WUR DMP Example Fictitious Project v03

Authors: Danny de Koning – van Nieuwamerongen
https://orcid.org/0000-0002-8264-4541

Irene Verhagen
https://orcid.org/0000-0001-5588-1333

Date: 20240502

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

DMP template: Example based on the WUR DMP template version 09

DMP example fictitious project:

- Use this example of a data management plan (DMP) in the WUR template to get ideas, inspiration, and insights in how to fill in a DMP for your own project.
- Some of the answers given apply only to the fictitious project. Make sure that if you copy-paste sections from this DMP, that they specifically apply to your workflow and project.
- Question? Contact data@wur.nl.

A. Describe the research project [info]

1. Name researcher:

Aulus Agerius

2. What is the name of your department(s)?

- Animal Sciences
- Social Sciences

3. What is the name of your chair group(s) or business unit(s)?
Please copy-paste the English name and abbreviation for:
- chair groups from this page.
- business units from this page (expand to Wageningen Research and keep expanding to find your specific division / group).
Example: Bioprocess Engineering (BPE) or Green Economy and Landuse (GEL).

- Adaptation Physiology Group (ADP)
- Cell Biology and Immunology Group (CBI)
- Business Economics Group (BEC)

4. Describe the organisational context of your research project.

DMP version (or date last modified)	20230404
Supervisor/(co-)promotor(s) (WU only)	Important Person
Graduate school (WU only)	WIAS
Start date of project	20220707
End date of project	20260707
Project number	123456789
Funding body	WUR Library

5. Give a short description of your research project.

Title: Effects of housing and dietary conditions on health and welfare in growing pigs and the perception of farmers.

Summary: This entire study description is fake and made up within 5 minutes, so take the study aims with a grain of salt. Health and welfare in pigs is an increasingly important subject within Dutch society. Although the subject of health and welfare in commercial pig husbandry has been researched on various occasions, results of these studies are not widely applied. We aim to study the effects of several housing conditions (barren housing vs various types of enriched housing) and dietary conditions (feeding levels, and dietary protein and sugar contents) on the health and welfare of growing pigs (assessed by behaviour, lesions, antibody titres, hormones, body weight). Additionally, we aim to assess the perception of farmers on pig health and welfare, what they consider poor health and welfare, and what they would take from the results of these studies (through recorded interviews and online surveys).

6. List the individuals and / or parties responsible for the following data management tasks.

Data collection	Aulus Agerius (Postdoc) John Doe (research assistant) Jane Doe (research assistant) Jaynie Everywoman (lab assistant) Ray Public (PhD candidate)
Data quality	Aulus Agerius (Postdoc) Ray Public (PhD candidate)
Storage and backup	Aulus Agerius (Postdoc) Ray Public (PhD candidate)
Data archiving/publishing	Aulus Agerius (Postdoc) Ray Public (PhD candidate) Numerius Negidius (Professor ADP) Nomen Nescio (Professor CBI) Yamada Hanako (Professor BEC)
Data stewardship/support	Petey Awesomedatasteward (chair-group data steward). Lisa Awesomecoordinatingdatasteward (department data steward coordinator). WUR Library RDM support (data@wur.nl).

7. I have requested a review of this data management plan from:

   - WUR Library – Research Data Management Support (data@wur.nl)
   - The (coordinating) data steward of my chair group / business unit

8. Name of data management support staff and / or data steward consulted during the preparation of this plan and date of consultation.

Dr ir Danny de Koning-van Nieuwamerongen
WUR Library – Research Data Management Support
data@wur.nl
Date: 20240425

Petey Awesomedatasteward
Chair-group data steward
Date: 20240225

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

9. Will you use existing data for this project?

- Yes. Please specify below which data (e.g. DOI, URL, or storage location) and the terms of use (e.g. licence).

We will be using pre-existing unpublished data from our previous study focusing on health and welfare in commercial pig husbandry entitled 'Assessing Commercial Pig Husbandry Health and Welfare' (adding DOI + licence when published). All creators of these data agreed on data reuse for this project. The data from that study consist of:

- Lesion scores in pigs related to housing conditions.
- Video recording of pigs related to housing conditions.

