

---

## Plan Overview

*A Data Management Plan created using DMPonline*

**Title:** Understanding the ammonia emission and its underlying mechanisms of the broiler industry in the Netherlands

**Creator:** Shutong Dong

**Affiliation:** Wageningen University and Research (Netherlands)

**Template:** Data Management Plan | Wageningen University and Research

### Project abstract:

The Netherlands witnessed its transition from conventional farming to intensive and high-tech farming and has one of the biggest livestock industries in Europe nowadays. As more and more attention is paid to animal welfare, the Dutch broiler industry has decided to switch from fast-growing broilers to slow-growing ones, which however, will lead to more ammonia emissions at the same time. To manage this trade-off, it is therefore important to understand the emission mechanisms and how the emission is affected by the factors involved in Dutch broiler production. The research objectives include (1) validation of calculating ventilation rate using the CO<sub>2</sub> mass balance method for slow-growing broilers under representative conditions; (2) determining emission factors of the in-house and the covered outdoor run for slow-growing broilers; (3) understanding underlying emission mechanisms by setting up interventions of potential variables concerning ammonia emission; (4) determining the bio-physical reactions of the litter concerning ammonia emission and the connection between litter characterization and ammonia emission.

**ID:** 120057

**Start date:** 01-08-2022

**End date:** 31-07-2026

**Last modified:** 16-03-2023

### Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customise it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

# Understanding the ammonia emission and its underlying mechanisms of the broiler industry in the Netherlands

---

## A. Describe the research project

### 1. Describe the organisational context of your research project.

Name researcher	Shutong Dong
DMP version (or date last modified)	20230312
Chair group/Business unit	Farm Technology Group
Graduate school (WU only)	PE&RC
Supervisor/(co-)promotor(s)	Peter Groot Koerkamp; Albert Winkel
Start date of project	20220801
End date of project	20260731
Project number	4400003141
Funding body	Wageningen Livestock Research

### 2. Give a short description of your research project.

Title	Understanding the ammonia emission and its underlying mechanisms of the broiler industry in the Netherlands
Summary	<p>The Netherlands witnessed its transition from conventional farming to intensive and high-tech farming and has one of the biggest livestock industries in Europe nowadays. As more and more attention is paid to animal welfare, the Dutch broiler industry has decided to switch from fast-growing broilers to slow-growing ones, which however, will lead to more ammonia emissions at the same time. To manage this trade-off, it is therefore important to understand the emission mechanisms and how the emission is affected by the factors involved in Dutch broiler production.</p> <p>The research objectives include (1) validation of calculating ventilation rate using the CO<sub>2</sub> mass balance method for slow-growing broilers under representative conditions; (2) determining emission factors of the in-house and the covered outdoor run for slow-growing broilers; (3) understanding underlying emission mechanisms by setting up interventions of potential variables concerning ammonia emission; (4) determining the bio-physical reactions of the litter concerning ammonia emission and the connection between litter characterization and ammonia emission.</p>

### 3. List the individual(s) responsible for the following data management tasks.

Data collection	Shutong Dong (PhD candidate) Peter Groot Koerkamp (promotor) Albert Winkel (co-promotor)
Data quality	Shutong Dong (PhD candidate) Peter Groot Koerkamp (promotor) Albert Winkel (co-promotor)
Storage and backup	Shutong Dong (PhD candidate) Peter Groot Koerkamp (promotor) Albert Winkel (co-promotor)
Data archiving/publishing	Shutong Dong (PhD candidate) Peter Groot Koerkamp (promotor) Albert Winkel (co-promotor)
Data stewardship/support	Shutong Dong (PhD candidate) Peter Groot Koerkamp (promotor) Albert Winkel (co-promotor) WUR Library RDM support (data@wur.nl (WUR data support))
Any other role [...]	

**4. Name of data management support staff consulted during the preparation of this plan and date of consultation.**

Not applicable

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

**5. Will you reuse existing data for this project?**

- Yes. Please specify below which data (e.g. DOI/url) and the terms of use (e.g. licence).

