
Plan Overview

A Data Management Plan created using DMPonline.be

Title: GenoBreed: Welfare-driven livestock breeding through advanced genomics

Creator: Roel Meyermans

Principal Investigator: n.n.

Data Manager: Roel Meyermans

Project Administrator: Roel Meyermans

Affiliation: KU Leuven (KUL)

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

Template: FWO DMP (Flemish Standard DMP)

Principal Investigator: n.n. n.n.

Data Manager: Roel Meyermans  <https://orcid.org/0000-0002-5079-5097>

Project abstract:

Our livestock sector, both in the EU and in Flanders, is currently undergoing an important transition towards sustainability and animal welfare. Sustainability from a climate and an ecological point of view asks for livestock that is converting (for humans non digestible) plant resources into high value animal proteins in the most efficient way. In this respect, enormous achievements have been realized, both in the optimization of management as well as in the genetic improvement of the animals. In pigs (Pietrain) as well as in beef cattle (Belgian Blue), Belgian breeds are globally amongst the most efficient and performative, but they still suffer from genetic diseases and disorders leading to an impaired animal welfare. In pigs, genetic disorders such as scrotal hernia, cryptorchidism and porcine congenital splay leg syndrome impose serious welfare issues, while Belgian Blue cattle are extremely sensitive to psoroptic mange. For each of these conditions, prior research of our consortium has identified putative loci and candidate genes resulting from transcriptomic analyses and genome-wide association studies. The proposed project aims to use cutting edge genomic sequencing techniques to identify causal variants and underlying mechanisms based on this prior research. Total RNA transcriptomics will be combined with deep whole genome sequencing and several validation steps. These identified causative mutations can be used in selection programs. However, the availability of more genetic tests and (genomic) information sources, confronts the breeders with a lack of tools to make an sustainable breeding decision combining all these results at once. Therefore, we will develop an innovative approach to properly weigh and utilize genetic test results, breeding values and genomic information in practical breeding strategies. Overall, this project aims to harmonize efficiency and welfare through a genetic and hence sustainable approach in our two major livestock breeds.

S020024N

Principal Investigator: Prof. Dr. Nadine Buys (KU Leuven)

ID: 210214

Start date: 01-10-2024

End date: 30-09-2028

Last modified: 07-01-2025

GenoBreed: Welfare-driven livestock breeding through advanced genomics

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)

The project builds on previous projects and will collect new data:

- Phenotype and pedigree records of sampled cattle and pigs (new & re-use) (1-20 Mb's) (.txt, .doc, .csv, ... formats)
- Biological samples for DNA/RNA analysis (blood & ear notch) (new & re-use) (frozen -20 /-80 °C)
- DNA / RNA samples (new & re-use) (frozen -20 /-80 °C)
- SNP genotypes (raw and imputed) (new & re-use) (1-50 GB's) (.ped/.map format)
- Whole genome sequences (new & re-use of reference samples) (1-2 TB's) (.fastq format)
- Transcriptome data (new & re-use) (1-50 GB's) (.fastq format)
- Selection index simulations (digital) (1-100 Mb's)
- Scientific publications and reports to the stakeholders

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)

1. Responsible person: Dr. Roel Meyermans with PI: Prof. Nadine Buys.
2. Storage capacity/repository
 - During research: KU Leuven Shared (J) and LVS drives.
 - After the end of the research: KU Leuven K drives (Archive) and LVS drive
 - Biological samples will be stored in the CABG bio databank (freezers)
 - Data linked to publications available for open access publication will be deposited on KU Leuven RDR.
 - Storage costs during and after the research are covered in the project budget

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)

Biological samples (blood / ear notches) and extracted DNA/RNA could be discarded earlier (limited place in biobank, RNA can perish), only if no longer deemed necessary for future projects and when genotypes / transcription profiles have been generated. Other data received under contractual restrictions (e.g. pedigree information from third parties) will be deleted if legally obliged.

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)

The data does not impose issues concerning human participants or data with dual use.

Animal sampling will be conducted following the procedures approved by the ethical committee of KU Leuven.

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

We do not envision additional issues related to data management at this point.

GenoBreed: Welfare-driven livestock breeding through advanced genomics

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.