Additionally, we will be using data published openly in 4TU.ResearchData from researchers at Utrecht University (UU) Faculty of Veterinary medicine on behaviour in relation to dietary conditions (https://www.doi.org/12fakedoi3456/notreal; CC-BY license).

10. Will new data be produced?

- Yes.

Because not all required data is readily at hand, we will be producing new data.

- 12-hour video recordings of pigs per pen (10 pens) twice a week for 20 weeks.
- Behavioural observations (scan sampling) scored from the video recordings.
- Behavioural observations scored live with scan sampling 2 days a week for 20 weeks.
- Body lesion scoring once a week for 20 weeks.
- Blood samples collected at 10, 15, and 20 weeks of age, and analysed for antibody titers and serotonin levels.
- Body weight of the animals every 2 weeks.
- Interview recordings with farmers on perception of results in relation to pig health and welfare (will be transcribed).
- Surveys sent to farmers with questions on societal, government, and industry perceptions on health and welfare.
- Processing and analysis scripts.
- Figures to visualise results.

- Processed datasets (we will keep raw and processed data separated to ensure that the data can always be traced back to its original form).
- Informed consents.
- Documentation and metadata.
- Labjournals.

11. Please describe the data you expect to generate and / or use in the table below. Include reused existing data as well (as these are files that you manage and store).

File contents	Data type	Software	Format	Size per file	Nr of files	Max size
Existing data lesion scores	Tabular	Excel	.csv	50 MB	1	50 MB
Existing data behavioural video recordings	Video	N/A	.avi	5-10 GB	20	200 GB
Video recordings	Video	N/A	.avi	5-10 GB	400	4 TB
Scan sampling video	Tabular	Noldus	.txt	1-10 MB	1-10	100 MB
Scan Sampling live	Tabular	N/A	paper	N/A	40-80	N/A
Scan Sampling transcribed	Tabular	Excel	.csv	1-10 MB	2	20 MB
Body Lesions	Tabular	N/A	paper	N/A	4-80	N/A
Body Lesions transcribed	Tabular	Excel	.csv	1-10 MB	2	20 MB
Antibody titers	Tabular	ELISA reader	.csv	1-10 MB	3	30 MB
Antibody titers	Image (optical density)	ELISA reader	.tiff	5-15 MB	20	300 MB
Serotonin levels	Tabular	Fluorescence reader	.txt	1-5 MB	1-10	50 MB
Body weight	Tabular	N/A	paper	N/A	1-10	N/A
Body weight transcribed	Tabular	Excel	.csv	1-5 MB	2	10 MB
Interview recordings	Audio	N/A	.mp4	100-200 MB	20	4 GB
Interview transcriptions	Text	Word	.docx	1 - 5 MB	40	200 MB
Survey structure	textual	LimeSurvey	.lss	1 - 5 MB	1	5 MB
Survey archive	Textual and tabular	LimeSurvey	.lsa	5 – 10 MB	1	10 MB
Survey data	Tabular	LimeSurvey	.csv	1-5 MB	1	5 MB
Processed data	Tabular	Excel	.csv	1-5 MB	10-20	1 GB

Processing and analysis scripts	Textual	R and R studio	.R	0-1 MB	5-10	10 MB
Figures	Images	R (exports)	.jpg	1-10 MB	5-10	100 MB
Informed consent	Textual	NA	paper	NA	20	NA
Image informed consent	Image	Scanner	.jpg	1 – 5 MB	20	100 MB
Readme files	Textual	Notepad++	.txt	1 – 5 MB	5 - 10	50 MB
Metadata files	Textual	Yoda metadata editor	.json	1 – 5 MB	1 – 5	25 MB
Codebook	Textual / Tabular	Excel	.csv	1 – 5 MB	1	5 MB
Lab journal	Textual	NA	paper	NA	1	NA
Lab journal transcribed	Textual	Word	.docx	1 – 5 MB	1	5 MB
Pig stable journal	Textual	NA	paper	NA	1	NA
Pig stable transcribed	Textual	Word	.docx	1 – 5 MB	1	5 MB

12. Estimate how much data storage you require in total (e.g. by using the information in the table at question 11).

- >1000 GB

4000 - 4500 GB storage is required at max (summation of the max column of question 10).

