

Tips for writing a Data Management Plan (DMP) for researchers of Wageningen University & Research

This document explains a set of topics commonly covered by funders' Data Management Plan templates, and tips on how to address these. The templates of [Horizon 2020](#), [NWO](#) and [ZonMw](#) were used to create this overview.

If you would like feedback on your own (draft) Data Management Plan, please contact Data Management Support at data@wur.nl.

1. Data description

When you are asked to describe the data you will collect, it is good to be as concrete as possible. It may help to think of these three types of data:

- raw data (e.g. audio file of an interview, field measurements, experiment data from an instrument)
- processed data (e.g. transcribed and anonymised interviews, digitised field documents, cleaned experiment data in e.g. SPSS)
- analysed data (e.g. coded interview transcriptions, tables/figures of the analysed data)

In addition to the data, do not forget to describe what documentation you plan to keep [see also: **6**].

Some DMP templates also ask for an estimation of the expected data size (i.e. how many MBs, GBs, TBs). You can use existing files that you or your peers already have to come up with an estimate.

2. File formats

Some DMPs ask for the file formats you plan to use (e.g. TIFF, STATA, CSV). In this case, try to go for open file formats, that is, those for which no proprietary software is required (e.g. CSV rather than XLS (Excel)). This makes it easier for others to access and reuse your data. If possible, try to either:

- use open file formats throughout the research process, or
- use a closed format during the research process and convert it to an open format when archiving

Data archives often tell you which formats they prefer – these are usually open formats. You can use these lists of the archives [DANS-EASY](#) and [4TU Centre for Research Data](#) to find open file formats.

If your research requires specialised software with closed file formats, and no open format alternative exists for this, it is advisable to add to your DMP that the closed format is sufficiently standardised within your discipline for your peers to access it (if this is indeed the case).

3. Data storage during the research

In addressing data storage, try to:

- be as exact as possible (e.g. mention the use of SURFdrive rather than of 'a cloud service')
- follow [WUR's policy on data storage](#), and also mention in your DMP that you will stick to this policy
- use [institutional storage facilities](#), and also mention in your DMP that you will use these.

At WUR, safe storage during the research would be the W-drive, possibly in combination with IT-approved cloud storage such as Onedrive for Business (see [this page](#) for details). It is worth adding in your DMP that data on the W-drive is backed up automatically and stored at two separate data centres. Avoid the use of external storage such as hard drives and USBs – not only do these not follow WUR's policy, funders are also aware of the risks they entail and are unlikely to approve of their use.

4. Long-term archiving in a data repository / archive

Research data underlying publications [must be kept for at least 10 years](#). Data repositories are there to securely store the data for the long term and to make the data accessible.

WUR Library is 'front office' for the repositories [DANS-EASY](#) and [4TU Centre for Research Data](#). This means that the Library can help you in preparing and depositing your data set with these archives, as well as cover (part of) the archiving costs that may come up. DANS-EASY traditionally focused mostly on the social sciences and humanities but also contains life sciences data. 4TU focuses on the technical sciences. Both repositories are certified. If you choose one of these two repositories, it is advisable to mention in your DMP that WUR Library is front-office for it, and that you will get the Library's support in the data set preparation and submission.

You can also look for another (discipline-specific) repository. Perhaps the journal you will likely publish in recommends a certain repository. You can also search [Re3Data](#) for repositories.

Try to choose a repository that:

- assigns persistent identifiers to datasets [see also: **8**]
- is certified (check this [here](#))
- is commonly used in your discipline (especially if it does not meet the two points above). This allows for your peers to easily find the data, and likely means the repository meets discipline-specific needs on supported file formats and metadata.

If your chosen repository meets the above points, do mention this in your DMP.

When choosing a repository, also consider matters such as costs, the licences the repository lets you choose from when publishing your data, and the possibility to place an embargo on your data in case you need it. On [Re3Data](#) you can filter repositories by a set of criteria.

Our website gives more information on [data repositories](#) and [data publishing](#).

5. Selection of relevant data for long-term archiving

When listing what data is relevant for reuse and/or (public) archiving, think of:

- your publications: if the data has been used in publications, you have to archive it for 10+ years. This includes all the data needed to fully understand your research, redo it and verify your results. This also means your data documentation [see also: **6**].
- the time and costs involved: if the data collection/production was costly in terms of time or money, it is well worth keeping the data.
- the size of the data: if (raw) data files are extremely big, archiving all of it might be technically challenging or too costly.

Also note:

- If you work with third parties, your contract may stipulate what data you can and cannot share.
- When working with personal data, carefully consider what you should anonymise and can publish.

