



Tips & tricks when writing your DMP: general University Leiden template

Refers to the Data Management Plan (DMP) of Leiden University. Latest published version of the DMP template found at <https://www.library.universiteitleiden.nl/research-and-publishing/datamanagement/training-datamanagement> and <https://doi.org/10.5281/zenodo.3903267>

A link to the Leiden University regulations on data management found at: <https://www.library.universiteitleiden.nl/researchers/data-management>

Leiden Research Support Portal can help you find proper contact persons for research related questions: <https://www.researchsupport.universiteitleiden.nl/en>

0: Cover information

In this section you are asked to provide administrative information about your project and the project data. When your project is part of a larger or externally funded projects, check with the project leader/PI/supervisor to see if a project based DMP already exists so that you can make your version conform with it.

- The DMP should stand alone so that someone who knows nothing about your project can still understand what you are doing and why, and also how to trace the project data.
- Leiden University recommends researchers to create a ORCID to make sure all your research output can be connected to you. More information on ORCID: <https://www.library.universiteitleiden.nl/researchers/orcid>.
- Ethics issues can arise in many areas of research. The ethics review can precede the writing of a DMP but the DMP can also be one of the documents to submit for the review. If in doubt about the necessity of an ethical review, contact the appropriate ethics committee for your faculty (via the contact persons mentioned in the Research Support Portal).
- If you are creating or using any data that can identify a person, you must fill out the data processing register for research (sometimes called the DPIA) and identify the relevant risks and corrective measures. If need is, you can contact the privacy officer of your faculty. See the relevant pages on privacy and data protection at the staff members website: <https://www.staff.universiteitleiden.nl/ict/privacy-and-data-protection/general-data-protection-regulation-gdpr/data-processing-register> (after login)

Afterwards, you can tick the appropriate box on the DMP and keep a copy of your DPIA with your DMP.

- Research funders (such as NWO or ZonMW) require that you consult data management support staff when drafting your DMP. You can consult with support staff at your faculty or at the Centre for Digital Scholarship. For contact details, check the [Research Support Portal](#).
- The DMP is a living document and should be updated as your project changes and evolves. Use the version boxes to explain changes you make, for example, “Small changes made to correct small typos”, or “New data source from earlier project will be incorporated”. When you produce a publication package at the end of the project, you could update your DMP to reflect the decisions you made.

1: Data collection and creation

In this section you will describe your research data, whether it is already existing data collected by others or data that you will create yourself. It is important to know where your data comes from and it enables you to take the appropriate steps for the management and for the future use of the data. Remember people reading this plan have not necessarily read your research proposal.

1.1-2 Will the project use existing or third-party data?

- Will you use datasets that have been created by somebody else (third-party data or data from previous research)?
If so, describe what data was created, when, and by whom; what kinds of files does the dataset comprise; how you will get access to the data? This gives the ‘provenance’ to the data.
- Is there a license or user agreement that describes how you or others can access or use the data in the future?
- Do you need to cite the data in a particular way?

1.3-6 What type(s) of data will you collect or create, and how will you do that? In what file formats?

- Will you create your own data? If so, what is it that you do to create your dataset? This is where you describe your own research activities: will you make (speech) recordings, take pictures, set up a database, perform interviews, use GPS recorders?
- What type of files will you create: audio or video files, spreadsheets, images, texts, survey results, etc? Are there any options for the file type you could use (for example, .xls or .csv).
- Also detail here what files you will create when processing your data, and what files you will produce to archive your data.
- Tip: the following web page gives suggestions on the best formats for archiving your data after the project, but these may not be the best to use during your research when you need to work on the data:

- Social Science / Humanities (DANS/Dataverse):
<https://dans.knaw.nl/en/about/services/easy/information-about-depositing-data/before-depositing/file-formats>
- Science (4TU): <https://data.4tu.nl/info/en/use/publish-cite/upload-your-data-in-our-data-repository>
- Will you follow a methodology documented by others?
- Tools might include online tools or software, and physical equipment, machines, please see also the ICT and Research pages on the staff website: <https://www.staff.universiteitleiden.nl/ict/it-and-research>. For contact points, consult the Research Support Portal
Your Faculty Information Officers, <https://www.staff.universiteitleiden.nl/ict/help-and-support/helpdesks-and-contact/information-managers> are happy to help you out with questions on tools and software (for instance for recording interviews)
- How much space are you going to need to store your data **during and after** your research? This is important to make sure that you have enough storage space available, but it is not always easy. If you do not know, can you make an estimation, for example, 10 Megabyte versus 100 Terabyte? Sample files can help to calculate the total size. Or use file size calculator tools for your file formats on the web. To give an example: for videorecordings of interviews, estimate how many people you will interview, for how many hours. Depending on the quality of the recording, one hour of videorecording can take up to 1 gb.
If you expect to have costs for storing these data, mention these in Section 6.
- Tip: if you are gathering data without access to the university facilities, it can be useful to think about any limitations on your equipment such as storage capacity, encryption possibilities, risks to damage or loss.