C. Storage of data and data documentation / metadata during research [info]

13. Where will the data, code and accompanying documentation / metadata be stored and backed up **during** the research project (see the WUR Data Storage Finder)? Include platforms you use to share data, collect data on, or send data to for processing or analysis.

- WUR SharePoint / Teams - only when an up to data version of the research data is also safely stored on the W:drive or Yoda@WUR.
- Yoda@WUR.
- W:drive Massive File Storage Disaster Recovery (WUR network drive).
- W:drive Enterprise File Storage (WUR network drive).
- Git@WUR (GitLab locally hosted at WUR).
- Other, please specify below the storage medium / system and describe back-up frequency, access management, and geographic location (e.g. within or outside the EU).

We will store all research data and documentation initially within Yoda@WUR. The data will be placed in the Yoda Vault at key moments within the research after appropriate Yoda metadata has been added at the folder level. Key moments in research are at least when raw data is collected, data is fully analysed, and at the end of the project. Data in the vault will permanently be available and represents a secure copy of the data at that point in time. Through Yoda, we can securely share data with our UU partners as well.

When required, non-research data (presentations, informed consents, contracts, etc) will be placed on the W: drive Enterprise File Storage (EFS). Once data is not required in high access (when experiments have been performed and published), data will be transferred to W: drive Massive File Storage with disaster recovery (MFSDR). EFS is more costly but offers high availability to the data while working with the data. MFSDR is less costly and suitable for long term storage, but has less quick data availability during data center disasters.

We will be using Git@WUR to share our processing and analysis scripts with project members to work together on these scripts. Once data will be published or archived, an export of the main branch will be added to the publication.

OneDrive and M:drive will not be used as these are WUR personal accounts that don't allow easy access to the data for the project members. MS Teams will only be used transiently to quickly share small data files within the project team at WUR where required, but is not used

as a long-term storage location. An up-to-date version of the data will always be in Yoda@WUR (and W:drive where required).

For characteristics of the storage media, see https://library.wur.nl/storagefinder/.

Video and audio recordings are first stored on external hardware (physical audio recorder for interviews and external hard drive for videos of behaviour). The data collected on external hardware is transferred to Yoda@WUR as soon as possible. Data collected on laboratory machines are transferred with external hard drives from the machine to the workspace of the researcher and transferred to Yoda@WUR. The time that data remains on external hardware will be limited as much as possible.

Data recorded on paper will be transcribed digitally the same day and stored in Yoda@WUR or the W:drive. Additionally, photos will be taken of the paper documents and stored as well. Paper documents containing the data will be archived for 10 years in a physical archive and afterwards destroyed (digital copies remain).

Photos of informed consent forms are taken and securely stored on the W:drive in a separate folder with limited accessibility. The physical informed consent forms are then destroyed immediately to limit redundant duplication of personal data.

Survey data resides on a separate LimeSurvey server managed by Utrecht University. Survey structures and the data are stored at Utrecht University, conform to their regulations for processing and handling sensitive data. Data is backed-up automatically and can be retrieved for up to 2 months. Data is also replicated to 2 data centers ensuring higher accessibility. Data is encrypted at rest and the data from the survey can only be accessed using MFA with SSO. After the surveys have been performed, the survey structures, archive, and tabular data are exported from LimeSurvey and stored in Yoda@WUR. Data is then removed from the LimeSurvey servers.

