Research Data Management (RDM) guidelines to accompany the PhD checklist

These guidelines and the accompanied checklist aim to aid PhD supervisors in advising their PhD candidates on how to manage their research data. The checklist and guidelines are divided in three phases (planning, doing and finishing your research) that make up the entire research cycle.

For this checklist the following resources were used:

- Barsky, E. (2017), Good enough research data management: a very brief guide.
- Website of the Wageningen Data Competence Center (WDCC)

Support

1. Contact the Data Desk for questions and DMP review: data@wur.nl
2. Website data management WUR: https://www.wur.nl/en/Value-Creation-Cooperation/WDCC/Data-Management-WDCC.htm
3. Website data policy WUR: https://www.wur.nl/en/Value-Creation-Cooperation/WDCC/Data-Management-WDCC/Data-policy.htm
4. WUR Library (information for PhD candidates): https://www.wur.nl/en/Library/PhDs.htm

Phase 1: Planning your research (first 6 months)

Data management plan

Within the first 6 months of starting their PhD the PhD candidate is required to write a Data Management Plan (DMP), preferably with input from the promoter and supervisor, and has this checked by Data Management Support within the WUR-Library (see 'Support' for contact info). This obliges the PhD candidate to think about how to store, organise, share and archive their PhD data correctly and in compliance with the WUR data policy. A DMP is a living document, meaning that it can be updated and edited continually (e.g. when a research sub-project is added during the PhD period and results in new data types). Be aware that there could be different requirements for (submitting) DMPs between the Graduate Schools and funders.

PhD candidates can follow a course in Research Data Management, which is offered four times a year via the Wageningen Graduate Schools and consists of three mornings in which topics such as DMPs, data organisation, safe and secure storage, data archiving and data management support are covered.

TIP: often a PhD candidate lacks knowledge about some aspects of data management practices and where to find information about it. Take the time to sit down together to go through the DMP and show where information can be found. Make the PhD candidate aware that writing a DMP is very valuable and should not be taken too lightly, as time spent planning research data management saves a lot of time in the future.

Phase 2: Doing your research (years 1-3)

Data organization

It is important to have and maintain a logical folder structure in which to store research data (and other documents related to the research) and avoid losing oversight and file redundancy. It is additionally important to employ a clear and consistent file naming practice and version control. The WDCC provides tips and examples on how to do this.

TIP: Check, before data is being collected, whether the folder structure and file naming practice proposed by your PhD candidate in his/her DMP is put into practice. There might already be templates within the

chair group for a folder structure, file naming structure and versioning system the PhD candidate can be provided with.

Data documentation and description
Before starting to collect data think about how to take notes about the data that is to be collected. There are several topics necessary to document for future understanding and reuse of your data. At the very least, several questions need to be answered. Furthermore, when publishing a dataset (see "*Data archiving and preservation*", below), your data documentation forms the basis of a README file and a methodology file, required by any reputable data repository.

TIP: Discuss with the PhD candidate what needs to be documented regarding the data (this can differ per experiment of course) and how these notes are to be kept (e.g. OneNote, lab journals that will be digitalised, E-lab journal etc.). Further, check (or let someone else check, e.g. a lab employee), especially in the beginning, whether research notes are complete. For the PhD candidate, it would be useful to start writing a README file and methodology file as soon as a specific (sub-)project starts. This ensures that data documentation is up to date.

Data storage and backup
The data collected during a PhD project is highly valuable and, as such, should be stored during your research in such a way that it is safe, secure with no risk of data loss, leaks or corruption. The WUR has a several storage options that comply with the WUR regulations and are fully managed (i.e. installation, maintenance and support) by the IT department. Note: never store or share (see "*Data archiving, sharing and preservation*" below) sensitive data (see next paragraph) on a commercial cloud service (e.g. Dropbox, Google Drive, WeTransfer).

Research data can thus be either public, internal, confidential or even secret. For these different categories different storage options apply. To help decide what type(s) of research data are collected and how to store them, there is a decision aid, which follows the WUR's research data policy. Not only is it important to store research data safely, but also preferably in open formats, which enables and enhances exchanging data with peers (in the future).

Further, in more detail, keep your raw data raw by saving your raw data read-only in its original format in a separate folder. Make a working copy of your raw data (input data, used for processing or analysis). This version can be identical to the original version. In some cases it will be a modified version. For example, modifications required to allow your software to read the file or removing explanatory notes from a table. The original and working copy of a data file should be given different names (see "*Data organization*" above). The changes you make to your original data files should be described in a readme file (see "*Data documentation and description*", above). If the original data file has been obtained from others, keep the metadata of it in a separate folder.

TIP: Determine with the PhD candidate to what category or categories the research data belongs, the appropriate format(s) to save the data in, where to store it (in compliance with the WUR research data policy), how frequently the data is backed up and who is responsible for backup.

Phase 3: Finishing your research (years 3-4)

Data archiving, sharing and preservation
There are several ways to publish and thus archive your data set, but archiving your final data set in a repository has several important advantages;

- data kept long-term (WUR policy requires at least 10 years),
- there is a licence applied, acknowledging data rights. For more information about copyright and licences visit the Copyright Information Point (CIP),

- a persistent identifier (e.g. DOI, accession number, hdl.handle.net) is assigned to your data set, which enables findability and proper citing,
- your data is, separately from your publication, promoted to other users(especially when archived in a disciplinary repository, see below).

WUR Library is the front-office of the certified data repositories DANS-EASY and 4TU.ResearchData for which Data Management Support (WUR Library) can assist in publishing and registering your dataset there. Additionally, costs for archiving data in these repositories is covered by the WUR-Library. Of course, not all data types 'belong' with DANS-EASY or 4TU.ResearchData, but are more suitable when archived in e.g. a more discipline specific repository (e.g. genome data in NCBI). Be aware to always choose a repository where a persistent identifier is assigned to your dataset (to make it findable) and determine which licence applies, because this may vary between repositories, but also funders my require you to choose a specific licence. The WUR has approved several repositories, by assessing them on certain criteria (e.g. secure and durable storage). For how to share data (collected by WUR or third parties), have a look at the WUR Data Sharing Guidelines. Make sure to register the data in Pure (or ask the Data Desk to do this for you).

For sharing data with (project)partners use OneDrive for Business, Team sites or Microsoft Teams
For sending files use SURFfilesender

TIP: Submit the dataset(s) underlying the PhD candidate's first publication together, and pay due attention to the type of repository, assignment of a persistent identifier, and licence type.