

Research Data Management Plan templates – instructions for use

The NFDI4Health offers three Research Data Management Plan (RDMPs) templates: One for epidemiological studies, one for clinical studies, and a short version for handing in with proposals.

A (research) data management plan is a useful tool for the planning phase of any research project but can and should be used throughout the duration of the whole project as a means of documentation. The plan describes the types of data researchers intend to collect or generate during their research process. It also states how the data will be used and stored during their lifecycle.

A DMP serves to answer questions such as the following:

- Description of the research project
- Type of data involved
- How the data will be organised
- Data usage
- Metadata and referenceability
- Storage and security
- Publication
- Legal aspects and anonymisation
- Digital long-term preservation

The three templates each contain a list of questions and discipline-specific answer options. You can decide and choose which questions apply to your research project and select which of the given answers is most accurate. Depending on your needs select one of the following templates:

- For an epidemiological study, please select the RDMP- epidemiological template.
- For a clinical study, please select the RDMP – clinical template.
- For a short RDMP for a proposal, please select the RDMP short – template.

Please note that a DMP/RDMP is a living document, i.e. some of the questions cannot be answered at the start of a project but only within the course of the research process, e.g. **1.5** *What is the project duration?*, **1.8** *Who is funding the project?*

Other questions such as those on the (expected) data volume (**3. Data Formats**) can only be answered with estimated values, as the actual volume cannot be known at the beginning of a research project. Other answers may also change in the course of time and should be adjusted accordingly to ensure that the RDMP remains up-to-date and accurate.

Below you find some definitions of terms that are used in the templates to avoid confusion:

Metadata/ Study data:

Metadata are data providing information about other data. Metadata are not the content of the data themselves, e.g. not an image or single values of a laboratory measurement.

A few examples of metadata are:

- Title, abstract, author, and keywords.
- Types, versions, relationships, and other characteristics of digital materials
- Resource type, permissions, and when and how a resource was created
- Process data, description of processes that collect, process, or produce statistical data
- Creator, copyright holder, and public licensing

In the given context, metadata are study data describing a study.

Data:

The actual content of the data itself, i.e. any data that is collected or generated during the research project or any data that is a result of the research project, such as images, values of laboratory measurements, urine samples, etc.

Raw data:

The unedited/unprocessed collected/generated data.

Cleaned data:

Data that has been edited and cleaned of incomplete, incorrect, inaccurate or irrelevant parts.

Processed data:

Data that has been structured and organized in a way to enable a following analysis.

Analysed data:

Data that has been inspected, cleansed, transformed, and modelled with the intent of discovering useful information.

Published data:

Data that has been published and made accessible as data sets, e.g., in repositories.

Store/d:

Data is stored/saved during the project duration, for a short or long time.

Archive/d:

Data is submitted to an archive; this concerns long-term archiving and preservation to make data available for the long term.

Generate:

Generating data describes the process of data collection. The researcher aims to produce rich and meaningful data for further analysis by activities such as searching for, focusing on, noting, selecting, extracting and capturing data.

Analyse:

Analysing data describes the process of inspecting, cleansing, transforming, and modelling data with the intent of discovering useful information.

Data reuse/ reused data

“Reused” means using data for other purposes than it was originally collected for. The data can originate from a previous study of yourself or from another source. This is also referred to as *secondary use*.

Metadata standard

A metadata standard is a document which establishes a common way of structuring and understanding data and includes principles and implementation issues for utilizing the standard. These can be generic (e.g., DataCite, Dublin Core) as well as subject specific.

Mapping

To map metadata to an existing standard, you assign information from the standard to your metadata.

Persistent identifier (PID):

A long-lasting reference to a document, file, web page, or other object. Examples of persistent identifiers are DOI (Digital Object Identifier), ORCID (Open Researcher and Contributor ID), and ROR (Research Organization Registry ID).

Terms of use:

A legal agreement between the creator of the data and the data end-user, e.g., an individual data usage agreement that states what data may be used under what conditions and for what purposes by the user who requested to use the data.

License:

A legal arrangement between the creator of the data and the data end-user(s), not an individualised agreement. The purposes for which the data may be used are defined in the corresponding licence that the data are minted with. An example of licenses is Creative Common Licenses.

If you have any questions or should you need support in filling in the template, please contact: contact@nfdi4health.de

DOI: 10.4126/FRL01-006483992

Version: V1_0

Release Date: 22.08.2024

Lizenz

Dieses Werk wurde unter der Lizenz „*Creative Commons Namensnennung 4.0 International*“ (CC BY 4.0) veröffentlicht. Den rechtsverbindlichen Lizenzvertrag finden Sie unter <https://creativecommons.org/licenses/by/4.0/legalcode>.