D. Structuring research data and information [info]

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

hewa = health and welfare
rec = recordings
wk = week
scansam = scan sampling
intvw = interviews
transcr = transcribed
bw = bodyweight
ab = antibodies
od = optical density
sero = serotonin
img = image
proc = processed
stat = statistical
fig = figures
UU = Utrecht University

- prj123456789_hewa

  - raw_data_hewa
    - pigs_rec_hewa_raw
      - pigs_rec_wk1_hewa_raw
      - pigs_rec_wk2_hewa_raw
      - etc….
    - scansam_pigs_rec_hewa_raw
      - scansam_pigs_rec_wk1_hewa_raw
      - scansam_pigs_rec_wk2_hewa_raw
      - etc….
    - scansam_pigs_live_hewa_raw
      - scansam_pigs_live_wk1_hewa_raw
      - scansam_pigs_live_wk2_hewa_raw
      - etc….
    - intvw_rec_hewa_raw
    - intvw_transcr_hewa_raw
    - bw_hewa_raw
      - bw_img_hewa_raw

        - bw_transcr_hewa_raw
    - survey_hewa_raw
        - survey_structure_hewa_raw
        - survey_archive_hewa_raw
        - survey_data_hewa_raw
    - ab_hewa_raw
        - ab_od_img_hewa_raw
        - ab_titer_hewa_raw
    - sero_hewa_raw
    - lesion_hewa_raw
        - lesion_img_hewa_raw
        - lesion_transcr_hewa_raw

- reused_data_hewa
    - UU_rec_hewa_reused
    - UU_lesion_hewa_reused

- scripts_hewa
    - proc_scripts_hewa
    - analysis_scripts_hewa

- proc_data_hewa
    - scansam_hewa_proc
        - scansam_rec_hewa_proc
        - scansam_live_hewa_proc
    - ab_hewa_proc
    - sero_hewa_proc
    - intvw_hewa_proc
    - survey_hewa_proc
    - bw_hewa_proc
    - lesion_hewa_proc

- results_hewa
    - stat_output_hewa
    - fig_hewa
    - tables_hewa

- journals_hewa
    - pig_stable_journal_hewa
    - lab_journal_hewa

Further subfolders may be created when desirable to retain a structural overview. For example, processing scripts might be feasible with 1 script per type of data, which does not require further subfolders (clearly distinguishable from filenames which script belongs to which data type). If, however, multiple scripts will be necessary to process a single data type, then subfolders per data type might be required for better organisation of files. The same holds true for analysis scripts. Most of the processed data will not need further subfolders as the processing per data type will lead to 1 or a couple of files of that data type as a clean and merged overall data file. The results folder may require further subfolders depending on the outcomes of the analyses. This will be determined at the stage of analysis.

Folder names or the structure may be slightly modified if the project requires this for better practicality. The project abbreviation is mentioned within the folder name so that the folder can immediately be identified when the folder is misplaced.

All data analyses are performed using scripts. As the data is relatively simple in nature, running scripts will not take long and, therefore, intermediate forms of the data during analyses are not stored (the scripts can easily recreate the intermediate forms).

Informed consent forms are stored on the W drive in a separate limited access folder of the chair group.

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

We will us a pre-defined structure where feasible:

[subject_specifics]_[projectname]_[date]_[version].[extension]

The file name contains the project abbreviation so that it is clear to which project a file belongs. The date will be supplied in the format yyyymmdd to ensure proper sorting on date (i.e., 20220707) and conform to the international standard for using dates.

Example:
abtiter_wk15_pig_hewa_20220707_v01.csv

When more elements are required in file names, abbreviations will be used to keep the file name at a recommended length of 30-40 characters to limit the length. When abbreviations are applied, these will be explained in the documentation.

For files that are generated automatically by machines, such as for video recordings, the filenames will be renamed using batch renaming software or cmd prompt (Windows) script where possible. When not possible, filenames are appropriately documented (as done with all filenames).

When no indication in the file is given concerning 'results', 'output', 'analysis', etc., then the version vRAW will indicate the raw unaltered version while a version number will indicate an altered version of the raw data. In all other cases, a filename will contain an identifier to help identify the purpose of the file more (as in 'results', 'output', 'analysis').

16. How will you distinguish between versions of files (multiple answers possible)?

- 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.

E. Data documentation and data quality [info]

17. Describe below what data documentation and metadata will accompany the data to help make the data findable, understandable, and reproducible.

