

***This data management plan template is designed by the library and adapted for use by ESA MSc thesis students by Nynke Hofstra and Wichertje Bron. Please ask them if you have any questions. This is an appendix to the MSc data management policy available from the MSc thesis brochure, available from*** [***this URL***](http://www.wageningenur.nl/en/Expertise-Services/Chair-groups/Environmental-Sciences/Environmental-Systems-Analysis-Group/Education/Brochures-and-forms.htm)***.***

**Format for a data management plan**

* *The completed DMP will be an appendix to your MSc thesis contract and will be subject to review by your thesis supervisor.*
* *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 are mostly for much larger projects than an MSc thesis. Keep that in mind when you look at them.*
* *Any remaining questions? Contact* *data.library@wur.nl*.

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

**1.** **Describe the organizational context** [(info)](#organizationalcontext" \o "information)

|  |  |
| --- | --- |
| **Name** |  |
| **Date** |  |
| **Chair group** | *Environmental Systems Analysis Group* |
| **Supervisor/ (co-)promotors** |  |
| **Start date of project** |  |
| **File name of this DMP**  | *Date\_DMP\_MScthesis\_lastname.docx* |

**2. Give a short description of your research project** [(info)](#shortdescriptionofresearch" \o "information)

|  |  |
| --- | --- |
| **Title** |  |
| **Abstract** |  |

 **3.** **Define data management roles**

[(info)](#Datamanagementroles" \o "Information)

|  |  |
| --- | --- |
| **Roles** |  |
| 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 according to ESA guidelines (after finishing the publication, thesis or completion of the project).*
 |
| Who is **analysing** 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).*
 |
| **Other**(Do other persons contribute, for example by writing code?) |  |
| 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* |

**4. Give an overview of expected** **type of research data** [(info)](#Typeofresearchdata" \o "Information), **software choices** [(info)](#Softwarechoices), **data size & growth** [(info)](#Datasizeandgrowth)

|  |  |  |  |
| --- | --- | --- | --- |
| **Data stage** | **Specification of type of research data** | **Software choice**  | **Data size/growth**  |
| Raw data  |  |  |  |
| Processed data |  |  |  |
| Models/code  |  |  |  |
| Other? |  |  |  |

**5.** **Short term storage solutions\*** [(info)](#Shortermstorage)Describe where the data will be stored physically and how the back-up is organised.

|  |  |  |
| --- | --- | --- |
| **Data stage** | **Storage location** | **Backup procedures**(storage medium and location/ how often?)  |
| Raw data  | *M-drive* | *IT daily* |
| Processed data | *M-drive* | *IT daily* |
| Models/code  | *M-drive* | *IT daily* |
| Other? | *M-drive* | *IT daily* |

*The data will be saved on the personal M drive of the researcher, to ensure daily back up by ICT-WUR. If storage capacity is too limited the D drive could be used for storage, but the researcher is responsible for sufficient back-up.*

**6.** **Structuring your data and information** [(info)](#Directoryandfilenaming" \o "Information and examples)Give a visual representation of the system for directory- and file names you intend to use. See the [examples](#ExamplesDirectory) for inspiration.

*My proposed 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.*  |

Does your workflow provide for version control? If not, describe how you intend to keep versions apart.

|  |
| --- |
|  |

 **7.** **Documentation and metadata\*** [(info)](#DocumentationandMetadata" \o "Information)

Describe how you are going to document your data collection process, what the resulting data files comprise and how they will be processed further. Think about documenting the:

*1. content (what does your dataset contain?)*

*2. context (who, what, why, where and how will the data be collected and analysed)*

*3. process (are there specific processes and does it make sense to organise notes by process?)*

*NB: Good data documentation ensures that:*

* *The data are organized in such a way that it facilitates efficient data analysis and minimizes errors.*
* *Data are self-explanatory and understandable to supervisors/project partners in order to facilitate communication and teaching.*
* *Data can be retrieved, understood and properly interpreted, now and in the future, as relevant context is available.*

|  |
| --- |
| *See the* [*ESA thesis brochure*](http://www.wageningenur.nl/en/Expertise-Services/Chair-groups/Environmental-Sciences/Environmental-Systems-Analysis-Group/Education/Brochures-and-forms.htm) *for detailed guidelines on organisation of the data files and folders. In the main folder* ***add a readme file*** *(preferably ASCII file format) explaining the data files, content and structure. Use the folder headings: ‘Original data’, ‘Calculations/scripts/interpretation’, ‘Final data’, ‘Documentation’.* |