I will use the pre-existing unpublished data from our previous study focusing on reduction of ammonia emission of the laying hen manure using microorganism products; and data from a study in Geel, Belgium, about ammonia emission.

**6. Will new data be produced?**

- Yes

**7. When producing new data, describe the data you expect in terms of type, software and format.**

Data type (e.g. numerical, video, etc.)	Software (e.g. Excel)	(Open) file format (e.g. csv)
Textual	Word	.docx
Numerical	Excel	.xlsx
Images	Image viewer	.png/.jpg/.jpeg
Statistical	R	.csv

**8. Estimate how much data storage you require in total.**

- 100-1000 GB

## C. Storage of data and data documentation during research

### 9. Where will the data and accompanying documentation/metadata be stored and backed up during the research project?

- Sharepoint/Teams (WUR collaborative platforms)
- OneDrive for Business (WUR cloud storage)
- W:drive (WUR network drive)

Data will be stored at W:\ASG\WLR\_VenO\Medewerkers\winke022. This storage media is managed by WUR (datacenters at WUR) in which data is automatically backed-up, has secure access management (only accessible to specific added WUR credentials), folder access can be set per folder, integrity checks are performed, has disaster recovery (when one datacenter fails), and is encrypted at rest.

We will be using Git@WUR (see description of Git@WUR at <https://library.wur.nl/storagefinder/>) to share our processing and analysis scripts with project members to work together on these scripts. This storage media is managed by WUR (datacenters at WUR) in which data is automatically backed-up, has secure access management (only accessible to WUR credentials or those given specific access), and has disaster recovery (when one datacenter fails).

## D. Structuring your data and information

### 10. Give a representation of the folder structure you intend to use, or the link.

- Regio\_Deal\_Food\_Valley
- Validation\_CO2\_prj123456
  - manuscripts
- drafts
- submitted
  - raw\_data
- CO2\_production\_data1
  - CO2\_production\_data2
- processed\_data
- CO2\_production\_data1
  - CO2\_production\_data2
- results
- figures
- models
- tables

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

### 11. Describe the file naming conventions you intend to use.

We will use 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: Regio\_Deal\_Food\_Valley Validation\_CO2\_CO2production\_20220707\_v01.xlsx

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.

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).

## **12. Describe the file versioning system you intend to use.**

Raw files are designated with the word RAW at the end of the filename. Any other files that result from modifications, merging, processing of the raw files, will be designated with the letter v followed by 2 numbers which increase every time a new version of the data results from the modifications. Additionally, the date will be added for minor modifications which will be useful for modifications that do not elicit a new major version. Example:

Validation\_CO2\_CO2production\_20220707\_RAW  
Validation\_CO2\_CO2production\_20220712\_v01  
Validation\_CO2\_CO2production\_20220731\_v01  
Validation\_CO2\_CO2production\_20230115\_v02

For scripts and code within Git@WUR, versioning is an integral part of the system. Files with the exact same name can be compared between different branches and commits (updates to files made). It is therefore essential in Git to not change filenames between modifications as you will lose the inherent compare and version check. When files will be published, the files will be exported and added to the data to be published, appended with the version v01 if it is the first publication, and with reference to the specific master branch commit id.

## **E. Data documentation and data quality**

### **13. Describe below what data documentation and metadata will accompany the data.**

In order to produce FAIR data (data that can be re-used by others and be fully understood), plenty of documentation needs to be provided for all files collected, stored, archived, and published.

The metadata (at least data set title, creator(s) + affiliation (s), contributor(s) + affiliation(s), short description of the data set, keywords, licence etc.) will be documented through the use of either the datacite metadata scheme <https://schema.datacite.org/meta/kernel-4.4/> or Dublin Core <https://www.dublincore.org/specifications/dublin-core/dces/> (see question 25).

