## DMP title

**Project Name** FertiHealth (KU Leuven DMP)
**Project Identifier** C24E/22/007
**Grant Title** C24E/22/007
**Principal Investigator / Researcher** Ben Aernouts
**Project Data Contact** ben.aernouts@kuleuven.be
**Description** FertiHealth targets the smart integration of phenotypic and genotypic data that is available on dairy farms with the aim to create knowledge on the interaction between health, production performance and fertility and develop methodologies and algorithms to improve future monitoring and decision support systems.

### 1. Data Description

**What data will you collect or create? Fill out the table below and/or describe.**

| Type of data | Format | Volume | How created? |
|---|---|---|---|
| Farm management software backups (DeLaval DelPro) (historical – reuse) | .bak.zip | 350 GB | Daily/weekly automated backup from DelPro farm management software available on the research group's data server |
| DeLaval DelPro backup data (ongoing – generate) | .bak.zip | 100-200 GB | Daily/weekly automated backup from DelPro management software, accessed via LogMeIn remote software. |
| Milk recording data (cow ID, milk fat, protein, lactose, urea, somatic cell count, risk of ketosis and acidosis, standardized peak production, lactation value, milk production, expected milk production) | .txt/.csv | 10-20 GB | Obtained from the database of CRV (Cooperatie Rundvee Verbetering) |
| Insemination and pregnancy check-up details (cow ID, day, time, bull) | .txt/.csv | 10-20 GB | Obtained from the database of CRV (Cooperatie Rundvee Verbetering) |
| Genetic data (pedigree information), genetic and genomic breeding values and single nucleotide polymorphism (SNP) genotypes | .txt/.csv .bam | 10-20 GB | Obtained from the database of CRV (Cooperatie Rundvee Verbetering) |
| Treatment and disease registers | .xlsx | Max 1 GB | Export of administrative records of diseases and treatment generated by farmers and veterinarians in either a management software (CRV Dier/Digiklauw, UniformAgri, …) or a digitalized logbook. |
| Generated/processed data | .txt/.csv .png/.jpg | 100 GB | Data tables containing output of data processing algorithms generated during the project |
| Scripts for data processing and developed software algorithms | .py | 0.5 – 1 GB | Software code in python to analyse the abovementioned data. |
| Manuscripts, articles, presentations and visualisations | .docx/.pdf/.pptx | Max 1 GB | Peer-review articles, scientific presentations and further dissemination of the project by the researchers involved in the project. |

**Do you intend to reuse existing data?**

I will generate and collect new data and I will make use of existing data.

**Do you use personal data (i.e. all data possibly identifying an individual)?**

- No

No. The animal data will be anonymized so it cannot be linked to the farm and farmer.

## 2. Documentation and Metadata

**Describe the documentation that will be created for the data. This section deals with the way in which you will document how the dataset was created and subsequently processed.**

At the project level, an extensive metadata file will be assembled which describes in detail the setup and the algorithms used to obtain the data, the generated source code and a comprehensive description of the collected data. The algorithms themselves will not be described in this metadata file, but will be available and documented on a KU Leuven GitLab repository, allowing for full replicability of all results. Processed data and final code will also be deposited on a preprint server, such as Zenodo, to allow reproduction of results.

**Describe the metadata for the data. This section deals with metadata: information contained in your dataset about the research data.**

Initially, the metadata will be registered in .xml files following the DCMI metadata standard. If datasets are shared on open access databases, e.g. on Zenodo, the metadata will be converted to the metadata standard required by the respective open access database.

## 3. Ethical, Legal and Privacy Issues

**Are there any ethical issues concerning the creation and/or use of the data?**

No.

**Did you consider all issues about copyrights and IPR?**

Yes, the project proposal and phases have been formulated such to maximize applicability in global markets, by focusing on using data sources that are to a large extent readily available on farms. Farmers will be the end-users of the developed tools, and are owner of the data entered in them. Therefore, no barriers for valorisation of this project were identified. For the developed software, we will set the correct licenses and explore IP protection via patents where possible and confidentiality where needed in consultation with the technology transfer offices of KU Leuven.

