

REG RESEARCH DATA MANAGEMENT PLAN

Background

In response to the need for more adequate and transparent data management and data storage, Wageningen Graduate School (WGS) has developed a data management framework. This implies that since April 1st 2014 it is compulsory for each research group to have a Research Data Management Plan (RDMP) at the group level. This document contains these general agreements for data management in REG. When developing an individual RDMP for your project (whether a PhD, an MSc project or otherwise) you may refer to this general REG RDMP; only special conditions in your project need to be described specifically.

Principles

The principles behind the need for RDMPs are defined in The Wageningen Code of Conduct for Scientific Practice, chapter 2.3, page 9 (<https://www.wageningenur.nl/en/show/The-Wageningen-Code-of-Conduct-for-Scientific-Practice.htm>):

- Research must be replicable in order to verify its accuracy. The choice of research question, the research set-up, the choice of method and the reference to sources studied is accurately documented.
- The quality of data collection, data input, data storage and data processing is guarded closely. All steps taken must be properly reported and their execution must be properly monitored (lab journals, progress reports, documentation of arrangements and decisions, etc.).
- Raw research data are archived in such a way that they can be consulted at a minimum expense of time and effort.
- Raw research data are stored for at least ten years from the date of publication.

For whom?

The REG RDMP is applicable to any research project done with one the staff member of REG as project leader or supervisor that results in publication in the form of peer reviewed papers, reports, a PhD thesis or an MSc thesis.

- *PhD or project leaders*: As of 1 April 2014 the way data are stored and managed must be described at the start of each PhD project, and included as part of the research proposal (obligatory by graduate school). For PhDs a template for an individual RDMP is available (Appendix A). This template can also be used by other researchers, e.g. project leaders. The template also contains background information and many tips and examples on different aspects of the RDMP.
- *MSc students*: MSc students within REG must follow similar procedures as PhDs, i.e. formulating an individual RDMP as part of the research proposal. The MSc RDMP is a much simplified version of the PhD RDMP. Agreements on data management should be included in the MSc thesis agreement under point 7: "Arrangements on reporting". The following sentences are provided as an example for the thesis agreement: "Data produced in a REG project belong to both REG and the prime investigators involved. At completion of the thesis the research data have to be handed in to the supervisor in line with the REG guidelines for data files (see Guidelines in Appendix B of the REG RESEARCH DATA MANAGEMENT PLAN, which can be found on the website of REG)".

General agreements on RDM within the Resource Ecology Group (REG)

When developing your individual RDMP you can refer to the general agreements within REG as given in Appendix B. Any deviation from the REG consensus should be described according to specific project needs. Note that general REG agreements only apply to certain sections of the individual RDMP (Section: 3. Define data management role; 5. Short term storage; 7. Documentation and metadata; 8. Sharing and ownership; 9. Long-term storage). The remaining sections of Appendix B have to be described individually for all projects.

SECTION 3: Data management roles

→ *Role of researcher:*

- *Collects, processes and analyses 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 REG guidelines (after finishing the publication, thesis or completion of the project).*

→ *Role of supervisor/project leader:*

- *The supervisor or project leader checks the data files before long-term storage. The supervisor is responsible for the data quality, clarity and completeness according to REG guidelines.*

SECTION 5: Short-term storage

NB: Short-term storage refers to the storage of data during the project/thesis work, before the data is published or results are reported (e.g. in a thesis).

→ *The data will be saved on the drive of the researcher from which a daily backup is made (M-drive for MSc students, for staff and PhD students the drive that is indicated for daily backup). Too limited storage capacity for MSc students should be discussed with the REG supervisor.*

→ *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: (YYMMDD). In this way it facilitates sorting the files according to the version's date and previous versions are not overwritten. More useful tips on versioning management can be found in Appendix A.*

SECTION 7: Documentation and metadata

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.*

→ *Detailed guidelines for design, organization and documentation of data files have been developed within REG and it is strongly recommended to use those (Appendix B).*

SECTION 8: Sharing and ownership

NB: ownership is not so much about 'property' (to whom do the data belong). It is 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 (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.*

SECTION 9: Long-term storage

NB: The Code of Conduct for Scientific Practice requires that you retain your data for ten years after you have published your article/thesis and that you make it available upon request for verification purposes.

