Plan Overview

A Data Management Plan created using DMPonline

Title: DMP_example_FictitiousProject_v02

Creator: WUR RDM support

Affiliation: Wageningen University and Research (Netherlands)

Template: Data Management Plan | Wageningen University and Research

Project abstract:

This entire study description is fake and made up within 5 minutes, 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 (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 titers, 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 away from the results of these studies (through recorded interviews and online surveys).

ID: 121594

Start date: 07-07-2022

End date: 07-07-2026

Last modified: 07-12-2023

Copyright information:

DMP_example_FictitiousProject_v02

A. Describe the research project

1. Name researcher (please, add your full name):

Danny de Koning-van Nieuwamerongen

2. What is the name of your department?

- Animal Sciences
- Social Sciences

3. What is the name of your chair group or business unit?
Please copy-paste the English name and abbreviation for:
- chair groups from this page.
- business units from this page.
Example: Bioprocess Engineering (BPE) or Contract Research Organization (CRO).

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-)promotors	Important Person
Graduate School (WU only)	RDMsupport_is_awesome
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, 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 (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 titers, 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 away from the results of these studies (through recorded interviews and online surveys).

6. List the individuals responsible for the following data management tasks.

Data collection	Aulus Agerius (Postdoc) John Doe (research assistant) Jane Doe (research assistant) Jaynie Everywoman (lab assistent) 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)
Data stewardship / support	Petey Awesomedatasteward (chair-group datasteward). Linda Awesomecoordinatingdatasteward (department datasteward coordinator). WUR Library RDM support (data@wur.nl (WUR data support)).

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

Dr ir Danny de Koning-van Nieuwamerongen
WUR Library – Data Management Support
data@wur.nl
Date: 20230404

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

8. Will you reuse 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 Helfare (adding DOI + licence when published). 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 with 4TU.ResearchData from researchers at Utrecht University Faculty of Veterinary medicine (https://www.doi.org/12fakedoi3456/notreal ; CC-BY license) on behaviour in relation to dietary conditions.

9. 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 twice a week for 20 weeks.
- Behavioural observations (scan sampling) scored from the video recordings.
- Behavioural observation 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 wk of age, and analysed for antibody titers and serotonin.
- 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 send to farmers with questions on societal, government, and industry perceptions on health and welfare.
- Processing and analysis scripts.
- Figures to visualize 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).

10. 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).

Please scroll to the right in dmp.wur.nl to view the entire table.
(Text continues below the table.)

Subject	Type	Software	Format	Nr files	Size of each file	Max size
Existing data lesion scores	Tabular	Excel	.csv	1	50 MB	50 MB
Existing data behavioural video recordings	Video	N/A	.avi	20	5-10 GB	200 GB
Video recordings	Video	N/A	.avi	50-100	5-10 GB	1 TB
Scan sampling video	Tabular	Noldus	.txt	1-10	1-10 MB	100 MB
Scan Sampling live	Tabular	N/A	paper	40-80	N/A	N/A
Scan Sampling transcribed	Tabular	Excel	.xlsx, .csv	2	1-10 MB	20 MB
Body Lesions	Tabular	N/A	paper	4-80	N/A	N/A
Body Lesions transcribed	Tabular	Excel	.xlsx, .csv	2	1-10 MB	20 MB
Antibody titers	Tabular	ELISA reader	.csv	3	1-10 MB	30 MB
Antibody titers	Image (optical density)	ELISA reader	.tiff	20	5-15 MB	300 MB
Serotonin levels	Tabular	Fluorescence reader	.txt	1-10	1-5 MB	50 MB
Body weight	Tabular	N/A	paper	1-10	N/A	N/A
Body weight transcribed	Tabular	Excel	.xlsx, .csv	2	1-5 MB	10 MB
Interview recordings	Audio	N/A	.mp4	20	100-200 MB	4 GB
Interview transcriptions	Text	Word	.docx, .txt	40	1 - 5 MB	200 MB
Survey	Tabular	LimeSurvey	.csv	1	1-5 MB	5 MB
Processed data	Tabular	Excel	.xlsx, .csv	5-10	1-5 MB	50 MB
Processing and analysis scripts	Textual	R	.R	5-10	0-1 MB	10 MB
Figures	Images	R (exports)	.jpg	5-10	1-10 MB	100 MB

During working with the data we will make use of non-preferred formats and software required to facilitate working on the data (e.g., Excel to work with tabular data in .xlsx format). When archiving the data and / or publishing the data, all file formats will be converted to their preferred format equivalent to meet the FAIR principles requirements. We will consult https://dans.knaw.nl/en/file-formats/ for the equivalent preferred formats. When conversion is not feasible (even with a simple copy paste), we will make sure that the contents of the files are thoroughly described.

11. Estimate how much data storage you require in total (e.g. by using the information in the table at question 10).

- >1000 GB

1000 - 1500 GB storage is required at max (summation of the max column of question 10).

C. Storage of data and data documentation / metadata during research

12. Where will the data 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.

- W:drive Enterprise Files Storage (WUR network drive).
- W:drive Massive File Storage Disaster Recovery (WUR network drive).
- Yoda (data management platform; SURF hosted WUR instance).
- WUR SharePoint / Teams - only when an up to data version of the research data is also safely stored on the W:drive or Yoda.
- 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 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 analyzed, 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 will be placed on the W drive Enterprise Files 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 a lesser quick availability to the data.