**Are the collected data considered to be â€œdata containing personal informationâ€ and are all the requirements about the collection of these data met?**

Yes, a 3rd party agreement ‘Individual Agreement Data Collection and Processing’ was made with each of the dairy farmers that participate in this project. In this agreement, the farmers declare that they give permission to the authorized representatives of the Livestock Technology group (KU Leuven), for the collection of data in a period of maximum 5 years and up to no later than 31/12/2027. Additionally, they allow the use, processing and anonymization of the data within the framework of the FertiHealth project and further research in the Livestock Technology group after the data collection period. These contracts directly concern data derived from the farm management software (DeLaval Delpro) back-ups and treatment and disease registers (specified under §2). In addition, in this agreement, the farmers declare that they give permission to the authorized representative of the Livestock Technology group to collect their milk recording data, insemination and pregnancy check-up details and genetic and genomic data that are stored in the CRV database.

## 4. Data storage and Backup during Research

**How and where will the data be stored during research?**

- Centrally on storage facilities of the research unit
- In a cloud service offered by the university

1. All raw data is stored on a by the group managed Synology data storage unit running in RAID5 and backed-up weekly to the groups SharePoint (WET – Livestock Technology/Documents/LT Data Cloud Storage/).
2. The raw data are merged into a PostgreSQL database that is stored on a by the group managed server running Ubuntu and backed-up weekly to the Synology data storage unit (/Database/) and the groups SharePoint (WET – Livestock Technology/Documents/LT Data Cloud Storage/Database/).
3. Processed data & results are stored on the laptop of the researcher and synched to the researcher’s KU Leuven-hosted OneDrive share (e.g. '/Users/Dyan Meuwissen/OneDrive – KU Leuven/PhD Research/').
4. Software code is stored on the laptop of the researcher and pushed to the researcher’s KU Leuven-hosted Gitlab repository (e.g. 'https://gitlab.kuleuven.be/livestock-technology/dmeuwissen/ PhDResearch/).
5. Processed data and final code will also be deposited on a preprint server, such as Zenodo.

**Which back-up procedures are in place?**

1. The by the group managed Synology data storage unit is running in RAID5.
2. For the by the group managed server (database) a weekly backup is made to the Synology.
3. A weekly backup of the Synology is stored on the groups SharePoint (WET – Livestock Technology/Documents/LT Data Cloud Storage/).
4. The laptops of the researchers are synched with their KU Leuven-hosted OneDrive share (e.g. '/Users/Dyan Meuwissen/OneDrive – KU Leuven/PhD Research/').
5. The groups SharePoint and the researcher's OneDrive cloud storage are hosted by KU Leuven and thus underly the Universities back-up procedures.

The KU Leuven hosted services (SharePoint - scalable, OneDrive - 2 TB, GitLab - scalable) are sufficient for data needs. Further the server used to host the database (1 TB SSD + 2TB) and the connected Synology data storage unit (24 TB) offer sufficient storage to store and backup all raw and processed data.

**Describe the data security procedures and who has access to the data.**

Secure access to all KU Leuven hosted services (GitLab, OneDrive, SharePoint) is only given to authorized personal KU Leuven accounts. Both the server and Synology data storage unit are only accessible through the KU Leuven-hosted VPN and additionally secured by personal accounts/passwords. All data that will be shared or disseminated to third parties will be anonymized, as specified in the data agreements (specified under §3).

## 5. Data selection and Preservation after Research

**What is the long-term preservation plan for these dataset(s)?**

All data described under §1 will be retained until at least 5 years after the end of the project.

All data will be archived on the Synology data storage unit (RAID5) of the research group for at least 5 years, conform the KU Leuven RDM policy. This Synology is backed-up weekly to the group's SharePoint hosted by KU Leuven.

Given that the storage unit/server are self-hosted, the annual costs are mainly regarding the maintenance of these systems. These costs will be covered by follow-up projects.