- The WUR readme file template (see template at https://doi.org/10.5281/zenodo.7701727).
- The Yoda metadata form (see the public Yoda metadata editor at https://utrechtuniversity.github.io/yoda-portal/).
- The WUR codebook template (see template at https://doi.org/10.5281/zenodo.7701727).
- Elaborate documentation and notes within scripts / code.
- Other, please specify below.

As we are also doing laboratory work, a lab journal is maintained. Currently, this is a paper journal transcribed digitally. We are investigating whether eLabJournal can be used within our project. Additionally, journals within the pig stables are maintained on paper and transcribed digitally. These journals add in the documentation of the project and give insight in how data was attained.

Where required, the WUR readme templates will be expanded upon to provide enough information to increase reproducibility and understandability, and, therewith, the FAIRness of the data.

18. Describe what data and analysis quality controls will be used?

- We will perform preliminary (pilot) experiments to validate intended experimental methods.
- We will use standard and validated protocols where appropriate.
- We will use repeated measurements to validate results (e.g. duplicate or triplicate analysis, multiple observer agreement, measurements taken over time, etc.).
- Statistical model assumptions are adhered to and assessed (e.g. (residual) distribution analysis, outlier analysis, (accounting for) independence, homogeneity of variance, etc.).
- We will use a statistical power analysis before and after the experiment.
- We will consult statisticians.
- Supervisors or peers will review the data and results for any anomalies (e.g. unexpected inconsistencies, outliers, correct labelling of data and / or treatments, correct and consistent coding applied, etc.).

- We will use standardised coding and terms of data throughout all experiments so that data descriptions are equal throughout various datasets created.
- Other, please specify below.

For laboratory assessments, there always is a reference control present on each assay, which helps determine the quality of the data.

Two observers will independently collect data from a selection of video material and live observations to determine interobserver agreement, before actual data collection will take place. A low score would indicate that the behaviours scored are very dependent on the person making the observations, making the data less reliable. Improvements to the protocols will then be made to increase agreement and reliability of observed behaviours.

F. Working with sensitive data (personal data, ethics), data (rights)holder(s), sharing and access [info]

19. Who is the (rights)holder of the data (commonly known as the owner of the data)?

- WUR is the (rights)holder of the data.
- Other, please specify below.

UU remains the rightsholder of the reused data, but are freely usable and openly accessible.

20. What is the data classification for your project (for example as specified in SmartPIA) taking into account the (privacy) sensitivity of the data?

- Serious.

Data contains:

- Personal data of farmers.
- Faces of animal caretakers, students, and employees are visible on video recordings.
- Personal views are expressed concerning animal production and welfare.

21. Is this project registered in SmartPIA?

- Yes.

22. Please specify the (sensitive) data and privacy protection measures. Note that any measures undertaken should be consulted with the Information Security Officer (ISO) and Privacy Officer (PO).

- Access management to the data is either managed or approved by the project leader / supervisor of the project and contains clear documentation of who has access.
- Informed consents are present when information from humans is involved.
- Personal or other sensitive data will be removed when not required for verification of research.

- Personal and other sensitive data will not be made openly available and will at most be shared under formal agreements for which the ISO and PO are consulted.
- Personal and / or other sensitive data will be separately stored where possible to increase difficulty of linking data for those with unauthorised access to data.
- We will consult with the ISO and PO for appropriate measures to undertake.
- Other, please specify below.

In compliance with the General Data Protection Regulation (GDPR), participants will sign an informed consent form checked by the privacy officer of the science group before data collection commences, which states amongst others that we:

- provide information on what data is exactly collected.
- provide information on the intent to share and / or publish the data and the conditions for sharing.
- provide transparency about which information we will make available.
- provide information on the storage and archival period.
- provide transparency in the methods applied to reduce the risks of identification.
- the right to withdraw consent and collected data.