				Only for digital data	Only for digital data	Only for digital data	Only for physical data
Dataset Name	Description	New or reused	Digital or Physical	Digital Data Type	Digital Data format	Digital data volume (MB/GB/TB)	Physical volume
		<i>Please choose from the following options:</i> <ul style="list-style-type: none"> • Generate new data • Reuse existing data 	<i>Please choose from the following options:</i> <ul style="list-style-type: none"> • Digital • Physical 	<i>Please choose from the following options:</i> <ul style="list-style-type: none"> • Observational • Experimental • Compiled/aggregated data • Simulation data • Software • Other • NA 	<i>Please choose from the following options:</i> <ul style="list-style-type: none"> • .por, .xml, .tab, .csv, .pdf, .txt, .rtf, .dwg, .gml, ... • NA 	<i>Please choose from the following options:</i> <ul style="list-style-type: none"> • <100MB • <1GB • <100GB • <1TB • <5TB • <10TB • <50TB • >50TB • NA 	NA
Publication	Scientific publications	New data	Digital / Physical	Other	.doc, .pdf	<100 MB	NA
Report	Reports for stakeholders in the project	New data	Digital / Physical	Other	.doc, .pdf	<100 MB	NA
Biological samples	Blood and/or ear notch samples from sampled cattle and pigs for DNA analysis	New data and reuse of existing data	Physical	Other	NA	NA	1000-2000 1.5 mL eppendorf tubes (or equivalent)
Biological samples for transcriptome research	Tissue samples from pigs for RNA analysis	New data and reuse of existing data	Physical	Other	NA	NA	500 1.5 mL eppendorf tubes
DNA / RNA	Extracted DNA and RNA for analysis from the biological samples	New data and reuse of existing data	Physical	Other	NA	NA	1000 micro tubes or equivalent
SNP genotypes	SNP genotypes (50K or equivalent)	New data and reuse of existing data	Digital	Other	.ped/.map, .bim/.fam/.bed, .vcf, .finalreport	<100GB	NA
Imputed SNP genotypes	Imputed SNP genotypes of the 50K SNP data	New Data	Digital	Other	.ped/.map, .bim/.fam/.bed, .vcf, .finalreport	<1TB	NA

Whole genome sequences	Newly generated whole genome sequences of pigs and cattle, plus already available at KU Leuven and ULiège (+ international repositories)	New data and reuse of existing data	Digital	Other	.fastq	<5TB	NA
Transcriptome data	Transcriptome analyses of sampled tissues of porcine	New data and reuse of existing data	Digital	Other	.fastq	<1TB	NA
Simulation data	Simulation data for selection index analyses	New data	Digital	Simulation data	To be determined	<1GB	NA
Phenotype data	Phenotype data of sampled cattle and pigs	New data and reuse of existing data	Digital / Physical	Observational	.csv, .txt	<100MB	NA
Pedigree data	Pedigree data provided by the herd-/pigbook associations for all sampled animals	New data and reuse of existing data	Digital	Other	.csv, .txt	<100MB	NA

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:

Cattle:

Phenotypes and SNP genotypes of previous project results on the same genetic disorder. Results published under <https://doi.org/10.1016/j.animal.2022.100460> and <https://doi.org/10.1186/s12711-024-00921-7>. All data are stored on the KU Leuven J (Shared) and K (Archive) drives of the CABG research group. WGS dataset of Belgian Blue cattle (owned by University of Liège, regulated under the collaboration agreement and MTAs between KU Leuven and University of Liège).

Pigs:

Phenotypes, SNP genotypes and Transcriptome results data from previous research, relevant for this project are stored on the KU Leuven J (Shared) and K (Archive) drives of the CABG research group. WGS dataset of Pietrain pigs, owned by CABG research group, not made publicly available. Stored on LVS drives of KU Leuven.

When relevant, data of the FAANG GeneSwitch project will be used: <https://data.faaang.org/projects/GENE-SWitCH>.

Selection index research: no existing datasets will be used apart from the ones previously described.

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

Ethical approval is applied for and is currently under review by the KU Leuven Ethical Committee.

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

During sampling, farmers are visited (address information) and can be contacted (phone). All participating farmers sign a informed consent form. After sampling these personal data are deleted.
Animal IDs will be fully anonymized for the analyses.

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

This project could potentially result in the identification of major genes that could be developed into a genetic test, in collaboration with the project partners. In this case, KU Leuven LRD and IP guidance will be contacted before dissemination of the results.

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

Cattle: MTAs between herdbook companies (CRV & awé/elevéo) and KU Leuven are currently being drafted by KU Leuven Research & Development (LRD). CRV and awé/elevéo are part of the advisory committee of the FWO SBO project.
Pigs: MTAs between pigbook company (VPF) and KU Leuven are currently being drafted by KU Leuven Research & Development (LRD). VPF is part of the advisory committee of the FWO SBO project.
Selection index research: None

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.

