**Data Management Plan\***

This data management plan has been designed in accordance with the data management policy of WUR, for use by ELS MSc thesis students. Please ask your supervisors if you have any questions.

\* Please note that you have to submit a draft version of this form with your thesis proposal to your supervisor. Thereafter, at the end of your thesis, you have to submit the final version of this form with your final thesis report, along with all your corresponding data and syntax files (see specification below). You may bring your data and syntax files to your supervisor/ELS Secretariat in a USB stick or via SURFfileSender or WeTransfer.

**Format for a data management plan**

* *The completed DMP should be submitted at the same time when you submit your final thesis report, together with all the data that you used/worked with. Thus it is useful to already start thinking and planning (and also filling out this plan) on how you will organise and maintain your data and analysis documents while carrying out your research.*
* *This format is intended to give you a helping hand in writing your DMP. You are free to add content elements when your particular research project requires it. The DMP should however include answers to the nine sections below.*
* *Hover over (info) for more information about each section. It will give you a hyperlink (CTRL-click) to additional information in the Appendix to this format. The examples may sometimes be of much larger projects than an MSc thesis, but still relevant. Keep that in mind when you look at them.*
* *The words in italics are meant to guide you in filling out the spaces. The size of the boxes are not indicative of how much text you need to type; please feel free to use more space if needed.*
* *Any remaining questions? Contact your supervisors or* *data.library@wur.nl*.

 *(you may delete the text above after completion of your DMP)*

## Organizational Context [(info)](#_1._Organizational_context)

|  |  |
| --- | --- |
| **Name of student** |  |
| **Student Registration Number** |  |
| **ELS Thesis Code** |  |
| **Date** |  |
| **Chair group** | **Education and Learning Sciences** |
| **Other persons involved** | *e.g. Please include name of your supervisor(s) along with any other external parties involved, as well as their role in the project* |
| **Start date of project** |  |
| **File name of this DMP** | *e.g. Date\_DMP\_MScthesis\_nameofstudent.docx* |

## Short Description of Research [(info)](#_2._Short_description)

|  |  |
| --- | --- |
| **Title of Thesis/Project** |  |
| **Abstract** |  |

## Define Data Management Roles [(info)](#_3._Data_management)

|  |  |
| --- | --- |
| **Roles** | **Who is responsible for what?** |
| **Who is collecting the data?***Even in a modelling study, someone, most likely you as MSc thesis student) will collect the required input data for the model, even if these are already available.* | *Name of researcher: ...**Role of the researcher:**• Collects and processes the data in an organized way. (S)he is responsible for the quality and authenticity of the data and ensures that they are safely stored during the project (i.e. short-term storage). The researcher will allow the supervisor and others involved in the research access upon request.**• Responsible for documentation and submission of the data files for long-term storage (after finishing the publication, thesis or completion of the project).* |
| **Who is analyzing the data?** | *Name of researcher: …**Role of the researcher:**• Analyses the data in an organized way. (S)he is responsible for the quality and authenticity of the results and ensures that they are safely stored during the project (i.e. short-term storage). The researcher will allow the supervisor and others involved in the research access upon request.**• Responsible for documentation and submission of the result data files for long-term storage after finishing the thesis).* |
| **Others***Are there other people contributing to data collection and analysis, for example by assigning codes?* |  |
| **What is the role of your supervisor?** | *Name of supervisor: …**Role: The supervisor or project leader checks the data files before long-term storage. The supervisor is responsible for the data quality, clarity and completeness* |

## Overview of research data, software choices, data size [(info)](#_4._Overview_of)

*Please only complete cells that are relevant for your research; you may add cells or change the categories accordingly*.

|  |  |  |  |
| --- | --- | --- | --- |
| **Data Stage** | **Specification of type of research data** | **Software used** | **Data Size** |
| **Raw Data***(e.g. data from experiments or observations)* | *(e.g. notes from observations; hard copies of paper surveys handed out)* | *(e.g. MS Word)* | *(e.g. 6MB)* |
| **Processed Data**(e.g. raw data that has been entered, cleaned and processed) | (e.g. transferring survey responses onto programs on the computer; results of analysis etc) | (e.g. SPSS, MS Excel) |  |
| **Model & Syntax**(e.g. codes or formula of your model, as well as the syntax to explain what the various codes/formula mean) | (e.g. Model developed for XXX etc) | (e.g. MS Excel, R) |  |
| **Others (if any)** |  |  |  |

## Structuring your data and information [(info)](#_5._Structuring_your)

Describe the system for directory and file names of the documents you submit. See the examples in appendix for inspiration.

My directory- and file naming structure:

|  |
| --- |
| *To avoid confusion about different versions of data files, dates of last change should be added at the beginning of the file name using the following format: (YYYYMMDD). This way facilitates sorting the files according to the version’s date and previous versions are not overwritten.* |

## Sharing and ownership [(info)](#_6._Sharing_and)

|  |  |
| --- | --- |
| **Sharing and Ownership** | **With whom, what and how?** |
| **Data sharing** - Do you expect that others may be interested to re-use your data? Do you have plans to share your data with these parties? - How are you going to make sure your data files will be accessible once you leave the department? Who will take care of your data? |  |
| **Data ownership**- Any funder’s requirements to share you data, or to impose an embargo? - Are there agreements on how the data will be used and shared within your group or with other parties involved in this research? (outside your group or outside Wageningen UR) |  |
| **Privacy****-** Are there any agreements being made with regards to whether the data should be kept private and confidential?- Who can or cannot see the data? |  |

**For ELS office use:**

|  |  |
| --- | --- |
| **Name of ELS staff** |  |
| **Files received via** | USB Stick / Email / WeTransfer / Others, Please specify: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |
| **Where is the data centrally stored?** | *(e.g. Give name and folder of location)* |
| **Any specific storage instructions?** | *(e.g. duration of central storage, who has access, who can grant access to others etc)* |

## Appendix with additional information

*(this section may be deleted after completion of your DMP)*

### 1. Organizational context

A data management plan should leave no doubt as to which research/er(s) it belongs to.