2: Data storage and security

In this section you will describe how your data is stored, backed-up, and accessed during your project:

Where will your data be kept? If you have ethical or commercial issues, what do you do to make sure that the right people can access your data, and the wrong people can't? Do you have an estimate of the size of your data and do you have enough space for it?

- Consider digital as well as non-digital data.
- You can reduce the sensitivity of the data: anonymise/pseudonymise, aggregate
- Think about use of passwords, encryption, firewalls, etc. and make sure the transport of data and the deletion of data is secure.
- Your Faculty Privacy Officer can provide more assistance if you are dealing with personal data, data restriction, security issues, or even a data breach.
<https://www.staff.universiteitleiden.nl/ict/privacy-and-data-protection/personal-data/privacy-and-data-breaches/service-units/leiden-university-libraries?cf=service-units&cd=leiden-university-libraries>

2.1-3 Risks and measures

- Tip: you can find more information about working securely online, using mobile devices, and secure printing, on these web pages:
<https://www.staff.universiteitleiden.nl/ict/privacy-and-data-protection>
- If you are handling personal data, or data that should otherwise be protected, you must think about measures to mitigate risks; if not, then most of these measures are not applicable.
- Risks might vary per research phase – data collection, data processing, data analysis, data sharing – or the location where you are working.
- Copyright or intellectual property can be relevant if you are using someone else’s data.

2.4-6 Storage and back-ups

- Procedures should depend on the circumstances, the value of the data, and the level of risk (see 2.1-3) considered appropriate.
- If you store your data on the university network, a back-up will be made every night.
- If you are collecting or storing data outside Leiden University, then you will need to think about how you store, back-up, or transport the data securely:
 - Will you have access to the internet? If not when will you be able to upload your files and in the meantime, do you need to think about a temporary backup?
 - If, as a last resort, you must store files on mobile storage, is there a password or encryption to prevent unauthorized access?
 - If you do have internet access, do you know how to transfer files securely? Leiden University offers EduVPN for secure access while using wifi and SURFfile sender for secure transfer of files.
- Tip: you can use Surfdrive, the safe university cloud service, to store/access/share your files securely, but be aware that you need to take extra measures (see 2.1-3) if dealing with highly sensitive data.
Surfdrive does not guarantee a backup, although they may be able to retrieve lost files in an emergency. Always transfer or copy your files to the university network regularly/ as soon as you return to Leiden.
Research groups may use Research Drive, which is similar to Surfdrive but for larger volumes of data.
More information about sharing and sending files:
<https://www.staff.universiteitleiden.nl/ict/ict-workplace/software-and-collaboration-tools/sharing-and-sending-documents/>

3: Documentation and metadata

In this section you will write what you do to describe, document and to structure your data to help others understand what you have done. Think about file names and folder structure, software needed to read your data, information about where your data originate and how you collected it. You may record all this in a README file that is stored with your data, but most important is that you follow the way of working that is standard in your discipline.

3.1-2 How will you name and structure files, folders and different versions?

- The important thing is to think in advance about your file naming and folder organization so that you can be consistent when creating new files and folders, and you can locate your own data easily, saving yourself time in the long-run.
- Consider how to differentiate the different stages of your data, from raw data to processed data files.
- Tip: a nice, but elaborate example is the file naming system used in the Leiden Academic Centre for Drug Research, which you can find in our hand-out on file naming and folder structuring, https://www.library.universiteitleiden.nl/binaries/content/assets/ul2ub/research--publish/cds/rdm-reference-materials/file_naming_and_coding_garp-example_20151211.pdf

3.3-5 Metadata, documentation and other additional supporting information.

Not everybody uses the word 'metadata', but these questions mainly aim at acquiring information about the way you document your data. Do you use a readme file, do you describe your data in a spreadsheet, in a word file?

- Do you use a standard way of documentation that is common in your discipline, or that is prescribed by the archive in which you aim to deposit your data.
- You can find our hand-out on metadata at: https://www.library.universiteitleiden.nl/binaries/content/assets/ul2ub/research--publish/cds/rdm-reference-materials/metadata_20160211.pdf
- If you have used any anonymization or pseudonymisation techniques, you should have a log of what changes you have made and, in case of pseudonimisation a key that must be kept in a separate location.
- Examples of supporting information are lab journals, codebooks, survey questions. Analysis software might offer the functionality of exporting codes.
- The Institute of Psychology developed a list of what information to include in a 'publication package' in order to enable replication of a study. <https://www.organisatiegids.universiteitleiden.nl/binaries/content/assets/sociale-wetenschappen/psychologie/organisatiegids/instruction-for-archiving-publication-packages-july-2020.pdf>
- You should also make a note of any data that has been removed, or remark on any erroneous results.
- NB: The Guidelines for archiving in the Social Sciences require that a Readme file is archived with the data. This document records how data files are organized and internally structured, allowing users to replicate the analyses or perform new analyses. The README file also records excluded subjects, corrupted data files, and any other exceptions to the standard analysis pipeline, to enhance understanding of the data and the data analysis. <https://www.organisatiegids.universiteitleiden.nl/binaries/content/assets/sociale-wetenschappen/psychologie/organisatiegids/guidelines-for-the-archiving-of-academic-research-for-faculties-of-bss-n..pdf>.
- Tip: take a look at the following examples on how to create a good README file: <https://data.research.cornell.edu/content/readme> or closer to home:

4: Data access, sharing and reuse

In this section you will describe when and how you will share your data. Do you have any restrictions with regards to personal data, a consortium agreement, copyright, commercial interest? Do your informed consent forms allow sharing? For whom will your data be useful?