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 master 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 data 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 (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 or the W:drive as soon as possible. Data collected on laboratory machines are transferred with external hardware from the machine to the workspace of the researcher and transferred to Yoda or the W:drive. The time that data remains on external hardware will be limited as much as possible.

Data recorded on paper will be transcribed the same day digitally and stored in Yoda or the W:drive. Additionally, photos will be taken of the paper documents and stored as well.

D. Structuring your data and information

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

(Text continues below the representation)

HeWa = health and welfare

- HeWa_prj123456
  - HeWa_raw_data
    - HeWa_video_observations
    - HeWa_scansampling
    - HeWa_interview_recordings
    - HeWA_interview_transcriptions
    - HeWa_bodyweight
    - HeWa_surveys
    - HeWa_antibodies
    - HeWa_serotonin
    - HeWa_lesions
  - HeWA_reused_data
    - HeWa_UU_videorecordings
    - HeWa_UU_lesions
  - HeWa_scripts
    - HeWa_processing_scripts
    - HeWa_analysis_scripts
  - HeWa_processed_data
    - HeWa_behavioural
    - HeWa_laboratory
    - HeWa_interviews
    - HeWA_surveys
    - HeWa_bodyweight
    - HeWa_lesions
  - HeWa_results
    - HeWa_stat_output
    - HeWa_figures
    - HeWa_tables

Further sub-folders may be created when desirable to retain a structural overview. 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.

14. Describe the file naming conventions you intend to use.

We will us a pre-defined structure where feasible:

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

The date will be supplied in the format yyyymmdd to ensure proper sorting on date (i.e., 20220707) and conform the international standard for using dates.

The version numbering will be supplied in a 'v' followed by 2 numbers (even below version 10), the first a so-called 'leading zero', to ensure proper sorting on version (i.e. v01, v02, v09, v10, v11).

Example:
HeWa_ELISA_15wk_pig_20220707_v01.csv

When more elements are required in file names, abbreviations will be used to keep the file name at a suggested length of 30-35 characters to limit the length. When abbreviations are applied, these will be explained in the readme file (see question 13).

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).

Within file naming we take into account that the filename needs an indication of where it is stored so that no question remains where the file should be located once it is accidentally misplaced. For example, when multiple projects exist all using some form of behavioural video recordings and a file is accidentally misplaced, it should be immediately visible that the file is at a wrong location. Example: pig_behaviour_day1.avi and pig_behaviour_day2.avi versus projectA_pig_behaviour_day1.avi and projectB_pig_behaviour_day1.avi (in the latter part it is immediately visible to which project one of the files belongs to).

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

- Dates within file names are updated when files are modified.
- A version number in the format 'v01' will be added to all file names which increases after file modification.
- 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

16. 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 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 / codes.

17. 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.
- Supervisors or peers will review the data and results for any anomalies (e.g. unexpected inconsistencies, outliers, correct labeling 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.

If possible, we will use the same vocabulary from discipline specific metadata standards (to be decided).
For laboratory assessments, there always is a reference control present on each assay, which helps determine quality of the data.

F. Working with sensitive data (personal data, ethics), data ownership, sharing and access

18. 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.

The reused data published from UU colleagues remain under UU ownership but are freely usable and openly accessible.
Any other regulations made are stipulated within the consortium agreement and signed between UU and WUR respecting UU and WUR policies.

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

- Serious.

20. Is this project registered in SmartPIA?

- Yes.

21. 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).

- Only WUR provided storage, processing, and analysis platforms are used as consulted with the ISO and PO where applicable.
- 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 are 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.

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.

Anonymization 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 (pseudonymization and anonymization 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. As common in animal commercial husbandry measurements, sounds (incl. voices) are not recorded (there is no microphone alongside the camera's).

22. 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 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. Any publication of results will thoroughly be discussed within our project on proper handling and ethics of the publication.

23. Will there be any intellectual property (IP) rights 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

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

- Privacy / GDPR.
- Ethics.

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

- Due to sensitivity of data we will need to archive (part of the) data underlying publications or reports internally. 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. Where possible, anonymization / pseudonymization efforts will be undertaken.

26. 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 10.
- 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:
- anonymized survey data
- video recordings without humans in the picture or anonymized (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

That 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. 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 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 with the legal department of the Animal Sciences group.
The reused data from Utrecht University will not be published again as it already is published with open access. A reference to the reused data (the doi) will be included in the documentation / metadata.
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.

27. 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.

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

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

- DANS-EASY.
- Other, please specify below.

Sensitive data will remain in Yoda.
If the publication module of Yoda is active by the time we will consider publication, then we will use Yoda as our repository.

29. 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-EASY, 4TU.ResearchData and / or Zenodo (which often are the DublinCore or DataCite standard).

30. Which [licence/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 open data.
Restricted for sensitive data.
See question 26.

31. If software is generated in this project, describe your publishing strategy below.

We are not producing any analysis software. Any processing and analysis scripts on Git@Wur will be exported and included in the data publication through the aforementioned repositories.

H. Data management costs

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 10 year data storage and access management to that data after journal publication or report are covered by the research group / project.
- Costs relating to publications up to 100 GB in DANS-EASY, 4TU.ResearchData, Zenodo are covered by WUR Library.
- Costs relating to publications over 100 GB 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.