adjustments to the software and is a complex process that cannot be carried out without human resources. It is essential that you consult us beforehand!

### 1. Data description

#### How does your project generate new data?

This means measurements, surveys, photos, films, analyses, calculations, etc. Describe as comprehensively as possible which methods you will use and what kind of data will be generated or used.

#### Is existing data reused?

If data are used that do not originate in your project, indicate the context from which the data originate, if possible as a citable source (e.g. via a DOI).

Otherwise, it is advisable to indicate that no suitable data are available after your own research.

The use of existing data, even if it is additional or serves as a comparative material, is in any case a positive aspect for your proposal.

#### Which data types (in terms of data formats like image data, text data or measurement data) arise in your project and in what way are they further processed?

Describe the entire process of data collection, from data origin, e.g. measurement, to use. It should be visible that the data is handled in a planned manner. Also describe how you document the individual steps. It can also be helpful to describe a system for naming files and folders.

### **To what extent do these arise or what is the anticipated data volume?**

On the one hand, give information on the number (e.g. of intended measurements, interviews to be conducted or images to be analysed) and on the other hand, on the expected amount of material per measurement. For example: "10 measurements are planned, each measurement will generate approx. 10 Mbyte of data, for the entire project this results in a data volume of approx. 100 Mbyte."

However, also indicate which data volumes are only needed for the work in the short term and which need to be stored in the long term.

## **2. Documentation and data quality**

### **What approaches are being taken to describe the data in a comprehensible manner (such as the use of available metadata, documentation standards or ontologies)?**

The data description should always include information on the origin of the data, measuring instruments used, day, time, location, persons carrying out the measurements, etc. This information should be documented in a measurement or data protocol.

Describe the data, indicate the units used. Someone who does not know your data must be able to understand and use the data on the basis of your description. Use only known abbreviations or create a list of abbreviations.

### **What measures are being adopted to ensure high data quality?**

Provide information on the calibration of measuring instruments, test and comparison measurements, plausibilities, etc.

### **Are quality controls in place and if so, how do they operate?**

Describe the mechanisms you use to ensure high quality data. Random data or misinterpretations should be prevented.

### **Which digital methods and tools (e.g. software) are required to use the data?**

Document which software was used for the evaluation. In the case of standard software, it is sufficient to state the name and version. If you use self-developed software for evaluation, it should be made available for further use at the end of the project together with the data and the corresponding documentation.

Measurement data that are only available in a specific format of the device manufacturer and can only be read and processed with this device should be described accordingly. Check the possibility of converting such data into non-proprietary formats. For example, digital images from cameras in the manufacturer-specific RAW format can always be saved as DNG, TIFF or JPEG. This increases the reusability!

## **3. Storage and technical archiving the project**

### **How is the data to be stored and archived throughout the project duration?**

Describe the storage media and processes used. In any case, backups must be provided. If you use the central network drives of the Regional Computer Centre of the UHH (RRZ, [www.rrz.uni-hamburg.de](http://www.rrz.uni-hamburg.de)), backups are created automatically. In the case of mobile data storage devices (USB hard drives, integrated memory in devices, memory cards, hard drives in laptops or PCs), you must make a backup copy as often as possible. To do this, you should only connect the mobile data storage devices that are used as backup media to the measuring device or laptop for the creation of the backup and also keep these data storage

devices secured in a separate location if possible.

If you are on a research trip, you should always create a daily backup of your data and store it separately, e.g. in a hotel safe.

If possible, use storage media from different manufacturers and production batches to be protected against possible production and serial errors in the storage media.

### **What is in place to secure sensitive data throughout the project duration (access and usage rights)?**

Describe which persons have access to the data and in what form, and what measures you take to prevent the data from falling into unauthorised hands. If, for example, you store sensitive data on a USB hard drive, then this hard drive should be protected (password, encryption if necessary) so that in the event of loss or theft, unauthorised use of the stored data is excluded.

For sensitive data, especially when using personal data, it is also advisable to include advice from the data protection office (<https://www.uni-hamburg.de/uhh/organisation/beauftragte/datenschutz.html>).

## **4. Legal obligations and conditions**

### **What are the legal specifics associated with the handling of research data in your project?**

Special legal features may exist in the case of

- personal data,
- business data,
- data with copyright protection,
- patents,
- data whose publication could pose a threat to persons, institutions, groups or objects (e.g. exact location information for endangered animal and plant species or valuable objects; naming of persons or groups in the case of religious or political endangerment),
- data provided to you for your research only under restrictive conditions.

Please note: Data acquired in an illegal form (e.g. data theft, use of data without consent, photos without permission) cannot be subsequently legalised and do not enable legal research work. In particular, your research results are not verifiable.

### **Do you anticipate any implications or restrictions regarding subsequent publication or accessibility?**

Depending on the conditions mentioned above, there may be restrictions. Describe these as transparently as possible.

