EUR Data Management Plan - Guidelines

Version 4.3

Table of content

The data management plan format can be found in a separate document; questions in the dmp format marked with * refer to the guidelines.

General

Ideally, each research project should start with a Data Management Plan [DMP]. In a DMP you write what type of data and how much data you will be collecting, what tools you will be using, who has access to the data, how backups are made, etc. Thinking about what is needed in each stage of the research project before you start the project will help make the project run more efficiently and prevents issues from occurring. Writing a DMP is part of good research practice and most research funders have made writing a DMP mandatory.
You can find all funder-specific formats in DMP Online. In case there is no funder or funder-specific format, you can use the EUR DMP Format [this format], which can also be found in DMP Online. This document contains the guidelines belonging with the EUR DMP Template.

Administration & Project Description

This section of the DMP is meant to provide an overview of who is involved in the project, the research topic, research question(s) and methodology, and track the version of the DMP. This information serves as a reference for interpreting choices made in the subsequent sections of the DMP.

- A DMP is meant as a living document that is reviewed and updated about once a year. To keep track of these updates, the question **Date of update DMP and version** was added. If the current version is version one, this question may be skipped.
- An ORCID is an author identifier designed to aid the researcher in distinguishing his/her work from other researchers with a similar name, as well as making it more easy for interested parties to keep track of the work of a specific author. Some journal require all authors or specific authors [e.g. the corresponding author] to have an ORCID before accepting a paper for publishing.

Preparation: Legal arrangements and policy

Legal arrangements and policy are the backbone of research: without a legal base, research involving research participants would not be possible and entering a partnership without a contract is unwise.

- Legal arrangements can be versy specific and initiated by the research team such as arrangements regarding data ownership in case of multiple parties involved, intellectual property rights, or an Informed Consent. In contrast, some legal arrangements are not initiated by the research team but must be upheld nonetheless. Examples of these are funder-specific requirements, Terms of Service from e.g. social media channels, or university or discipline-specific policies relating to e.g. the use of IT facilities or requirements for publication of data associated with published papers. In case legal arrangements need to be made, it is recommended to consult the legal officer of the faculty associated with the research project.
- Funding bodies also have specific policy on research data management which you will need to uphold.
    - ZonMW policy on research data management
    - NWO policy on research data management

- Ethical approval is required in case research participants are involved in the study to ensure what is asked of research participants does not infringe on participant health, safety, or wellbeing, in any way. Each faculty has its own ethical board and review process.

During research: Collecting and analyzing

Where the first two sections mainly concern the framework for conducting research in, this section focuses on the content of the research and aims to help assess whether all tools and protocols needed are in place.

- File format and size can differ significantly and may be reason to choose alternative tools. Video files, for example, can be several GB each, implying that standard storage capacity might not be enough. Here you can find a tool for calculating video file size and information on preferred file format. For this question it is sufficient to estimate the total file size [e.g. <1GB; 1-5GB; 5-10GB; 10-15GB].
- Many research projects involve research participants, and when research participants are involved, many times [sensitive] personal data is collected. Personal data is a piece of information or combination of pieces of information that may lead to the direct or indirect identification of the research participant. Examples are name, address, student number, and IP-address. Sensitive personal data is a special category of personal data and consists of data that may cause harm to the participant were it to made public. Examples are health status, religion, political preference, and sexual orientatation. If you collect (sensitive) personal data, privacy law [GDPR] applies, and it is recommended to consult the faculty privacy officer to determine what, if any, measures are needed to safeguard privacy.
- Sometimes it is necessary to share a datafile during the research project e.g. when working with a third party that needs to transfer the collected data to the EUR for analysis. In case this contains [sensitive] personal data, the data must at least be either pseudonymized or anonymized to protect the participants's privacy. If you need to transfer [sensitive] personal data, it is recommended to consult the faculty privacy officer for more information on how to do this and to determine whether additional measures are required such as sending the files encrypted. These measures depend on the degree of sensitivity of the data.
    - Pseudonymous means that the participant can only be indirectly identified, when, for example there is only a random participant ID in the datafile and a separate file [the key-file] contains the random ID and the personal data such as participant name.
    - Anonymous means that no-one can directly or indirectly identify the research participant. This includes the research team. There is no key-file. In practice it is nearly impossible to truly anonymize a dataset.
- Several tools are available to store data during the research project. Which one you need depends on the type of data you have and personal preference. For example, SURF drive and SURF research drive may be used to store and share personal data, but not sensitive [personal] data, and Mediasite is best suited for video files. It is recommended to use EUR supported tools as almost all are cloud-based

[thus accessible from anywhere], safe, and regular back-ups are automatically made. However, sometimes these tools do not meet all needs, resulting in private tooling to be used. If this is the case it is recommended to consult the faculty privacy officer to determine whether measures need to be taken to ensure safety and privacy such as using a tool with encryption or setting passwords.