- Yes

MTAs are being drafted between KU Leuven and the suppliers of data (herdbook/pigbook associations).
Other potential issues are covered by the collaboration agreement between KU Leuven, UGent and ULiège.

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).

Daily analyses, changes and logs will be curated by the and frequently reported in written reports and meeting reports.
All finalized analyses will be concluded with a scientific publication, and metadata will be recorded in standardized README.txt files.

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.

- Yes

If possible a standard metadata file will be supplied. E.g. for WGS and SNP data using the full meta-data model (<https://www.kuleuven.be/rdm/en/rdr/full-metadata-model>) , README files will be created following <https://www.kuleuven.be/rdm/en/guidance/documentation-metadata/README>, and <https://libraries.ou.edu/content/how-make-readmetxt-file> and <https://data.research.cornell.edu/data-management/sharing/readme/>.

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

Where will the data be stored?

All data generated in the project will be stored on KU Leuven internal servers (J: (Shared) , K: (Archive) and L: (Large Volume Storage) drives). Before digitalization, records of sampling will be stored in physical folders and scanned as PDF during digitalization.

Biological samples (blood, ear notches) and extracted DNA/RNA will be stored in the freezers (-20°C / -80°C) of the CABG Bio Sample Database, if possible in different aliquots, stored over different freezers on different electric circuits.

How will the data be backed up?

Standard back-up provided by KU Leuven ICTS.

Biological samples: if possible, all samples will be aliquoted (3 or 4 aliquots) and divided over different freezers, all with temperature alarm signals. If possible, additional dry storage of blood samples will be performed (blood cards).

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.

- Yes

The estimated volume we expect to generate in this project is within lines of the available space on the KU Leuven storage servers. If insufficient, additional server space will be acquired (funds from research group).

Biological samples: Currently sufficient room in the different freezers. If space would become limited, additional freezer space will be acquired. This is overviewed by the responsible lab technician.

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

Access to the digital data is restricted to the team members of the research group CABG.

Biological samples are stored in a locked (freezer) room with restricted access only to authorized members of the research group CABG.

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

Server storage: for LVS € 95.14 / TB / year. Expected total (2Gb): 1€/year

for Shared: € 450.76 / TB / year. Expected total (2Gb): 1€/year

for Archive: € 450.76 / TB / year. Expected total (100Gb): 50€/year

In the project budget, a part of the consumables is allocated to computing and data storage.

Biological samples: If necessary, additional freezers will be acquired (e.g. also replacement of older devices if needed to avoid breakdown). Can be covered by the project budget, and if not, by the research group.

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...).

Digital records of phenotypes, simulation data, SNP genotyped (raw + imputed), WGS data and transcriptome data will be retained for at least five years after the project.

Data from third parties: pedigree data is provided by the respective herd/pig-books and is subjected to restrictions in time of use during the project, therefore this data will be deleted after the research according MTA (MTA is currently being drafted).

Biological samples (blood / ear notches) and extracted DNA/RNA could be discarded earlier (limited place in biobank, RNA can perish), only if no longer deemed necessary for future projects and when genotypes / transcription profiles have been generated.

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

KU Leuven K and LVS drives.

Data directly linked to open access publications will be stored and curated on KU Leuven RDR.

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

Costs are expected to be in range with 50-150 € year.

In the project budget, a part of the consumables is allocated to data storage. If insufficient for the long-term storage, additional costs will be covered by the CABG research group.

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.

- Yes, in a restricted access repository (after approval, institutional access only, ...)

Depending on the MTAs and the collaboration agreement to be signed between all partners and data providers.

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

As agreed in the MTAs (currently under development by LRD).

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, Intellectual Property Rights

MTAs between all associated parties are currently being drafted by LRD.

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

KU Leuven Research Data Repository ([KU Leuven RDR](#))

When will the data be made available?

Upon publication of the scientific manuscript for each part of the research.

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

CCBY4.0 if not further restricted in the MTAs

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

Not available yet.

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

These are currently not foreseen in the project budget, but if needed may be covered by the research group.

6. Responsibilities

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

Dr. Roel Meyermans

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

Dr. Roel Meyermans

Who will manage data preservation and sharing?

Dr. Roel Meyermans

Who will update and implement this DMP?

Dr. Roel Meyermans