### **What is in place to consider aspects of use and copyright law as well as ownership issues?**

This also depends on the previously mentioned conditions. If you use only your own data, you can choose a licence for further use that is as open as possible. If you use third-party data, you should, as far as possible, refer to the source of the data. If this data is already stored in

an accessible repository, for example, enter the URL or DOI; it is not necessary to store it again.

You may not simply publish data that are subject to copyright protection.

**Are there any significant research codes or professional standards to be taken into account?**

Research disciplines, research institutions and even larger projects have guidelines for handling data. Here you can refer to such guidelines and existing documents in your application.

**5. Data exchange and long-term data accessibility**

**Which data sets are especially suitable for use in other contexts?**

Well prepared and documented data can be used for further research and verification of your research results. Intermediate results, results from test calculations or simply reproducible data, e.g. from model calculations, should not be earmarked for subsequent use.

Follow the FAIR Data principles:

[https://www.forschungsdaten.org/index.php/FAIR\\_data\\_principles](https://www.forschungsdaten.org/index.php/FAIR_data_principles)

Describe how you will implement or fulfil these principles in the project. Not all requirements have to be fulfilled and FAIR does not mean that all data have to be freely available.

**Which criteria are used to select research data to make it available for subsequent use by others?**

Describe why data is selected by you. For example, it may make sense to provide the raw data and the edited final version of your data.

**Are you planning to archive your data in a suitable infrastructure? If so, how and where?**

Check whether there is a secure, long-term and accepted storage option for your research field. (e.g. GESIS in the social sciences, DKRZ for climate data, etc.). If such a repository exists and it accepts your data, then your data should be stored there in any case.

If there is no such repository or if your data is not suitable for such repositories, then use the Research Data Repository (FDR) of the UHH ([www.fdr.uni-hamburg.de](http://www.fdr.uni-hamburg.de)).

Describe in your application where you will store which data.

**If you do not plan to archive and make available any data at all, you must give detailed reasons for this and the reviewers and the DFG will always regard this as a critical point.**

The only exception is the use of data already stored and available elsewhere, but this can be well justified.

## **Are there any retention periods? When is the research data available for use by third parties?**

In many repositories, including the UHH FDR, you can set embargo periods. Your data does not have to be available immediately after the end of the project; you can also wait for further research or the publication dates of relevant publications.

The first priority is the potential availability and discoverability of the data. Make your research data available as open as possible and as closed as absolutely necessary. The argument "this is my data" without comprehensible reasons will lead to a negative evaluation of your application.

## **6. Responsibilities and resources**

### **Who is responsible for adequate handling of the research data (description of roles and responsibilities within the project)?**

In the first instance, of course, each individual researcher is responsible for his/her own data. In larger projects, however, it is also possible to have a central responsibility that ensures that the regulations for handling data are adhered to. In your application, it is a positive signal if such regulations and responsibilities are provided.

The reference to close cooperation with the FDM Centre (possibly also confirmed by a Letter of Intent (LOI)) also underpins the procedure.

In Collaborative Research Centres, Research Groups and Research Training Groups, resources should always be available for the training of graduate students. The FDM Centre can support you.

### **Which resources (costs; time or other) are required to implement adequate handling of research data within the project?**

Take advantage of the DFG's existing funding opportunities; especially in large collaborative projects, it is also a good idea to create a position for IT and data management. Do not regard research data management as a side issue that is dealt with "somehow". The DFG and the reviewers attach great importance to this topic.

Are there costs for storing the data? Normally, resources are available free of charge via the RRZ. However, particularly large needs may require a financial contribution; this should be discussed with the RRZ in advance.

Long-term storage of data in the UHH-FDR is free of charge. However, if data is stored with third-party providers, costs may be incurred. In particular, permanent costs after the project duration must be clarified in advance, as these costs are not usually covered by the UHH.

Permanent operating costs and expenses for the maintenance of websites and applications and databases are also not covered by UHH.

### **Who is responsible for curating the data once the project has ended?**

Curation of data means 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 e.g. by migration into other data formats.

As a rule, there will be little or no resources left for these tasks after the project has been completed.

A distinction must be made between two basic cases:

- **Databases and web applications**
  - After appropriate joint planning at the beginning of the project and handover of documentation and the necessary access codes at the end of the project to the FDM Centre, or when using the Heurist database instance offered by the FDM Centre, the operation of the application is ensured as far as technically possible. You may continue to maintain the contents of the database or the web application. However, the curation of the content will not be taken over by the FDM Centre.
- **Research data stored in the FDR of the UHH or another repository**
  - Operators of repositories do not change the stored data. There is no migration to new file formats; rather, physical preservation, the so-called bitstream preservation method, is applied. The permanent usability and interpretability of the data cannot therefore be guaranteed, as this can depend on the availability of the respective data formats or the necessary software. A curation of the contents, e.g. an error correction or addition, cannot take place. Corrected and changed data can only be saved as a new version of the data. This procedure also complies with the regulations for DOI allocation.