WUR_MSc_DMP_TemplateAndGuidance_v01

Author(s): Irene Verhagen
https://orcid.org/0000-0001-5588-1333

Luc Steinbuch
https://orcid.org/0000-0001-6484-0920

Date: 20240726

Organization: Wageningen University & Research
https://ror.org/04qw24q55

WUR data management plan template for MSc students

- Use this template to fill in a data management plan (DMP) for your MSc thesis.
- You are free to add topics to this template to better align with your MSc thesis. However, the original topics must be retained.
- To get to additional information in the appendix for each section, hold keyboard key CTRL + left-click [info] or right-click [info] and select 'open hyperlink'.
- For more information, visit the WUR research data management website: https://www.wur.eu/rdm.
- For questions or a review of your DMP, contact your supervisor and / or the data steward of the chair group.
- If further support is required, contact data@wur.nl.
- **Please note that questions can have multiple answers checked where applicable.**
- **Make sure to attach the DMP to your MSc thesis proposal upon submission.**

A. Describe the MSc thesis [info]

1. Your name and master program:

2. Describe the organisational context of your MSc thesis.

Course code of your thesis	
Date of finalising this DMP	
Chair group(s)	
Other organisation(s) involved (when applicable)	
Supervisor / co-supervisor	
Start date of MSc thesis	
Planned end date of MSc thesis	
Umbrella project number (when applicable)	

3. Give a short description of your MSc thesis.

(Working) title	
Summary	

4. List the individuals and / or parties responsible for the following data management tasks (e.g. you, your supervisor(s) and / or the data steward).

Writing this DMP	
Supervising this DMP	
Reviewing this DMP (name supervisor(s) and / or data steward) + date of consultation	
Data collection	
Data quality	
Storage and backup	
Data archiving / publishing	
Any other role [.....]	

B. Describe the data to be collected, software used, file formats and data size [info]

5. Will you reuse existing data for your MSc thesis?

☐ Yes. Please specify below where you will get the data from (e.g. DOI, URL, or storage location) and the terms of use (i.e. a written statement on how you are allowed to use the data).

☐ No.

6. Will new data be produced during your MSc thesis?

☐ Yes.

☐ No.

7. Please describe in the table below all the data you need to store. This includes generated as well as reused existing data and code files that you manage and store. This information is useful – amongst others - to get an overview of how much storage space is needed.

File contents	Data type	Software	(Open) file format	Estimated size of each file (range)	Estimated number of files (range)
(e.g. lab analysis, gene sequence, interviews, lesion scores, etc.)	(e.g. numerical)	(e.g. Excel)	(e.g. .csv)	(e.g. 20 – 50 Mb)	(e.g. 50 – 100)

C. Structuring research data and information [info]

8. Give a (visual) representation of the folder structure you intend to use.

9. Describe the file naming conventions you intend to use. Please give one or multiple example(s).

10. How will you distinguish between versions of files?

☐ Dates within file names are updated when files are modified.

☐ The designation 'vRAW' is added to file names that contain raw unaltered data (before any processing and cleaning). Any alteration of RAW data is done on a copy of the RAW data and appended with a version number which increases with each file modification (e.g. v01, v02, v03, etc.).

☐ We will use Git versioning for code / scripts.

☐ Other, please specify below.

D. Storing data, code and readme file [info]

11. Where will you store the data, code and readme file during your thesis? Include platforms you use to share data (e.g. with your supervisor), collect data on, or send data to for processing or analysis. This includes any used external hardware or computers.

☐ WUR OneDrive for Business (automatic backup and possibility to share with supervisors, which is strongly advised).

☐ Git@WUR (GitLab locally hosted at WUR; dedicated to saving and updating programming code).

☐ A WUR storage solution, organised by my supervisor (e.g. W:drive, Yoda@WUR, Sharepoint, MS Teams). Please, specify.

☐ Other, please specify.

E. Personal or otherwise sensitive data [info]

12. Do you collect / process personal data or otherwise sensitive data for your MSc thesis?

☐ Yes.

☐ No. Please, continue to question 15.

13. Please specify the privacy or sensitive data protection measures. Think of data minimisation, pseudonymisation / anonymisation / informed consent forms / separate storage / access management).

14. Are there other ethical issues (e.g. when working with animals or when working with people in a medical setting) that need to be considered and for which your supervisor needed approval from ethical committees?

☐ My supervisor indicates that further ethical approval is not required.

☐ My supervisor requested approval at the appropriate ethical committee(s).

F. Preserving data, code and readme file [info]

15. Where at WUR will the data, code and readme file from your MSc thesis be archived when finished?

☐ Data will be send to my supervisor via SURFfilesender and archived on the W:drive.

