
Development of a digital twin for pig growth and behaviour

A Data Management Plan created using DMPonline.be

Creator: Jens Slootmans

Affiliation: KU Leuven (KUL)

Funder: Fonds voor Wetenschappelijk Onderzoek - Research Foundation Flanders (FWO)

Template: FWO DMP (Flemish Standard DMP)

Grant number / URL: 1SHGF24N

ID: 203778

Start date: 01-11-2023

End date: 31-10-2027

Project abstract:

Increasing productivity has been the main focus of the pig farming sector, however often at the cost of pig welfare. Recent technological developments allowed for the rise of a range of precision livestock farming (PLF) technologies that can automatically monitor the pigs and assist the farmer. By combining state-of-the-art PLF technologies, data about both production (i.e. pig growth) and welfare (i.e. various pig behaviours) will be collected into an extensive dataset. Using mechanistic modelling techniques, a reference growth curve for each individual pig will be estimated, which allows to predict the weight at time of slaughter. Additionally, data-based dynamic modelling will be applied to create a time series model for the individual weight of each pig based on real-time data during the growing phase. By comparing the dynamic short-term weight prediction with the reference growth curve an early warning system for growth deviations will be developed. Based on the combined models, a monitoring and predictive digital twin (DT) of the individual pig will be created. To gain insight into the further development of a prescriptive DT for the pig sector, the application potential will be explored by consultation of pig farmers and veterinarians. Finally, a user interface is created to allow easy interaction between the pig farmer and the DT for pig growth and behaviour.

Last modified: 17-05-2024

Development of a digital twin for pig growth and behaviour

DPIA

DPIA

Have you performed a DPIA for the personal data processing activities for this project?

- Not applicable

Development of a digital twin for pig growth and behaviour

GDPR

GDPR

Have you registered personal data processing activities for this project?

- Not applicable

Development of a digital twin for pig growth and behaviour

Application DMP

Questionnaire

Describe the datatypes (surveys, sequences, manuscripts, objects ...) the research will collect and/or generate and /or (re)use. (use up to 700 characters)

Data will come from measurements performed in animals and the environment. For each individual animal, we will merge physiological, behavioural, and environmental data. We expect to generate more than 5 million samples. The experimental data will include numerical sensor data and pig information (.csv format), and videos (.mp4 format). The data manipulation and statistical analysis will be performed using R, Python or MATLAB and all created scripts will be saved.

Link to shared excel file with datatypes tabel: [Overview-datasets.xlsx](#)

Specify in which way the following provisions are in place in order to preserve the data during and at least 5 years after the end of the research? Motivate your answer. (use up to 700 characters)

The project promotor, professor Tomas Norton, will be responsible for the data management of project during and at least 10 years after the end of the research.

All files will be stored on hard drives. Data under processing will also be stored on the protected and personal network drive at KU Leuven. Furthermore, all raw data generated in the experiment will also be stored on the designated network drive at ILVO.

The network drives provide daily back-up to external servers and ensure long-term data preservation.

What's the reason why you wish to deviate from the principle of preservation of data and of the minimum preservation term of 5 years? (max. 700 characters)

This project does not deviate from the proposed principle. The data will be stored for beyond 5 years.

Are there issues concerning research data indicated in the ethics questionnaire of this application form? Which specific security measures do those data require? (use up to 700 characters)

Not applicable

Which other issues related to the data management are relevant to mention? (use up to 700 characters)

KU Leuven's data repository Lirias will be used for archiving the data made accessible by the project. KU Leuven's data policy demands retention of data for a period of at least 10 years after the end of the project. KU Leuven supports its researchers in the area of RDM (research data management) by the provision of a customized and free data management plan and a dedicated RDM support desk advising on data storage, metadata and preparing data for sharing. KU Leuven also continues to invest in affordable long-term storage and curation.

Development of a digital twin for pig growth and behaviour

FWO DMP (Flemish Standard DMP)

1. Research Data Summary

List and describe all datasets or research materials that you plan to generate/collect or reuse during your research project. For each dataset or data type (observational, experimental etc.), provide a short name & description (sufficient for yourself to know what data it is about), indicate whether the data are newly generated/collected or reused, digital or physical, also indicate the type of the data (the kind of content), its technical format (file extension), and an estimate of the upper limit of the volume of the data.

Link to shared excel file: [Overview-datatypes.xlsx](#)

If you reuse existing data, please specify the source, preferably by using a persistent identifier (e.g. DOI, Handle, URL etc.) per dataset or data type:

Historical data from ILVO, obtained from the same pig barn. Data is stored on ILVO internal network drives, but does not have a persistent identifier. Existing documentation about the ILVO pig barn will be used as the foundation for the documentation of the new datasets.

Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? Describe these issues in the comment section. Please refer to specific datasets or data types when appropriate.

- Yes, animal data

The newly generated sensor data and sample data is gathered under the supervision of the ethical committee of ILVO (Flanders Research Institute for Agriculture, Fisheries and Food). They approved the ethical committee application with the designated code 2023/435.

Will you process personal data? If so, briefly describe the kind of personal data you will use in the comment section. Please refer to specific datasets or data types when appropriate.

- No

No

Does your work have potential for commercial valorization (e.g. tech transfer, for example spin-offs, commercial exploitation, ...)? If so, please comment per dataset or data type where appropriate.