Anonymisation will be investigated, but as this is often difficult to attain, we doubt that this will yield non-identifying data for individual files. Only aggregated personal data will be made publicly available in which single point data is not available and individuals cannot be identified. If required, single-point data (for example a single transcribed interview) will be made openly available only when the privacy officer is satisfied about anonymity. Access to these data is carefully monitored by the project leader, Postdoc, and the PhD candidate. Only the project leader (primary contact), Postdoc (secondary contact when primary is unavailable), and chair-group holder (when others are not available) will be allowed to grant access when requested. Access will be removed when not required anymore. These types of data will not be made openly available in its raw form (pseudonymisation and anonymisation processes will be investigated).

Video recordings will only be made available for the parts in which humans are not visible. When humans are visible at a specific time section in which data happens to be collected, they will be blurred from the videos. Video recordings in the pig stable do not include sounds (incl. voices) as there is no microphone alongside the cameras.

Audio recordings of interviews will not be made available.

23. Are there other ethical issues that need to be taken into account which may include approval from ethical committees?

- I work with animals and will seek / have approval of the ethics committees involved (Animal Welfare Body (IvD), Animal Tests Committee (DEC), Central Animal Testing Committee (CCD)).
- I work with humans in a social sciences aspect or that doesn't fall under Medical Research Involving Human Subjects Act (WMO) and will seek / have approval of the ethics committees involved (WUR Research Ethics Committee).
- The data and results of my research could negatively impact nature, wildlife conservation efforts, humans, society, policy creation or decision making. Please explain below.

The current societal impact and views of commercial husbandry of livestock can reach some tense levels within The Netherlands. The current research project collects data on the views of farmers on various aspects (society, husbandry, government). These types of data may create a larger polarity between different parties. Hence, our handling, sharing, and publishing of data may contain ethical properties. These will be taken into account with the steps previously described. Publication of results will be discussed within our project on proper handling and ethics of the publication.

24. Will there be any intellectual property (IP) rights or alternative applications or routes to impact (such as commercial interests) associated with the data?

- No.

It is not expected that there will be any IP rights resulting from this study. If there are any rights that may result from this study, they are stipulated within the consortium agreement.

G. Data archiving and publishing [info]

25. Are there reasons to restrict access to the data or limit which data will be made publicly available?

- Privacy / GDPR.
- Ethics.

26. Describe what data from question 11 will be archived (e.g. WUR network drive / Yoda) and not published, for a minimum of 10 years? Include the exact name of the storage medium chosen (see WUR Data Storage Finder).

- Due to sensitivity of the data, we will need to archive (part of the) data underlying publications or reports. Please specify below which data and the chosen storage medium.

Data that cannot be made public, such as sensitive and personal data, will be archived in Yoda@WUR. Along with that archived data, a reference to the data publication (the data that can be made public, see next question) will be present (to avoid duplicate storage).

The data concerns interview recordings and associated transcripts; raw video recordings of animal behaviour that include persons such as animal caretakers, students, and researchers.

Journals from the pig stables will have names removed of the animal caretakers.

Informed consents are not made public.

Where possible, anonymisation / pseudonymisation efforts will be undertaken.

27. What data will be published and made available for reuse via a data repository?

- Data underlying publications or reports. Please specify below which data listed in question 11.
- Only the metadata is published in a data repository as the data are too sensitive to openly share.

The raw data that will be published openly concerns:

- tabular transcribed data on lesion scores.
- tabular data on behavioural scan sampling (both live and from video).
- tabular data on antibody titers.
- tabular data on serotonin levels.
- tabular transcribed data on body weight.

The processed data that will be published openly concerns:

- anonymised survey data.
- video recordings without humans in the picture or anonymised (blurred faces).

Other data that will be published openly concerns:

- R scripts (including an export of the master branch of our Git@WUR project repo).
- Documentation (including anonymised stable journals and lab journals).

Data that will not be published openly concerns:

- raw and processed data from interviews as anonymity cannot be guaranteed.

For all data openly published, it will be checked by all WUR project members that data does not contain any personal data. A privacy officer will be consulted.