6. Data documentation

Data documentation is any contextual information that you add to your data set. You should document all information that others would need to make sense of your data, so that they could reuse it and reproduce/validate your study.

When asked about data documentation, explain: 1) what information you will document, and 2) how.

1) You can document information on two levels:

- the study level: e.g. roles and responsibilities, methodology, relationships between data files
- the data level: e.g. meanings of codes, reasons for missing values, variable definitions

2) You can document your data in different ways. Think of using:

- README files: these often list the files in a data set, what these files contain, and how they relate.
- a data dictionary: a list of variables and their meaning. This can be part of the README file.
- separate text files that give relevant information on e.g. methodology
- exports from a (lab) notebook

7. Metadata standards

A metadata standard is a structured way of describing data. It exists on two levels:

1) at the level of the whole data set: think of the creator of the data set, geographical location covered, and key words. This makes the data set easily findable and globally understandable.

2) at the level of the files within the data set: think of standard variable names and units of measurement used. This allows for the combining of the data set with other data sets using the same standard.

The first type of metadata is added to a data set when it is deposited to a repository. Repositories often use an existing standard. DANS-EASY and 4TU, for example, both use [Dublin Core](#). You can find out which standard a repository uses by checking their website or searching for the repository on [Re3Data](#) or [Fairsharing](#). Add the metadata standard used by the repository to your DMP. At times, a repository does not use an existing standard, but uses their own descriptors, often tailored to the type of data they host. In this case, you can mention this.

The second type of metadata is not always added to data sets, but is useful if you want your data set to be linkable to others and/or queried/aggregated. To find such metadata standards, browse [Fairsharing](#). Some repositories also allow you to add this type of metadata to your data (e.g. in [GBIF](#) you can easily apply the [Ecological Metadata Language](#) (EML) standard).

8. Persistent identifiers (e.g. DOIs)

Many DMP templates ask you if your data set will get a persistent identifier (often a [DOI](#), but [other forms exist](#)). A persistent identifier makes sure that the data is always accessible, even if the online location of the dataset changes. This makes it more durable than a traditional URL. Many data repositories assign persistent identifiers to the data sets they host. You can look at data sets on their website to see if this is the case, or look up the repository in [Re3Data](#) (once you have found the repository, look for a little blue icon right above its description). On this same website, you can also look for repositories that have persistent identifiers by using the filter 'PID systems'. Mention in your DMP that the repository will assign a persistent identifier, and which form.

9. Data management costs

With most funders, you can reserve budget for your data management costs. Visit [this webpage](#) for more details.

To get an idea of the costs involved in data management, you can use this [comprehensive guide](#) covering the full research cycle. You can also go through this [checklist](#), which focuses primarily on the costs involved in (preparing for) publishing data.

Costs of FB-IT storage at WUR (e.g. the W-drive) can be found in the IT Service Catalogue available on [this Intranet page](#). For the long-term archiving of your data, repositories might charge a fee. Wageningen is front-office for [DANS-EASY](#) and [4TU Centre for Research Data](#) and this means that the Library will in most cases cover the costs of archiving data sets with these two repositories.

10. The WUR data policy

WUR has a [data policy](#) that covers Data Management Plans, the storage of data, data archiving, and dataset registration. These policy regulations establish good data management practices. We therefore advise that you refer to the institutional policy in your DMP, and explicitly state that you will be following it throughout the research project.

11. FAIR data

Some DMP templates explicitly refer to the importance of data being 'FAIR', while others do not mention the term. But in almost all cases, the [FAIR principles](#) underlie the questions in a DMP. While filling in the DMP, it is therefore advisable to keep aiming for FAIRness in your answers.

FAIR research data is:

- Findable: others can find your data (it is in a repository, with metadata and a persistent identifier [see also: **4**, **7**, and **8**])
- Accessible: others can access (part of) your data set, if issues such as privacy do not hinder this
- Interoperable: people and machines can open the files [see also: **2**] and can combine this data set with other data sets through common (metadata) standards [see also: **7**]
- Reusable: the above three, plus: others can understand the data and know how they can reuse it (i.e. the data is documented [see also: **6**], [licensed](#), and its provenance and references are clear)

12. Working with personal data and complying with the GDPR (Dutch: AVG)

A DMP template might ask you how you plan to handle personal data. [This Intranet page](#) gives a lot of information and practical tips on this topic, also in relation to the GDPR (European legislation on working with personal data).

This document was last updated on September 9th 2018.

Questions or comments on this document? Please contact data@wur.nl.