- Tip: don't forget to check (or ask your project leader/PI/supervisor) if any restriction or obligation to share your data has been stipulated in a contract by your funder or consortium.

4.1-3 Possible restrictions on sharing?

- If you have restrictions, due to the GDPR, or for other reasons, explain why you cannot share (part of the) data you are collecting.
- Check that your answer corresponds with what you wrote in 2.1 and 5.1.
- If applicable, give details of the part on data sharing in the consent forms that you used for your participants.

4.4-6 Conditions for sharing, ownership and the audience for re-use?

- This is important to discuss with your project leader/PI/supervisor so that it is clear what rights to the data you and other members of the team have, during the project or if you leave Leiden.
- How the data should be cited in publications?
- Are people in your team allowed to share your data with others?
- The most obvious audience is the wider academic community, but you might also have interest from commercial companies, or from practitioners and even the wider public depending on the area of your research. How will they find out about the existence of your data?

5: Data preservation and archiving

In this section you will describe what you will do to archive your data after your project is finished. Are there any requirements about what you do with your data when your project is completed? Which data do you want to keep for future use, either by yourself or by others? And which data will you discard? Where are you going to archive the data? Do you have a specific data archive in mind? What are your faculty's or institute's requirements for archiving data? Who will be responsible for your data after you leave?

- The Leiden University regulations state that you need to make your data FAIR. The concept of FAIR data (findable, accessible, interoperable and reusable) has been widely adopted. There are many resources to explain how to make data FAIR in the different disciplines, both for humans and machines. Read about the first steps to take on the Centre for Digital Scholarship's website:
<https://www.library.universiteitleiden.nl/researchers/data-management/fair-data>.

Find a full explanation of the FAIR principles: <https://www.go-fair.org/fair-principles/> or check the FAIR principles for software: <https://fair-software.nl/home/>

- One crucial step for FAIR data is to store your data in a certified archive.
 - Storage on the University network (J-drive) is not a certified archive
 - DANS Easy, 4TU.ResearchData and Zenodo are suitable general archives to look at.
 - Several institutes use Dataverse to archive datasets: <https://dataverse.nl/dataverse/leidenuniversity>.
 - Our catalogue of Research Data Services lists repositories per faculty: <https://digitalscholarship.nl/rds/>
 - Or look at www.re3data.org , a portal with information on repositories that you can browse by discipline.
- One of the main advantages of using a data archive is that your datasets will get a DOI or other permanent link (persistent identifier) that you and others can easily cite in a publication.
- File selection: although the Leiden regulations state that in principle research data should be kept for at least 10 years, not every single file (or version of that file) that you produce needs to be archived. As a researcher, you have the most in depth knowledge of the value of the files, but also of any limitations (for instance, patient files may have legally enforced requirements about destruction). Data underlying publications should in most cases be kept for the long term.
- FAIR data is not the same as OPEN data. We encourage you to think about how much you can deposit in a publicly available archive, but also warn you to protect information that cannot be made public, such as personal data. Those files can still be made FAIR by registering their existence and location on a metadata level.
- If the archive that you chose has specific requirements on your metadata/documentation, look at the answers you provided in 3.3-5.
- FAIR data implies that datasets have a clear license. More information on licenses for research data in this guide: <https://www.openaire.eu/how-do-i-license-my-research-data>
- Tip: as for section 4 of the DMP, check if your funder has any requirements.
- Tip: when you know in which journal your article will be published, also check the requirements your journal or publisher may have.

6: Costs

- Costs for storage: for most of research, the network capacity provided by the University is sufficient so there is no need to make any extra costs, please see: <https://www.staff.universiteitleiden.nl/ict/ict-workplace/hardware/storage-and-data-recovery>
- For projects generating Big Data, however, extra capacity may be needed, and that may involve costs. Funders such as ERC/NWO are willing to fund these costs for the duration of the project, provided you already budgeted them at the stage of your proposal.

- Also, personnel costs for data stewardship tasks (such as proper management, and curation of data) can be included in the budget.
- Costs of archiving data after the project can only be paid by the project budget is paid for before closure of the project.
- Be aware that costs for dissemination after the project, for instance in an operational database that needs maintenance, need to be budgeted and paid for in advance.
- If you think that the project data will require extra budget, you can check this tool for costs: <https://www.ukdataservice.ac.uk/manage-data/plan/costing>