**Data Selection: Which data will have long time value for the research and will be preserved?**

All raw data, processed data, scripts, reports, presentations and manuscripts.

## 6. Data Sharing

**Are there any restrictions for sharing the data?**

Yes, raw data from the farms needs to be anonymized before it can be shared. It can only be accessed, used and published under the requirements specified under the data agreements. Any additional use of the data that is not specified in these agreements or that does not comply to these agreements, needs to be authorised by the respective dairy farmers and CRV.

Following data will be made publicly available by the end of the project:

1. Preprints of articles containing anonymized research results will be available on a preprint server, such as Zenodo, upon submission.
2. Published articles containing anonymized research results will be made available by the publisher conform to their restrictions (open-access or not).
3. All source code is stored on KU Leuven GitLab and access is shared with supervisors and authorized colleagues. Scripts, functions and snips of code that are not valorizable will be made publicly available on the KU Leuven Gitlab by the end of the project.

(Un)processed data from and about the farms will not be made publicly available, as specified in the data agreements. To authorized personnel of Livestock Technology (KU Leuven) and other researchers/collaborators (KU Leuven or non-KU Leuven), the data will be available immediately after collection, when the data is stored on the Synology and integrated in the database on the local server.

**If there are no restrictions, which mechanisms will be in place to assure that the data are discoverable, accessible and intelligible?**

PostgreSQL database containing the merged and linked raw and processed data on the local server (backup on Synology). Metadata file on the database structure, tables and headers.

Logical folder structure for the raw data files.

Metadata file to explain the folder structure and the content of the raw data files.

**How will you share the data?**

- Repository
- Publication

1. The full dataset with documentation will not be publicly available, but can be made available for reuse by authorized personnel of Livestock Technology (KU Leuven) and other KU Leuven and non-KU Leuven researchers/collaborators if in agreement with the active data agreements or upon renewal of these agreements.
2. Preprints of articles containing anonymized research results will be available on a preprint server, such as Zenodo, upon submission.
3. Published articles containing anonymized research results will be made available by the publisher conform to their restrictions (open-access or not).
4. All source code is stored on KU Leuven GitLab and access is shared with supervisors and authorized colleagues. Scripts, functions and snips of code that are not valorizable will be made publicly available on the KU Leuven Gitlab after the end of the project.

**With whom will the data be shared?**

- On request with peers only

All members of Livestock Technology (KU Leuven) and other researchers/collaborators that comply with the data agreements will be able to access the raw and processed data as ‘authorized representatives’ of these research groups. All members of Livestock Technology (KU Leuven) will be able to access all software algorithms, metadata and anonymized research outputs that are generated within the project.

1. The full dataset with documentation will not be publicly available, but can be made available for reuse by authorized personnel of Livestock Technology (KU Leuven) and other KU Leuven and non-KU Leuven researchers/collaborators if in agreement with the active data agreements or upon renewal of these agreements.
2. Preprints of articles containing anonymized research results will be available on a preprint server, such as Zenodo, upon submission.
3. Published articles containing anonymized research results will be made available by the publisher conform to their restrictions (open-access or not).
4. All source code is stored on KU Leuven GitLab and access is shared with supervisors and authorized colleagues. Scripts, functions and snips of code that are not valorizable will be made publicly available on the KU Leuven Gitlab after the end of the project.

## 7. Responsabilities and Resources

**Who is responsible for Data Management during the project? This will be the person who might receive questions on the data management aspects of the research project.**

The PhD student (Dyan Meuwissen) will be responsible for the data data storage, backup, documentation, metadata and sharing during the project, but will cooperate closely with her promotor (prof. Ben Aernouts) to ensure the data will still be preserved and shared once the project is finished. After the end of the project, the promotor (prof. Ben Aernouts) is responsible for data preservation and sharing for 5 years.

**Which additional resources are needed for the execution of the Data Management Plan?**

None.

**Did you read the KU Leuven Data Management Policy? (find the link to the policy in the guidance).**

- Yes