☐ I have been provided with access to Yoda@WUR where the data will be archived.

☐ Other, please specify below.

G. Discussion points for / with supervisor(s) [info]

For the following discussion points, please have a look at the guidance. When applicable, any agreements etc. can be written down in the fields below.

☐ **Intellectual Property rights.**

☐ **Data rightholder(s) and access.**

☐ **Publishing data and choosing a licence.**

Guidance: additional information (may be deleted after completion)

A guidance. Describe the MSc thesis.

1. Name MSc student and master programme.

Please add your full name and the master programme you follow.

2. Organisational context of your MSc thesis

A data management plan should always include contextual information. Then, it is clear to whom / which MSc thesis the filled in data management plan belongs and to which data the described data management practices apply. Please, add the chair group(s) your MSc thesis is affiliated with. Please carefully check the spelling. Preferably copy and paste directly from aforementioned website.

3. Description of your MSc thesis

Providing the (working) title and a short description of your MSc thesis helps the reader to understand your work and put the description of management of research data into context.

4. Data management responsibilities

Identifying persons or parties who play a role in your daily data management practices helps to clarify the data collection process. Identifying these roles are also important after completion of your MSc thesis. For example, for long-term accessibility of the data, responsibility for the archived data should lie with more than one person. Feel free to add roles.

B guidance. Describe the data to be collected, software used, file formats and data size

5. Reuse existing data

When others generate data, whether archived at e.g. WUR or published in a repository, this gives you the opportunity to reuse these data within your MSc thesis. However, make sure that you are aware of the terms of use (e.g. the licence) of the data that you want to reuse. The terms of use provide you with what you are allowed to do with the data. Want to find existing data? Have a look here for places to start.

In the case of not reusing data: think of the following potential reasons why the reuse of existing data was considered, but not implemented. Is the data you want to reuse not publicly available or otherwise restricted in access? Is the data not yet collected? Would the costs be too high (e.g. acquiring specific software)? Is the data too big in size (e.g. TBs)?

6. New data

Indicate whether you will collect / generate new data in the research project.

7. Description newly generated data.

Data type refers to whether the data can be classified as for example textual, numeric, audio, film etc. Note that data includes any processing or analysis scripts, protocols, or any other output resulting from research (figures, statistical output, processed data etc.). Try to use software that let you also save files in an open format, so that others can open the files (now and in the future) even if they don't have the software. Providing the estimated number of files and file size helps to indicate for example the storage requirements and costs.

C guidance. Structuring research data and information

8. Folder structure

Designing a logical folder structure (for tips, see here) ensures that you and e.g. your supervisor can easily locate data now and in the future. Provide how you plan to organise your (sub)folders. If you already have a folder structure in place, instead of having to go through the whole folder structure and write it down is via the following powershell workflow (note: the following steps do not work with zip files):

- Go to your file explorer
- Open the parent folder
- In the address bar, type and press enter afterwards: powershell
- A blue screen pops up. Continue in that screen
- Type and press enter afterwards: Get-ChildItem | tree > foldertree.txt
- There is now a file your folder [foldertree.txt], from which you can copy the content in the answer field.

If you're using Git Bash on Windows, go to your parent directory (or Main local Git directory) and use:

- cmd //c tree //a > foldertree.txt
- There is now a file in your folder [foldertree.txt], from which you can copy the content in the answer field.

If you are using Linux:

- sudo apt install tree.
- Go to your project parent directory.
- tree -d > foldertree.txt
- There is now a file in your folder [foldertree.txt], from which you can copy the content in the answer field.

9. File naming conventions

Applying a consistent and descriptive file naming convention (i.e. a systematic file naming method) helps to identify the content of a (data)file without opening it, easily and quickly locate, retrieve and filter (data)files, even if they have changed folders and easily sort and browse through your (data)files. When creating a file name convention, consider the following (more tips here):

- Keep file names short (max. 30-35 characters).
- Use abbreviations to avoid long file names.
- Using the international standard for date (YYYYMMDD).
- Apply a leading zero to the version number (e.g. v01).
- Avoid using spaces, periods and special characters. Use dashes and / or underscores instead.

A basic example of a file name:

[project]_[filesubject]_[subsubject]_[date]_[version].[extension]

10. File versioning

Make sure that you have a system in place to keep track of file versions. A simple and effective system is to incorporate version numbers (e.g. v01, v02 etc.) in your

file names. Additionally, depending on the type of files you work with, you could use systems that keep track of versions of individual files (e.g. OneDrive, Sharepoint) or multiple files. For example, Git@WUR, a version control system for code / scripts / software which is also available for students. Note: for files in Git, dates and version numbers should not be added to the filenames. Git is version control software, keeping track of versions for you and without having to apply file versioning yourself. When applying versioning by changing file names, Git would not recognise a new version as such, but as a completely different file.