→ *After completion of the thesis/papers/project, data files will be submitted to the supervisor/project leader for long -term storage (at least 5 years).*

→ *Following the RDMP of REG:*

- *Data will be stored by the supervisor/project leader in a dedicated directory that is backed up on a daily basis.*
- *The directory contains one or more zip files with file names on the basis of which the content can be clearly identified (e.g. PhD chapter number, Paper author, year and journal acronym). Accompanying publications (paper, report, thesis) should be included in the zip file.*
- *After 5 years, it is up to the staff member to decide if longer storage at the same location is needed*
- *For PhD and MSc thesis work the daily supervisor(s) will check if the data files are uploaded in good order for long-term storage.*

It is up to the owners of the data to decide if they want to share data through interfaces such as DANS / EASY Science Data, 3TU Datacentrum, our own library (Metis/WaY) or scientific publishers. For some journals/funders it is compulsory. More information can be found here:

http://www.wageningenur.nl/en/Expertise-Services/Facilities/Library/Expertises/Write-cite/research_data.htm

Support and training for PhD students

WUR library provides support and training for data management:

- Training course (1 day) of data management for PhD students
- Information about data archiving in DANS or 3TU through WaY (Wageningen Yield)

Contact: Data.library@wur.nl

Appendices:

A. RDMP template PhD: <http://www.wageningenur.nl/en/Expertise-Services/Facilities/Library/Expertise/Write-cite/Research-data/Data-Management-Plans.htm>

B. Data management approach REG (given below)

Appendix B Data management approach REG

This Appendix describes the data management approach (DMA) of the Resource Ecology Group (REG), and provides support for developing individual data management plans (DMPs). The intention of this DMA is to facilitate data storage and access after research projects are finished, mainly for future research purposes. DMPs are aimed at individual research projects, where appropriate individual researchers are mentioned.

Data management roles

The thesis contract of BSc/MSc students should include a DMP which includes that the data belong to both the supervisor(s) and the student and that at completion of the thesis the research data have to be handed in to the supervisor in line with the REG guidelines for data files. BSc/MSc students have control over their data, after thesis completion the student provides the supervisor a folder including all relevant data with a readme description (nowadays practice in REG is that students should hand over their data before getting a grade). The supervisor is responsible for data management after the student finished the thesis. PhD students and Postdocs have control over their data and are responsible, but need to determine in the DMP which supervisor will share responsibility after they have left. Staff members need to hand over the responsibility for data availability to REG when they leave/retire.

Type of research data

Raw data: These are data that are free of errors, but corrections should include an explanation.

Processed data: It should be determined during the project to what extent processed data are interesting to be stored, also considering investments to derive these from the raw data.

Model code: Model name and version should be provided. If the model was developed within REG the model should be properly documented and model code should be stored. Within WageningenUR a repository is available to store software/model code.

Other data: All data or documentation of the research process required to understand the research project and data should be stored.

Amount of data and storage options

Manually obtained data will be in the order of megabytes, generated data, pictures or videos may reach gigabyte range, high-throughput methods will produce terabyte range. Preferably data will be stored after publication on [3TU](#) or [DANS EASY](#). For data not suitable for 3TU or similar a solution needs to be found (see also documentation and metadata below). Short term storage is the responsibility of the prime investigator, long term storage is the responsibility of REG.

Sharing and ownership/custodianship

Data produced in a REG project belong to both REG and the prime investigators involved. This should be explicitly mentioned in a DMP. In a DMP a period will be determined after which data can be used by others than the prime investigators. In the case of BSc/MSc students aspects of sharing and ownership will be determined by the supervisor.