[(Back to top)](#_Organizational_Context_(info))

### 2. Short description of research

Giving a short description of your research gives context to your data management. It makes it easier for the reader to understand without having to check your research plan.

[(Back to top)](#_Short_Description_of)

### 3. Data management roles

Identifying persons who are – or can be – of assistance in your daily data management practices, smoothens your data collection process. Maybe some people have special responsibilities regarding data management? (e.g. a division of labor between programmers and those who do observations?)

Having a closer look at data management roles places data collection in a broader perspective than your research project alone. Discussing both your roles as well as those of your supervisor and other colleagues prevents possible future issues concerning data ownership.

[(Back to top)](#_Define_Data_Management)

### 4. Overview of research data, software choices, data size

A) Type of research data

It is good to think about your possible research data before you actually start collecting these data. This helps to make sure no research output is overlooked. You can choose from:

- Raw data (e.g. data from experiments or observations);

- Tangible data (e.g. videos, hard copy surveys etc);

- Derived / processed data\*

- Models (including data from simulations)

\*If you used derived data, you should at least say how you handle the raw data. If you haven’t produced the data yourself, that may be of influence on what you are allowed to do with the data. For processed data, please also take note to include information to understand the processing e.g. SPSS syntax, coding steps, data cleaning logs, sufficient meta data to allow others to check and/or replicate the process leading to the results.

**Please note that no matter what kind of data you end up using and analysing, whether collected or derived/processed, (e.g. original collected data, the original data file(s), final data file(s) used for analysis, and the syntax files), please store all of it and submit them together with this data management plant.**

There is a diverse range of different types of research data and some of these include:

- Documents (text, MS Word), spread sheets

- Scanned laboratory notebooks, field notebooks, diaries

- Online questionnaires, transcripts or surveys

- Digital audio or video recordings

- Transcribed test responses

- Database contents

- Digital models, algorithms or scripts

- Contents of an application (input, output, log files for analysis software, simulation software, schemas)

- Documented methodologies and workflows

- Records of standard operating procedures and protocols

B) Software choices

What software choices affect whether current and future users can actually view and use the data you collected. If you use proprietary software, for example, it may not be possible for people outside your field to do anything with your data except getting an error while trying to read them. Also, some software produces data in specific file formats, and may come with its own systems for folders and file names. Think software choices through with future users in mind.

C) Data size

Give an estimate in (Mega-Giga-Tera) bytes. Knowing this will give an indication as to what kinds of data storage solutions are needed.

An example of how to fill out the table in this section can be found below:

|  |  |  |  |
| --- | --- | --- | --- |
| **Data Stage / Type of Research Data** | **Specification of type of research data** | **Software used** | **Data Size** |
| *e.g. Data from questionnaires* | *I will need to collect data through a questionnaire. The questionnaire will be done on Qualtrics and the results will be downloaded to be used on SPSS/excel.* | *Qualtrics,* *SPSS/Excel (.csv)* | 1Mb |
| *e.g. Model parameter values*  | *I will need to gather information to use as model parameter values, for example on pathogen removal rates for different types of sewage treatment, or the die-off under specific environmental conditions. I plan to make an overview file of different parameter values found in literature or from other sources.* | *Excel (.csv)* | 1Mb |

[(Back to top)](#_Overview_of_research)

### 5. Structuring your data and information

We all think we are going to remember how we named our files and where we store them. But the truth is, we never do. Time invested in giving thought about an ambiguous directory and file-naming systems pays off for your future self.

Some basic tips for **file-naming and version control**:

Use descriptive names for files

(not dataset1 but pathogenmeasurement130119\_V1.0.xls)

Indicate versions, e.g. \_v1.0 (master files/milestone files)

Describing your folder structure is meant as an exercise in logic. It is intended to help you structure your data collection process. Of course, the folder structure may be incorporated in different working environments.

Please, find two examples of proposed folder arrangements and file naming strategies from two recent Wageningen University DMP’s below:

***Example 1****: The proposed folder arrangement and file naming strategy from the DMP by Beatriz Ramírez, Earth System Science Research Group, Wageningen University*



Figure 1. Proposed folder arrangement and file naming strategy (Example 1)

Example 2: The proposed folder arrangement and file naming strategy from the DMP by Lucie Vermeulen, Environmental Systems Analysis Group, Wageningen University



Figure 2. Proposed folder arrangement and file naming strategy (Example 2)

[(Back to top)](#_Structuring_your_data)

### 6. Sharing and ownership

Legally, in principle, the data belongs to the university, unless other agreements have been made with external parties. Thus, it is important to have a sound understanding on what you are allowed to do with the data and how you will leave your data behind when the time comes to pursue your career at another University or organization. Therefore, ownership is not so much about ‘property’ (to whom the data belongs to). It is about custodianship: what is going to happen to the data when your project is finished? Who is the person responsible for taking care of your data and ensuring it can be accessed when you are gone? Can you still publish about the data and use them for further research when you have left the university? With whom are you going to make these arrangements and how is the rest of the world going to know? These questions are generally best discussed with your supervisor(s) and parties involved (e.g. funders etc).

[(Back to top)](#_Sharing_and_ownership)