- Yes

No immediate commercial valorization is foreseen

Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material/Data transfer agreements/ research collaboration agreements)? If so, please explain in the comment section to what data they relate and what restrictions are in place.

- Yes

Some of the used AI models are or will be developed in collaboration with a 3rd party, restricting the possibility to share data for reuse (which applies on data used to train the AI models as well as the data resulting from applying the AI models)

Are there any other legal issues, such as intellectual property rights and ownership, to be managed related to the data you (re)use? If so, please explain in the comment section to what data they relate and which restrictions will be asserted.

- No

2. Documentation and Metadata

Clearly describe what approach will be followed to capture the accompanying information necessary to keep data understandable and usable, for yourself and others, now and in the future (e.g., in terms of documentation levels and types required, procedures used, Electronic Lab Notebooks, README.txt files, Codebook.tsv etc. where this information is recorded).

The data collected in this project is mainly sensor data. The metadata of the sensor data is stored in a documentation folder (sensor type, positioning of the sensor, maintenance of the sensor, calibration dates, variables measured by the sensor etc.).

Furthermore, there will be pig observation/scoring data. For this type of data, there will be readme.txt files explaining the setup of the observation and/or the organization of document (.xlsx files) structure. The .xlsx files will always contain a spreadsheet with explanation of the variables/parameters used in the other spreadsheets.

Code will annotated and version controlled in the host institution GitLab. Readme.txt files or comment/markdown in the code will explain the code and the use of the code

Will a metadata standard be used to make it easier to find and reuse the data? If so, please specify (where appropriate per dataset or data type) which metadata standard will be used. If not, please specify (where appropriate per dataset or data type) which metadata will be created to make the data easier to find and reuse.

- No

Currently the use of metadata standard is not anticipated.

See the previous question for more info about the metadata that will be created.

3. Data storage & back-up during the research project

Where will the data be stored?

The data will be stored on ILVO network servers and a KUL Mango repository.

How will the data be backed up?

Both the ILVO network servers and the KUL Mango repository have build in backup capabilities. By storing the data both at ILVO and at the KUL, the data is also saved twice.

Is there currently sufficient storage & backup capacity during the project? If yes, specify concisely.

If no or insufficient storage or backup capacities are available, then explain how this will be taken care of.

- No

There is currently not sufficient storage to store all video data (200 TB expected during the course of the project), however it is not the goal to store all video data. The video data will be analyzed by deep-learning computer vision models, of which only the results will be stored long term. For certain moments of interest, where high quality data is expected, video data will be saved as well.

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

Data is stored on institutional network servers that provide the required tools for granting access and read/write rights. Using these tools, the authorization of persons can be supervised in order to make sure the data is only accessed or modified by authorize persons.

What are the expected costs for data storage and backup during the research project? How will these costs be covered?

KUL Mango provides 1 TB of free storage space, which is sufficient for this project (since no raw video or images will be stored here)

ILVO networkstorage is also provided for free. The research unit previously invested in data storage solutions resulting in no additional cost for data storage for this project.

4. Data preservation after the end of the research project

Which data will be retained for at least five years (or longer, in agreement with other retention policies that are applicable) after the end of the project? In case some data cannot be preserved, clearly state the reasons for this (e.g. legal or contractual restrictions, storage/budget issues, institutional policies...).

All data, except for the raw video and image data, will be stored for at least five years.

Due to the large expected size of the video and image data (± 200 TB), it is not sustainable to store all the raw video and image data. Only videos and images of important moments (around treatment/sampling/disease outbreak) will be stored. Otherwise only the results of the computer vision models analyses will be stored after the end of the project.

Where will these data be archived (stored and curated for the long-term)?

The data will be archived on hard drives at the KUL and on the ILVO network servers.

Possibly the data will also be written on tape for long term data storage. The required facilities for writing tape are available at ILVO.

What are the expected costs for data preservation during the expected retention period? How will these costs be covered?

No additional costs are expected for the preservation of data.

5. Data sharing and reuse

Will the data (or part of the data) be made available for reuse after/during the project? In the comment section please explain per dataset or data type which data will be made available.

- No (closed access)

The data will not be made available for reuse after/during the project.

If access is restricted, please specify who will be able to access the data and under what conditions.

Not applicable

Are there any factors that restrict or prevent the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)? Please explain in the comment section per dataset or data type where appropriate.

- Yes, Other

Data sharing of part of the dataset is restricted by a collaborating company

Where will the data be made available? If already known, please provide a repository per dataset or data type.

The data will not be made available for reuse after/during the project.

When will the data be made available?

Not applicable

Which data usage licenses are you going to provide? If none, please explain why.

Currently, the specific data usage license is not yet determined

Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, you have the option to provide it in the comment section.

- Yes

KUL RDR repository datasets automatically get an identifier

What are the expected costs for data sharing? How will these costs be covered?

There are no expected costs

6. Responsibilities

Who will manage data documentation and metadata during the research project?

Jens Sloomans

Who will manage data storage and backup during the research project?

Jens Sloomans

Who will manage data preservation and sharing?

Tomas Norton and Jarissa Maselyne

Who will update and implement this DMP?

Jens Sloomans

*