**8.** **Sharing and ownership\*** [(info)](#SharingandOwnership" \o "Information)

|  |  |
| --- | --- |
| **Sharing and ownership**  | **(With) who(m), 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 datafiles 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) | *NB: ownership isn’t so much about ‘property’ (to whom do the data belong). It’s about custodianship: What is going to happen to the data when the project is finished? Who is the person responsible for taking care of the data and ensuring it can be accessed.* *The project leader or supervisor is: …**The project leader or supervisor (permanent staff member), in consultation with other researchers involved in the data collection, is responsible for long term data storage and access/use by others, unless stated otherwise in the individual RDMP.* |
| **Privacy**- Are there privacy or security issues, and if there are, how are you dealing with them? |  |

**9.** **Long term storage\*** [(info)](#Longtermstorage" \o "Information)
Which part of your research data has value for long term storage? Do you intend to preserve these data for the long term?

|  |  |
| --- | --- |
| **Yes or no ?** | **Argumentation** |
| *yes* | *The code of conduct for scientific practice requires that you retain your data for five years after you have published your article/thesis and that you make it available upon request for verification purposes.* |

Which data archive do you intend to use?

|  |
| --- |
| *After completion of the thesis/papers/project, data files will be submitted to the secretary for long term storage (at least 5 years). For ESA data will be stored in a dedicated central location for ESA data storage on the server of ICT.*  |

**Appendix with additional information**(may be deleted after completion of your DMP)

**1. Organizational context**
A data management plan should leave no doubt as to which research(er) it belongs.

[[Back]](#organizationalcontextback)

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

[[Back]](#shortdescriptionofresearchback)

**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 labour 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](#SharingandOwnership).

[[Back]](#Datamanagementrolesback)

**4a. Type of research data**Identifying your possible research data before you actually start collecting those data, makes sure no research output is overlooked. You can choose from:

*- Raw data; (e.g. data from experiments or observations (e.g. a lab note book))*

*- Derived / processed data*

*- Models (including data from simulations)*

If you use derived data you should at least say how you handle the raw data. (NB If you haven’t produced the data yourself, that may be of influence on what you are allowed to do with the data (see [data ownership](#SharingandOwnership)))

To give you an example of the diverse outputs of research data, read the [following list](http://www2.le.ac.uk/services/research-data/documents/an-introduction-to-managing-research-data):

• 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

**Example**An example of how the question on type of research data and [software choices](#Shortermstorage) might translate into the DMP format, is given below. For the complete DMP by Lucie Vermeulen we took this excerpt from [follow this link](http://www.wageningenur.nl/web/file?uuid=547da348-12b4-49c2-aebe-b813645ab19f&owner=0164641e-ab10-40db-b4a7-876f39d24a28).

|  |  |  |
| --- | --- | --- |
| **Type of research data** | **Specification** | **Software choices** |
| **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) |
| **Model input data - gridded** | Existing datasets on for example climate and hydrology | Depending on how they are available. Runoff data are .grd files, for example, most others I don’t know yet |
| **Model input data –****country data** | Existing datasets on for example population and livestock density, land use | Depending on how they are delivered, perhaps a spread sheet format. I may all convert these to csv files |
| **Model input data - metadata**  | For all input data, one document containing all metadata will be created, specifying at least source, time period, region, measurement method, type of data, unit of measurement, access rights, date downloaded. All data will be checked for consistency, and any changes made to input data will be documented. | Excel (.csv) |
| **Etc.**  | See the [DMP by Lucie Vermeulen](http://www.wageningenur.nl/web/file?uuid=547da348-12b4-49c2-aebe-b813645ab19f&owner=0164641e-ab10-40db-b4a7-876f39d24a28) |  |

[[Back]](#Typeofresearchdataback)

**4b. Software choices**What software will you use to create, analyse and visualize your data? Are these choices common practice in your field?

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 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](#Directoryandfilenaming). Think software choices through with future users in mind.

[[Back]](#Softwarechoicesback)

**4c. Data size/growth**Give an estimate in (Mega – Giga – Tera) Bytes. Giving an educated guess on the size of your research data output, gives you an indication about where you should store your data. If you will produce terabytes of data, for example, a simple hard drive won’t do. So, data size influences [data storage solutions](#Shortermstorage).

[[Back]](#Datasizeandgrowthback)

**5. Short term storage**

You need to decide how you will keep your data safe in the short-term. Where will the data be stored physically and how will it be backed up? Do you follow the common practice in your research group, and if not, why?
The table below may be of assistance in making an informed choice for short term storage.