Documentation and metadata

Raw data are stored within an REG-directed environment (this can be a dedicated directory that is backed up daily, or disk space in the cloud, depending on data size requirements), together with a readme.txt. Data underlying publications will also be stored in [3TU](#) or [DANS EASY](#) environments if these data are not already linked to the article by the publisher (Be aware that the continuation of such a repository of the publisher is unclear).

Guidelines for a REG-DMP, based on the [WGS format](#)

All DMPs should mention that for all aspects of liability and ownership, all regulations of REG are applicable. After a DMP has been developed, it can be updated when needed.

1. Organizational context

| | |
|---|---|
| Name | Person responsible for the DMP of this project, this can be a BSc/MSc/PhD/Postdoc/Staff depending on the type of project. |
| Date | |
| Chair group | REG |
| Graduate school | PE&RC |
| Supervisor/ (co-)promotors /research group | If applicable supervisors can be mentioned, or members of the project's research group. |
| Start date of project | |
| File name of this DMP | A shortened Project title or fancy acronym suffices. |

2. Research project description

| | |
|-----------------|--|
| Title | |
| Abstract | Context of the project related to the DMP including project funding. |

3. Define data management roles

| | |
|------------------------------------|---|
| Roles | |
| Who is collecting the data? | |
| Who is analysing the data? | |
| Other | <p>Who has control over the data (you as a researcher, are there other colleagues involved, role of supervisor)? Is there a person in the research group with a specific responsibility for data management? Do other persons contribute, for example by writing code?</p> <p>Who is responsible for your data after you have left?</p> |

4. Type of research data

| Data stage | Specification of type of research data | Software choice | Data size/ growth |
|-------------------|---|--|--|
| Raw data | Observed/measured from experiments/model in&output/images/interviews... | After the end of the project data will have to be stored in a common format. Think about the types of software to be used. | Estimate in terms of mega-, giga-, or terabytes. |
| Processed data | Include calibration equations if applicable. | | |
| Models/code | Model name, code versions | | |
| Other? | Think of lab journals, project progress meetings etc. | | |

5. Short term storage solutions

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

6. Structuring data and information

An unambiguous directory and file-naming system enables other researchers including yourself in the future to find back information. Use descriptive names for files and indicate versions. A folder structure helps your data collection process, and can be implemented in various working environments (Sharepoint/ATLAS/electronic lab journal/etc.). Depending on the workflow, version control happens automatically or not, and you need to think how to keep track of various versions.

7. Documentation and metadata

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. When the project is ended, data are stored within an REG-directed environment (this can be a dedicated directory that is backed up daily, or disk space in the cloud, depending on data size requirements), together with a readme.txt. Data underlying publications will also be stored in 3TU or DANS EASY environments.

8. Sharing and ownership

May others be interested to re-use your data, and do you have plans to share them? Are there specific funder's requirements to share your data? Are they under embargo (STW has a form of embargo for example)? In case other parties (outside REG or outside WU) are involved with the research, what are the agreements on how the data will be used and shared? In case of social data, is there a need for randomized trial setup or anonymity of research subjects?

9. Short vs. Long term storage

Research data should always be retained for the short term (think of papers and hand-written forms as well). Each DMP should contain a backup plan (medium/frequency). For remote field work, when an internet connection is available, files can be uploaded to the m-drive through <https://citrix.wur.nl>, this drive is backed up daily. The Code of Conduct for Scientific Practices requires that you retain your data for ten years after you published your article. Long term storage needs to be considered when:

i) data underlie publications, ii) long term storage is required by funders, iii) there are legal requirements, and iv) it is likely other may want to re-use data. In addition, if data are unique, or involve considerable time /money to reproduce long term storage needs to be considered.

Example for guidelines of SOQ

<http://www.wageningenur.nl/en/Expertise-Services/Chair-groups/Environmental-Sciences/Department-of-Soil-Quality/Education/MSc-Thesis-and-Internship.htm> (see Guidelines for data files)