For the data that cannot be published openly, only the metadata will be published while the actual data remain archived in Yoda@WUR. This data can be requested from the project leader after publication. A data request will be evaluated by the project leader on the reason for the request and the underlying research hypotheses for which the data will be used. When a request is approved, a formal data sharing agreement is signed by the requesting party. The data sharing agreement will be set up by the legal department of the Animal Sciences group.

The reused data from UU will not be published again as it already is published Open Access. A reference to the reused data (the DOI) will be included in the documentation / metadata.

Other processed data and output / results will not be included in the publication as these can be generated using the R scripts that are included in the publication.

28. When will the data be available for reuse, and for how long will the data be available?

- Data will be available for at least 10 years as soon as the article or report is published and not required for any other article publication.

- Publication of data not underlying an article or report will be considered at the end of the project.
- Other, please specify below.

Any data that does not fall under the aforementioned categories, but underlies the PhD thesis and will not be used for journal publications, will be published after the promotion of the PhD candidate.

29. Which data repository do you intend to use to make the (meta)data findable and accessible (see the WUR Repository Finder)?

- DANS Data Stations.
- Other, please specify below.

Sensitive data will remain archived in Yoda@WUR.

If the publication module of Yoda@WUR is active by the time we will consider publication, then we will use Yoda@WUR as our repository.

30. Which metadata standard will be used to describe the data during internal archiving and / or depositing in a data repository?

- Yoda metadata (DataCite metadata standard).
- Metadata standard from DANS Data Stations, 4TU.ResearchData and / or Zenodo (which often are the DublinCore or DataCite standard).

31. Which license / terms of use will be applied to the data?

- Open access (Creative Commons Attribution licence (CC BY); anyone can access and reuse with attribution).
- Restricted access (custom licence text or data sharing agreement is required, dictating restrictions of access and reuse). When a data sharing agreement is required, the Privacy Officer or Information Security Officer is consulted.

CC BY for all non-sensitive data as required by the funder and recommended by WUR to stimulate open science and reusability of the research data.

Restricted access for sensitive data. To allow restricted access we arranged:

- A durable shared email address managed by multiple people of our chair group to guarantee continuity of access.
- Received data requests will be evaluated by the project leader, chair and another professor (independent of the project) of the chair group, the data steward of the chair group, and a Privacy Officer in the case of personal data. All data requests will be handled using the general data request assessment procedure which includes evaluating:
    - appropriateness of the research question and hypotheses.
    - methodology intended to answer the research question.
    - the amount of data requested (is it in line with the research question / sufficient to answer the questions, what will the power of the analysis be).
    - whether the intent of the research question weighs up against possible discomfort of research participants (no invasive measurements were done in this project on humans, so less relevant here) and the sharing of personal data outside the project group.
    - whether the requesting party has received (personal) data of the research participants before and how that impacts the mosaic effect.
    - whether the request conforms to the informed consents of the research participants.
- The data requests will be evaluated 'blind'. This means that any contact information of the requesting party is removed before the request is sent to the evaluation committee. This reduces any possibility of any favouritism / discrimination in handling requests which is not allowed by EU law.
- If the request is approved, a data sharing agreement is sent by the data steward to the requesting party. The data sharing agreement stipulates that data remains confidential, cannot be shared outside of the requesting party, that the chair group receives any processed data and conclusions back, and that the requesting party is not allowed to store or archive the data after completion of their project (we will take care of proper archiving of processed data).
- Once a signed copy of the data sharing agreement is received, the data will be made available by the data steward by placing a copy of the data in a designated Yoda@WUR folder. The requesting party is given access to this folder.

H. Data management costs [info]

32. What resources (in time and / or money) will be dedicated to data management, data archiving or publication, and ensuring that data is reusable? Indicate as well how these costs will be covered.

- All costs for (long-term) data storage, data publication fees, and long-term access management to the data are covered by the research group / project.
- The PhD candidate and supervisor will spend at least 10% of their time on research data management to approach the FAIR principles as much as possible.

We will request extra funding from our funder to cover the data publication costs. In addition, our chair group will cover costs that are not fully met by the funders.