D guidance. Storing data, code and readme file

11. Where will data be stored during research.

Ensure that you store data safely. Safe storage solutions would be for example the storage solutions indicated in the WUR data storage finder. Not all storage solutions provided by WUR are (directly) accessible for MSc students. We recommend to use WUR OneDrive for Business, please see here how to get started / install. When other applications are needed to collect / process data, discuss with your supervisor(s) and check the ApprovedApps tool. Is the ApprovedApps tool advising against the use of the application you chose or not (yet) in the list, discuss with your supervisor how to continue. Avoid storing research data on hard disks, USBs, personal laptops (with the exception of WUR OneDrive for Business), and third party cloud services such as Google Drive and Dropbox (exceptions excluded).

Data includes eventual code and always a readme file. The readme file contains information about e.g. the steps that have been undertaken in processing and analysing data. In short: all information necessary to understand the data, reproduce research and verify results. WUR Library advises to use the WUR readme file template (DOI: 10.5281/zenodo.7701727) as the minimum required documentation to add to the data.

E guidance. Personal or otherwise sensitive data

12. Personal or otherwise sensitive data.

Working with personal data (i.e. any information that (indirectly) relates to an identified or identifiable living individual) or otherwise sensitive data (e.g. financial, governmental or company data) requires a considered approach to data sharing, storage, analysis and use. Indicate whether you will collect or process personal or otherwise sensitive data for your MSc thesis.

13. Privacy or sensitive data protection measures.

Working with personal or other sensitive data (and thus in accordance with the WUR policy, laws, privacy law, or GDPR), requires careful processing and storage. Describe which protection measures are applied to the data collected / processed for your MSc thesis. Examples of protection measures to apply are:

- Data minimisation; only collect the data necessary for the purpose of the MSc thesis.
- Informed consent form.
- Access management; restricting access to the data to only the persons that need to have access (i.e. legitimate users).
- Secure data storage and transfer (encryption).
- Pseudonymisation or anonymisation.

For more information about information security and privacy, have a look at this page or send an email to privacy.student@wur.nl.

14. Other ethical issues

Here, discuss with your supervisor whether you need or already have ethical clearance from e.g. an animal experimental committee or a medical or social ethics review committee.

F guidance. Preserving of data, code and readme file

15. Where will the data be preserved

Where will data be preserved for the long term? Where and if to preserve data also depends on its sensitivity. Discuss with your supervisor how and where to preserve data.

G guidance. Discussion points for / with supervisor(s)

Intellectual Property Rights
Getting intellectual property gives you the right to benefit from your new idea, innovation or product. The way intellectual property rights (IPR) are regulated by law has now been clarified for students. For more information, see this page. Further:

- Check your IPR position (with your supervisor) before you start your thesis: what is laid down in your contract / learning agreement?
- Read the IPR guidelines for students.
- Conclude clear agreements and record them in writing.

Data rightholder(s) and access
The party that is the (rights)holder over the data decides what others are allowed to do with it. You own the rights to the MSc thesis / Research Practice as stated in the Intellectual Property statement in the learning agreement. However, it could be that you have been provided with data from e.g. a company. As such, when data is created or provided by others, it is important to discuss with them who will be entitled to the data and what you are allowed to do with the data. A contract or agreement addressing data storage, sharing, access, confidentiality etc. might then be involved.

Publishing data and choosing a licence
Potentially, your MSc thesis results in a scientific article. Then, it is recommended to publish the data underlying the article in a data repository. Additionally, data not used for a publication can have (future) value for reuse and publishing these data in a data repository could be discussed with your supervisor. Data repositories compile, manage, preserve, share and provide access to data, metadata and documentation. Publishing the data and readme file in a data repository promotes data reuse and verifiability. Already, discuss in which repository data could be published (more info, see this page). Also discuss under which licence the data may be reused by others. Check the commonly used Creative Commons licences. WUR Library advices the CC BY licence for data, because WUR adheres to the motto 'as open as possible, as closed

a necessary' and researchers should at least be attributed for their scientific effort. CC BY means that other researchers should credit you as a creator of the data.

Before publishing data, take the following into account:

- Is (part of) the data personal data or otherwise sensitive? Then, these data can, very likely, not be published.
- Is there another rightsholder involved? Make sure to check the agreement (see Data rightsholder(s) and access).