As recommended by <https://www.wur.eu/rdm>, a readme.txt file will be supplemented to the archived and / or published data files which will include information on the:

- folder structure.
- files present and their relations.
- purpose of each files.
- file formats present.
- purpose of the research.
- explanation of all used abbreviations within file and folder names as well as within files.
- an explanation of all columns used (if any).
- description/explanation of measurement units.
- category descriptions when results are categorized.
- software requirements (including name, version, company).
- machine requirements (including name, version, company).
- steps undertaken in processing data (going beyond just what is present in materials and methods of a journal article which often does not go into detail of what was done).
- and any other information required to understand and reproduce the data within the supplied folders.

In addition, a license will be provided indicating the term of use / license on the data (if any is required / available / possible). Finally, where required, ample description will be given within processing files (for example within Excel files) or scripts to indicate to the reader what steps have been undertaken (or the same description will be given in other accompanying files).

We intend to use a CC-BY license as described at <https://creativecommons.org/licenses/>.

When practically more desired or feasible, csv files that explain column names or the use of various other files as codebooks or descriptions will be supplied within used files.

#### **14. Describe what data quality controls will be used.**

We will ensure that the same type of data is referred to similarly between files (we will, for example, not allow column heading 'subjectnr' in one file and 'animalnr' in another file when they describe the same data; the same goes for coding observations). This ensures uniformity across data files within our research. If possible, we will use the same vocabulary from discipline specific metadata standards (to be decided).

In accordance with standard research practices, we will collect data according to pre-described validated protocols. Where such protocols do not exist, we will perform pilots to validate the collection methods.

In accordance with standard statistical analysis practices, we will check the distribution of data (raw and especially the residuals after statistical analyses) to determine validity of statistical models, outliers, or any other inconsistencies within the data (for example the wrong coding or missing values).

Laboratory analysts and research assistants will assist in checking the validity and quality of collected data where appropriate. This may include re-doing a random sample of observations to determine agreement with the original observations, or perform exploratory (descriptive) statistics. 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

### 15. Are there reasons (privacy, ethics, contractual agreement, commercial interests, public security, IP rights) to restrict access to the data or limit which data will be made publicly available?

- No

### 16. Will you process and/or store personal data during your research project?

- Yes. Please, specify below which measures you will take to ensure data protection and safeguard the privacy of the participants in your project.

By default, all projects at WUR that collect personal data will be screened on potential risks (by checking a checkbox on personal data usage when registering a project in MyProjects at WUR). The screening process is performed by the application SmartPIA and consists of a questionnaire that will be filled in by the project leader. Based on the outcome of this screening, a Data Protection Impact

Assessment (DPIA) may be carried out along with the privacy officer of our department as mentioned on <https://intranet.wur.nl/umbraco/en/about-wur/policy-regulations/privacy-personal-data/> at 'who can I approach if I have questions about privacy?'. This impact assessment describes the type of data collected, the storage media and its characteristics, the risks and impact of data loss / breach / theft, and the measures undertaken to mitigate the risks.

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.

Where possible, we will delete any personal data that is not further required for the research. In addition, where possible, we will pseudonymize the data (currently under investigation if possible). 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).

### 17. Is this project registered in SmartPIA?

- Yes.

As all projects at WUR need to be registered in MyProjects to release the funds of the project, we will check the checkbox on whether personal data is involved in the project. This will then send a notification to the SmartPIA application, which in turn sends a questionnaire to the project leader with some more in-depth questions on the type of data. This questionnaire will then calculate a data classification score and when sufficiently high, the privacy officer of the science group will be notified and create a DPIA along with the project leader and project members.

## 18. Are there other ethical issues that need to be taken into account?

- Yes. Please, explain.

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.

However, as we are not performing research on, for example, endangered species and their last final location for poachers to find them, we do not expect limitation in publishing results.

The current research does not need to pass a social ethical board. Any experiments on animals have passed the animal experimentation ethics committee.

## 19. Who has ownership and controls access over the data?

All research data collected at WUR and all data files produced are under WUR ownership (the employer has ownership of the data unless otherwise stipulated). 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 project members respecting UU and WUR policies.

The responsibility of granting access to all data (including sensitive / personal data) remains at WUR for the data collected within this project and of which ownership lies with WUR (specifically with the chair group/business unit involved) as previously specified.