*USB-stick by* [*http://dpconline.org/*](http://dpconline.org/) *shared* [*via twitter*](https://twitter.com/digitalfay/status/411444578122600450/photo/1)

|  |  |  |  |
| --- | --- | --- | --- |
| **Storage solutions** | **Advantages** | **Disadvantages** | **Suitable for** |
| **Personal computer & Laptop** | Always availablePortable | Drive may failLaptop may be stolen | Temporary storage |
| **Networked drives**File servers managed by your research group, university or facilities like a NAS-server | Regularly backed upStored securely in a single place. Centralized storage makes it easier to maintain, backup.  | Costs | Master copy of your data(if enough storage space is provided …)  |
| **External storage devices**USB flash drive, DVD/CD, external hard drive | Low costPortability  | Easily damaged or lost  | Temporary storage |
| **Cloud services**Like Dropbox, SkyDrive, etc.  | Automatic synchronization between files online and folder on PCEasy to use and access  | It’s not sure whether data security is taken care ofYou don’t have direct influence on how often backups take place and by whom | Data sharing  |

[[Back]](#Shortermstorageback" \o "Back)

**6. 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 :-) Let alone, our fellow researchers. Time invested in giving thought about an unambiguous directory and file-naming systems pays of for your future self.

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

* *Use descriptive names for files*(not: dataset1 but pathogenmeasurement021213\_v01.xls)
* *Indicate versions, e.g. \_v01 (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 (like for example Sharepoint, ATLAS, an electronic lab journal, etc.). Depending on the workflow you use, versions of your datasets and documents may be kept automatically or not. **Examples**

* For inspiration on shaping your folder structure you can have a look at the recorded video of a [presentation](http://wurtv.wur.nl/P2GPlayer/Player.aspx?path=roadkit1/2013/10/08/2/video_post.wmv&mediaType=recordings) (01:10:30 until 01:16:30) by Mari Wigham (Wageningen UR Food & Biobased Research) on structuring your data. This presentation was part of the half-yearly [Data Management Course](http://www.wageningenur.nl/en/article/Data-Management-3.htm) organised by WUR Library.
* 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 (hier een link naar dit DMP online!) by Beatriz Ramírez, Earth System Science Research Group, Wageningen University*



 **Figure 1. Proposed folder arrangement and file naming strategy.**

***Example 2****: The proposed folder arrangement and file naming strategy from the* [*DMP by Lucie Vermeulen*](http://www.wageningenur.nl/web/file?uuid=547da348-12b4-49c2-aebe-b813645ab19f&owner=0164641e-ab10-40db-b4a7-876f39d24a28)*, Environmental Systems Analysis Group, Wageningen University*

 **Proposed folder structure of PhD Lucie Vermeulen**

**Papers**

 PDFs

 Paper MSc thesis 2012

 Paper 1

 Paper 2

 Paper 3

 Paper 4

**Data**

 Gridded

o Hydrology

o Point sources

o Diffuse sources

o Climate

o Geophysical

 Country

o Land use

o Population

o Livestock

o Sewage treatment

 Pathogen measurements

**Model**

 Model scripts

 Model input

 Model output

o Graphs
o Tables
o Maps

o Tests

**Administration**

 Meeting agendas and notes

 Financial

 Planning

 PhD proposal

**SENSE**

 General information

 A1 course

 Data Management course

 PhD Assessment

 …

**Conferences**

 Health Related Water Microbiology 2013

 …

**Miscellaneous**

 We Day

 IPCC review

 …

**Archive**

[[Back]](#Directoryandfilenamingback)

**7. Documentation and metadata**Good documentation [ensures your data can be](http://www2.le.ac.uk/services/research-data/organise-data/metadata):

*• Searched for and retrieved*

*• Understood now and in the future*

*• Properly interpreted, as relevant context is available.*

Information on how the research was performed may come in different forms: standardized protocols, manuals of equipment or software, field notes on paper, e-mails from colleagues etc.

Depending on the type of research, some simple solutions for data documentation may work. Some groups have developed best practices to store raw and processed data in different sheets of an Excel workbook, and use the first sheet to document the research process and the meaning of the subsequent sheets (legend). Other groups simple use shared drives and organise them logically like we discussed under 6. In combination with simple conventions (e.g. ‘everybody who uploads a file to a folder adds a few lines to ‘readme.txt’ to explain what it is) documentation is added.

Some research groups support more sophisticated solutions like an electronic lab notebook or a program like atlas.ti.(Groups that consider adopting an electronic lab notebook may want to use this site <http://campusguides.lib.utah.edu/ELNs> as a starting point). Even with such a system, you will have to go through the process of analysing our own processes and organise the application accordingly.