- Part of good research data management practice is ensuring access to files is managed, files are organized and named in an easy to understand way [preferably standardized according to reseach discipline], regular backups are made, and versions of files are tracked and named accordingly. This will help the project become more efficient and transparent as clear agreements have been made / protocols are available. It will also help with future archiving of the data or depositing data underlying publications in a data repository.

After research: Data sharing and re-use

The EUR requires data underlying publications being made available for verification, replication, or re-use in similar research projects. This principle is becoming the standard for many research funders, journals, and even research domains. Good organization of research data is essential in getting data ready for sharing.

- To be able to re-use data or verify results, the data files must be accompanied by **documentation** describing e.g. timeframe of collection, collection methods, variable names or codebook, keywords, or other 'need to know' information. This is called meta-data: data describing data, and most often they are written in a separate readme.txt file accompanying the data. For some research types or research domains, there are so-called **meta-data standards** available that describe a uniform way to assign metadata [for examples; see FAIRsharing or DCC listing], thereby contributing to interoperability, but these can get very technical very quickly.
- There are many different types of data repositories. There are university-, funder-, and journal specific repositories, but also domain- or data type [e.g. code or MRI-scans] specific repositories, all of which can either be free to use or require payment. The EUR also has its own data repository which can be used to deposit data [in all formats] underlying publications. To help make this data more interoperable and reusable, a data curator checks all files and makes suggestions for e.g. file format and metadata description. If possible, all files underlying a publication should be placed in the EUR Data Repository; if not possible, then the metadata should be there. Please note that for each file restrictions may be set such as making it confidential. In case of confidential files, the terms under which data may be accessed and/or reused must be present. These and other features allow the majority of datasets to be placed in one repository. If a domain- or data type specific repository is more suited for your files, make sure you enter the metadata and the digital object identifier [DOI; the 'digital address' of your files] in the EUR Data Repository.
- When you share data, you want to make sure that others who re-use it give you the credit you deserve and only use it for its intended purpose [e.g. yes or no to commercial use, share alike]. There are many different types of licenses, but most common are the Creative Commons [i.e. CC-license]. The standard license in the EUR Data Repository is CC-BY, which lets others distribute, remix, adapt,

and build upon your work, even commercially, as long as they credit you for the original creation. This is the most accommodating of licenses offered and is recommended for maximum dissemination and use of licensed materials. If you want, you can choose a different CC license or a different license all together [e.g. a software license for code].

For ZonMW and NWO projects, the following data sharing and re-use requirements apply:

As much as possible, research data should be made publicly available for re-use. As a minimum, NWO requires that the data underpinning research papers should be made available to other researchers at the time of the article's publication and be assigned a persistent identifier, unless there are valid reasons not to do so. In this case, metadata must be made available. The guiding principle here is 'as open as possible, as closed as necessary.' Due consideration is given to aspects such as privacy, public security, ethical limitations, property rights and commercial interests. In relation to research data, NWO recognizes that software (algorithms, scripts and code developed by researchers in the course of their work) may be necessary to access and interpret data. In such cases, the data management plan will be expected to address how information about such items will be made available.

After research: Archiving

Technically, archiving is a process of depositing data somewhere. Here, it specifically pertains to storing the project data for the long-term [5, 10, 15 years] after the project has finished, as is often required by a research funder.

- The EUR required archiving term is 10 years.
- Files selected for archiving are typically encrypted before storing, are not accessed unless in case of emergency [or after the archiving term has expired], and contain files of the entire project and not just files associated with a published paper [these are deposited in a data repository]. Files that you archive can be e.g. raw data files, informed consents, contracts or agreements that were signed, the data management plan, ethics forms, and other important documents. Thus, this is a different type of archiving compared to depositing data in a repository for re-use.
- Currently, the EUR is working on a pilot for a data archive that is available for use; similarly there may be faculty specific archives or archiving protocols available.
- In case there are research partners from outside the EUR, data may be deposited in their respective archives. Please then list the name and a link to the archive in the data management plan.

For ZonMW and NWO projects, the following archiving requirement applies:

NWO expects you to preserve the data resulting from your project for at least ten years, unless legal provisions or discipline-specific guidelines dictate otherwise.