## 20. 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

### 21. Do you have selection criteria, which allow you to determine which part of the data should be preserved once the project has ended?

- Yes. Please, elaborate below.

Only the data that underlies the journal publications, reports, and the dissertation of the PhD candidate will be preserved. At the end of the project we will discuss with our department whether the data collected but not underlying a publication needs to be archived as well (for possible future publications). In any case, we will preserve all raw files, processed files, transcripts, scripts, and analysis output belonging to the aforementioned publication types (see the folder structure previously described). Along with the archived data, plenty of documentation will be provided as described before. Personal data from participants will be destroyed after the retention period mentioned in the informed consent form.

Any code or data within Git@Wur will be exported and archived along with the data files.

Any other publicly available data (such as in reference articles or in journal articles) and the reused data (the data publication of UU as described in point 5) within our project will not be stored again as these are openly available. We will only store the references (persistent identifiers) to these materials and will clearly be stated in the accompanied documentation (readme files).

**22. What data will be archived internally (e.g. WUR network drive) for a minimum of 10 years?**

- Other. Please specify below.

Data that cannot be made public, such as sensitive and personal data, will be archived in the W-drive of 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).

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

- Other. Please, specify below.

All non-sensitive and non-personal data belonging to a journal publication will be made publicly available using the below specified repository (this includes the raw data including video observations when non-personal data is present). Data present on Git@Wur will be exported and added to the data publication (as it contains processing and analysis scripts).

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

- Data available after completion of the project (with embargo).
- Data available as soon as the article is published.

Data underlying a journal publication will be made available as soon as the journal publication is published and the data does not underlie another publication still needing to be published.

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.

**25. Which data repository do you intend to use to make the data findable and accessible?**

We will either use 4TU.ResearchData, Zenodo, or Yoda@WUR. The final solution will be discussed within the project (and depends on whether the publishing function of Yoda is available).

**26. Which metadata standard will be used to describe the data during archiving / depositing in a data repository?**

The minimum metadata standard used will be the DataCite metadata scheme 4.x as described at <https://schema.datacite.org/meta/kernel-4.4/>, which in turn is used by Yoda@WUR. The exact metadata terms used of the metadata schema is visible within Yoda or an example can be seen at: <https://www.uu.nl/en/research/yoda/guide-to-yoda/i-am-using-yoda/documenting-your-data>.

**27. Which licence/terms of use will be applied to the data?**

We will apply the CC-BY license (<https://creativecommons.org/licenses/>) to the data that can be published.

Sensitive and personal data will not be shared (closed access), or will be available for restricted access under a custom license (as specified to participants in the informed consent). The custom license will consist of a legal Data Transfer Agreement (DTA) that stipulates what can and cannot be done by the requesting party with the data. We have not decided yet to make use of a DTA as arranging long term access to restricted data relies on long-term involvement and availability of project members long after the project is finished (we may not be able to guarantee continued actively controlled and managed access to requested data as people may leave WUR and not have access to WUR storage anymore).

**28. If analysis 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

### 29. What resources (in time and/or money) will be dedicated to data management and ensuring that data is reusable?

The creation of FAIR data requires quite some extensive time. The PhD candidate will spend 10 - 20% of the time correctly handling and managing the data. The Postdoc will spend approximately 10% on data management. Currently, no financial costs are required to store the data during the project as that is covered by the chair-group. Publication of data may need extra funding as over 1 TB of data is available (for a large part the video recordings) needing 10 year storage. Yoda costs 270 euro / TB / Year which is not covered by WUR by default. We will discuss financial contribution within the project for publication of the data.

Any costs for archival on W drive of WUR may be minimal as long term storage can possibly be transferred to tape costing 12 euro / TB / year and will be covered by our chair-group (once tape storage becomes available within WUR). If tape storage is not available, the W-drive Massive File Storage Disaster Recovery will be used for 10 years and is also covered by our chair group.

### 30. If there are additional costs related to preparing the data for reuse, how will these costs be covered?

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.