Catalog DFG checklist

Comment: This is the PDF version of the [DFG checklist](#) as RDMO catalog

Section Data description

Page Contentwise data description

Help: <u>General notes</u>: This catalog is the checklist published by the DFG, supplemented by help texts, option sets with suggested texts and further information. The DFG checklist must be taken into account when submitting an application and when completing the project and is intended to help you to structure your data management. If you have any questions, please contact the central research data management team at your institution. The following questions are intended to describe the datasets that are created and/or used in the project. The definition of a data set is an important conceptual decision that must be made individually for each undertaking or project. Choose a logical unit of your research data as the data set, for which the same information regarding methodology, technology, accessibility, etc. applies. The division into data sets helps to assess the value of the data with regard to potential subsequent use and later archiving. Since all questions have to be answered for each individual data set, an overly fine-grained structure does not make sense.

Question: How does your project generate new data?

Help: This information is relevant for all types of data in order to be able to understand their genesis. The way in which data are generated in your project depends on the disciplines involved - for example, it can be 3D modelling, the synthesis of a molecule, self-developed software, an online quantitative survey, etc. Please select the methodology used to generate the data.

Question: Are existing data reused?

Help: This refers to your own data from previous projects as well as to publicly or commercially available data. Re-using other scientists' data should also be mentioned here. If data are used that aren't generated by the project, please state the context from which the data originates, preferably as a citable source (e.g. via a DOI). Otherwise, it is advisable to indicate that no suitable data are available based on your own research.

Explanation: Some data can in principle be recreated at any time. Examples of this are scientific experiment data or digital copies of analog objects (as long as the originals are not lost). The effort and costs for this can of course be quite considerable. With regard to the question of the need for long-term archiving at a later date, the effort involved in creating a new document should be weighed against the effort involved in long-term archiving. Other data, on the other hand, cannot be collected again per se. This is the case, for example, with any kind of episodic observation, be it social science or natural science, as these depict a specific phenomenon at a specific point in time and/or place and are therefore usually not repeatable. Their value for subsequent use by others as well as the loss of long-term archiving that has not taken place or has failed is incomparably higher than that of reproducible data.

Page Technical Data Description

Question: Which data types in terms of data formats (e.g. image data, text data or measurement data) arise in your project?

> **Help:** For choosing data formats, you should also consider consequences for collaborative data use, archival and later re-use. It is recommended to use formats that are standardized, open and non-proprietary as well as common in general or in the specific community.

Question: In what way is the data processed in your project?

> **Help:** Please briefly describe how the data are evaluated, for example: mathematical calculations used, aggregations, preprocessing, references to other data sets, software used, etc.

Question: What is the anticipated data volume?

> **Help:** Please also state the unit of measurement, e.g. "10GB".

Section Documentation and data quality

Page Documentation

Question: What approaches are being taken to describe the data in a comprehensible manner (such as the use of existing standards for metadata or documentation, or ontologies)?

> **Help:** Also indicate the metadata standard (description system) if it is already clear which one shall be used, e.g. because the data repository you have chosen specifies one or you are bound to a subject-specific metadata standard. An example of a generic metadata schema for data is the [Datacite schema](#).

Question: What digital methods and tools (e.g. software) are required to use the data?

> **Help:** In order to be able to reuse data, in addition to the data itself, the software, if applicable (measuring) devices etc. and knowledge of special procedures for use are required. Since usage can vary depending on the discipline, either discuss with your colleagues or contact the [NFDI consortium](#) for your discipline if you don't know exactly what is needed.
>
> As for data formats: the more standardized, open and established methods and tools are, the easier it is usually to reuse data. The establishment of standards for methods, use of software, collection of research data and description of research results is an essential prerequisite for the comparability and transferability of research results.

Page Data Quality

Question: What measures are being adopted to ensure high data quality?

> **Help:** Depending on the discipline, the quality of the data can be ensured by various measures

Question: Are quality controls foreseen and if so, how do they operate?

> **Help:** Quality controls can be carried out in different ways. They can be heavily dependent on your discipline and your structures at the institute or in the project. It is important that the quality control is independent of the measures taken for quality assurance and refers to the product "data".

Section Storage and technical protection measures during the project

Page Data storage during the project

Question: How are data stored and secured during the project duration?