 **Examples**

1. To provide a **first example**, we show you how the dataset by dr. ir. P.A.J. v Oort on *Key weather extremes affecting potato production in the Netherlands* was documented. It is available at DANS [https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset:55620/tab/2](https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset%3A55620/tab/2)
Below you find two types of data documentation:
2. **Readme.txt** gives a description of all data files and of all documents describing the content, context as well as the process of data collection
3. **Methodology.txt** describes the data collection process

|  |
| --- |
| **Readme.txt**This dataset contains the underlying data for the studyVan Oort, P. A. J., B. G. H. Timmermans, H. Meinke, and M. K. Van Ittersum. "Key weather extremes affecting potato production in The Netherlands." European Journal of Agronomy 37, no. 1 (2012): 11-22.<http://dx.doi.org/10.1016/j.eja.2011.09.002>Purpose and method of data collection is described in methodology.txtBibliographic details of the reports used for agronomic data can be found in metadata.csv The current adresses of meteorological data sources that were used for the study can be found in knmistationsdata.txtNote that data for other crops than potatoes have been collected, see crops.txtDatafiles:All data is provided in a propietary Excel 2013 workbook: verzameldedatasets\_Oort2WULibrary20130920.xls From this file non-propietary csv files (for the numerical data) as well jpg files (for the graphs) have been produced: consaard.csvmetadata.csvextremeyears.csvgraphs: sugarbeet.jpggraphs: winter wheat.jpgwijnandsrade.csvbedrijvenineigenbeheer.csvcranendonk.csvvredepeel.csvde\_schreef.csvgraph\_de\_schreef.jpgrivrodronten.csvrivrowageningen.csvcbs\_de\_jager.csvbietenstatistiek.csvwestmaas.csvsvp.csvgraph\_svp.jpgcbs\_flevoland.csvaardappel19731999minderhoudhoeve.csvgraph\_minderhoudhoeve.jpgcrops.txtknmistations.txtmethodology.txtreadme.txtNames correspond with the sources in metadata.csv |



1. To provide a **second example**, we show you the info and legend sheets of some datasets (in .xls) belonging to research by Lennart Suselbeek, Resource Ecology Group and Forest Ecology and Management Group, Wageningen University. Below you find two types of data documentation:

	1. The first sheet of the excel sheets with the data is an **information tab**. It gives details about the project and provides keywords
	2. The second sheet is a **legend** to the rest of the sheets. It entails a sheet explaining all the codes that are used in the DATA sheet.





[[Back]](#DocumentationandMetadataback)

**8. Sharing and ownership**
Legally, the ownership of research data isn’t very clear. So, it’s 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 organisation. Therefore, ownership isn’t so much about ‘property’ (to whom do the data belong). It’s 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 you supervisor and funders.

[[Back]](#SharingandOwnershipBack)

**9. Long term storage**The code of conduct for scientific practice requires that you ***retain your data for five years after you have published your article*** and make it available upon request for verification purposes. You may be able to fulfil such requests while you are in your present job, but to make data available for a longer period for re-use and verification you should store it in a data archive with proper documentation and in a sustainable data format.

Research data should always be retained for the short term. Additional long term storage is at least recommended if:

*1. data underlies publications
2. long term storage and sharing is required by funders
3. there is a legal requirement
4. it is likely that other may want to re-use the data*

In the flow chart below you can check whether your data are a possible candidate for long term archiving and re-use. If your answer(s) to the question(s) is YES you should at least consider storing these data for the long term.



Of course, practical considerations may make you decide not to pursue long term storage even though your data might be of future value. Think about the following:

*- Can the data be understood? Is the data documentation clear enough? If not, long term archiving is of no use.*

*- Are there any legal requirements NOT to share the data? (privacy, etc.)*

*- Does the expense of archiving justify the long term value?*

If you choose long term storage is there a common practice/institution in your field? Or do you intend to use the services provided by Wageningen UR? Wageningen UR Library has agreements with the repository from The Royal Netherlands Academy of Arts and Sciences, called [DANS EASY](https://easy.dans.knaw.nl/ui/home) as well as with the data repository for the technical sciences [3TU.Datacentrum](http://datacentrum.3tu.nl/en/).
You can contact  data.library@wur.nl for more information.

# **Examples**

# - For more inspiration [see how Frits van Evert](http://www.wageningenur.nl/en/show/Dataarchivingdoubleinterview.htm), researcher at Agrosystems Research, supports archiving research data in a sustainable way- The dataset by dr. ir. P.A.J. v Oort on *Key weather extremes affecting potato production in the Netherlands* is a good example of what sustainable data archiving in a data repository looks like [https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset:55620](https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset%3A55620)

[[Back]](#LongtermstorageBack)