> **Help:** Please note that depending on the storage location, your data are not automatically backed up. You can find out advantages and disadvantages of storage media at www.forschungsdaten.info. The DFG expects here a description of the infrastructure you use for data storage and backup. In particular, describe the use of storage and backup infrastructure offered centrally at your institution or justify your use of local or (subject)-specific solutions.

Page Technical data archiving during the project

Question: What is in place to address security of sensitive data during the project (access and use controls)?

> **Help:** Sensitive data can appear in different contexts, e.g. personal data or data that arise in economic cooperation projects or licensed data. If you have any questions about personal data and its security, please contact your institution's local data protection officer. In the case of economic cooperation projects, the access and usage rights are usually recorded in the non-disclosure agreements, while the usage conditions for data are recorded in license agreements.

Section Legal obligations and conditions

Page Publication restrictions

Question: Do you anticipate any implications or restrictions regarding subsequent publication or accessibility?

> **Help:** There are reasons that preclude a (direct) publication of the data. These can be, for example, the protection of personal data or the commercial or industrial use of the data. Please keep in mind that it may be possible to publish the data after a certain blocking or embargo period and that you can also prepare this publication now. If you have any questions, speak to your institution's research data management team.

Page Scientific specifics

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

> **Help:** In addition to general scientific codes, there can also be discipline-specific codes. If you do not know these, find out more from suitable NFDI consortia or professional associations. The DFG page on academic integrity also provides information. Here you will find references to laws, subject-specific comments on the guidelines and explanations as well as case studies.

Page Legal Restrictions

Question: How are aspects of use and copyright law as well as ownership issues considered?

> **Help:** The various legal aspects must be taken into account for each data set. Whether your data are subject to copyright protection depends, among other things, on the threshold of originality. Information on this can be found e.g. in the *Information sheet on copyright protection of research data* under *Copyright and licensing for research data* (https://zenodo.org/record/5243232).

Question: Which further legal specifics apply to the handling of research data in your project?

> **Help:** For example: Does the legal situation in different countries have to be taken into account? Does every employee have the same access rights?

Section Data exchange and long-term data accessibility

Page Reuse

Question: Are these data suitable for subsequent use in other contexts?

> **Help:** Please consider whether your data can be reused. The reusability of data does not mean its publication, but e.g. the value of the data for follow-up projects, new collaborations, etc. In the evaluation, it can also play a role whether data can be generated again. This is the case for digital versions of analogue objects (if the originals are still available), generally but often with great effort also for experiments in natural sciences. Such data do not necessarily have to be preserved. On the other hand, any time-related data - time series, episodic events such as earthquakes, also social science surveys - are not repeatable unique observations.

Question: Which criteria are used to select research data to make them available for subsequent use by others?

Question: When are the research data available for use by third parties?

> **Help:** Accordingly to the DFG Guidelines for Research Data Management, research data should be made available as soon as possible and ideally according to the FAIR principles. The importance of data publication is underscored once again by the Guideline 13 of the Code of Good Scientific Practice. Discipline-specific practices are pointed out here.

> **Recommended repositories**

> For the publication of research data, it is recommended to use subject-specific repositories. Recommended directories are re3data.org and FAIRsharing.org. If no suitable subject-specific repository exists, inquire whether your facility operates an institutional repository. Another alternative would be Zenodo.

Page Archival

Question: Are you planning to archive your data in a suitable infrastructure? How and where?

> **Help:** In accordance with good scientific practice, research data must generally be stored for 10 years (see the Guideline 17 of the Code of Good Scientific Practice). In order for the data to remain available and findable, they must also be curated beyond the project. If you archive the data in a data center or repository, the responsibility is usually assumed by the operator. Please check or clarify the responsibilities in advance. Archiving includes metadata, documentation and any relevant code/software.

Section Responsibilities and resources

Page Responsibilities

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

> **Help:** We recommend specifying the persons in the following form: title; first name, last Name; ORCID ID (if existing); professional position; institute; contact details; role in the project.

Question: Who is responsible for curating the data after the end of the project?

> **Help:** We recommend specifying the persons in the following form: title; first name, last Name; ORCID ID (if existing); professional position; institute; contact details; role in the project.

Page Resources

Question: Which resources (costs, time or other) are required to implement adequate handling of research data within the project?

> **Help:** This can involve technical or IT resources as well as expertise that is brought in, for example, by data manager or IT experts. If necessary, provide more detailed information on the respective items.