

## DMP title

**Project Name** DMP, A self-learning computing pipeline for disease decision support in dairy cattle using ensemble learning and fuzzy logic - DMP title

**Project Identifier** 1SE0922N

**Principal Investigator / Researcher** Martin Julius Gote

**Project Data Contact** +32 495 196664, martin.gote@kuleuven.be

**Description** The main objective of this project is the development of a dairy cow decision support tool (DST) based on on-farm sensor data to automatically detect health problems, predict their causality and provide decision support for treatment. Therefore, sensor data from on-farm automatic milking systems and additional sensors will be collected to implement a monitoring tool that can produce alerts based on deviating production and health parameters.

**Institution** KU Leuven

### 1. General Information

#### Name applicant

Martin Julius Gote

#### FWO Project Number & Title

1SE0922N: A self-learning computing pipeline for disease decision support in dairy cattle using ensemble learning and fuzzy logic

#### Affiliation

- KU Leuven

### 2. Data description

#### Will you generate/collect new data and/or make use of existing data?

- Generate new data
- Reuse existing data

**Describe in detail the origin, type and format of the data (per dataset) and its (estimated) volume. This may be easiest in a table (see example) or as a data flow and per WP or objective of the project. If you reuse existing data, specify the source of these data. Distinguish data types (the kind of content) from data formats (the technical format).**

Type of data	Format	Volume	How created
DeLaval DelPro backup data (historical - reuse)	.bak.zip	350 GB	Daily/weekly automated backup from DelPro management software.
DeLaval DelPro backup data (ongoing - generate)	.bak.zip	100-200 GB	Daily/weekly automated backup from DelPro management software.
DeLaval DelPro data (extracted from running databases - generate)	.csv	50-100 GB	Daily data extraction files created by automated python pipeline.
Treatment and symptom registers	.xlsx	max 1 GB	Administrative records of diseases and treatments generated by farmers and veterinarians in either a management software or on paper.

Farmer feedback	.csv	max 1 GB	Manual feedback by farmers through a web application.
Database consisting preprocessed data	.db	10 GB	Extracting the for the project necessary data from the historical and ongoing backups.
Generated/processed data	.txt/.csv .png/.jpg	100 GB	Generated/processed data: data tables containing output of software algorithms generated during the project.
Software algorithms Source code other code	.py .html/.css/.js .sql	1 GB	Software code in SQL to access the historical sensor data in the .bak files and software code in Python to analyse all the above mentioned data. Software code to generate a web application to communicate with farmers, written in html/css/javascript.
Manuscripts, articles and presentations	.pdf/.tex .pdf/.pptx	max 1 GB	Peer-review articles, scientific presentations and further dissemination of the project by the researchers involved in the project.

### 3. Legal and ethical issues

**Will you use personal data? If so, shortly describe the kind of personal data you will use. Add the reference to your file in KU Leuven's Register of Data Processing for Research and Public Service Purposes (PRET application). Be aware that registering the fact that you process personal data is a legal obligation.**

- No

Privacy Registry Reference:

Short description of the kind of personal data that will be used:

**Are there any ethical issues concerning the creation and/or use of the data (e.g. experiments on humans or animals, dual use)? If so, add the reference to the formal approval by the relevant ethical review committee(s)**

- No

**Does your work possibly result in research data with potential for tech transfer and valorisation? Will IP restrictions be claimed for the data you created? If so, for what data and which restrictions will be asserted?**

- Yes

Together with the technology transfer offices of KU Leuven and in consultation with KU Leuven

IOF manager Dr. Bart De Ketelaere, we will evaluate the valorization potential of the different research data throughout the project. If needed, proper actions will be taken (filing patent, transferring software code into Dynamic Link Libraries, ...) to protect and anchor potential IP. This IP could already be licensed nonexclusively to commercial companies during the project so that it doesn't influence the research and development.

**Do existing 3rd party agreements restrict dissemination or exploitation of the data you (re)use? If so, to what data do they relate and what restrictions are in place?**

- Yes

A 3rd party agreement 'Individual Agreement Data Collection and Processing' was made with each of the dairy farmers that participate in the FWO project. In this agreement, the farmers declare that they give permission to the authorized representatives of Livestock Technology (KU Leuven), for the collection in a periode of maximum 5 years and up to no later than 31/12/2026, and the use, processing and anonymization of the data below about and from their company within the framework of the FWO project and further research in Livestock Technology (KU Leuven) after the data collection periode.

The mentioned data in this agreement concists of dataset numbers 1 to 5 (see question 2: data description).

#### **4. Documentation and metadata**

**What documentation will be provided to enable reuse of the data collected/generated in this project?**

At the project level, an extensive metadata file will be provided which describes in detail the pipeline for data collection and integration, the herefore generated source code and an extensive description of the collected data ('/LTUsers/mgote/FWO/project\_description.xml', '/LTUsers/mgote/FWO/project\_description.md', '/LTUsers/mgote/FWO/project\_description.csv').

The algorithms themselves will not be described in this metadata file, but will be available and documented on a KU Leuven GitLab repository ('https://gitlab.kuleuven.be/livestock-technology/mgote/fwo').

The generated relational database containing all preprocessed data will be accompanied with a metadata file describing the data, preprocessing steps and general information about the use of the database ('/LTUsers/mgote/FWO/database/database\_description.md').

All algorithms developed during the project will be documented and made available on a KU Leuven GitLab allowing for full replicability of all results.

**Will a metadata standard be used? If so, describe in detail which standard will be used. If no, state in detail which metadata will be created to make the data easy/easier to find and reuse.**

- Yes

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

#### **5. Data storage and backup during the FWO project**

**Where will the data be stored?**

1. All raw data is stored on a by the group managed Synology data storage unit (/farmBackups/).
2. Newly aquired raw data is initially transfered via the KU Leuven iRods system, where an additional archive will be created (/set/home/set\_pilot013/).
3. The database is stored on a by the group managed server running Ubuntu.
4. Processed data & results are either stored on the Synology data storage unit or a KU Leuven hosted OneDrive share ('/LTUsers/mgote/FWO/').
5. Software code will be synched to a KU Leuven hosted GitLab repository ('https://gitlab.kuleuven.be/livestock-technology/mgote/fwo').

**How is backup of the data provided?**

1. The by the group managed Synology data storage unit is running in Raid 5.
2. For the by the group managed server (database) frequently a backup is made to the Synology.
3. All other services are KU Leuven hosted and thus the underlies the automatic daily back-up

procedures on the university's central servers.

**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 KU Leuven hosted services (iRods - scalable, GitLab - scalable and OneDrive - 2 TB) are sufficient for data needs. Further the server used to host the database (500 GB) and the connected Synology data storage unit (24 TB) offer sufficient storage to store and backup all raw and processed data.

**What are the expected costs for data storage and back up during the project? How will these costs be covered?**

Given the current data-volume estimates, the freely available data storage facilities will suffice and therefore, data storage facilities won't add extra costs to the project. However, if the freely available storage facilities become limiting, the current data storage unit can be extended with up to 24 TB more storage space, which would be covered with the FWO benchfee.

**Data security: how will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?**

Secure access to all KU Leuven hosted services (iRods, GitLab, OneDrive) is only given to authorized personal KU Leuven accounts.

Both the by the group hosted server and data storage unit are only accessible through a from Thomas More (KU Leuven campus Geel) hosted VPN and additionally secured by personal accounts/passwords.

All data that will be shared or disseminated to third parties will be anonymized, if not specified other in a given contract.

## **6. Data preservation after the FWO project**

**Which data will be retained for the expected 5 year period after the end of the project? In case only a selection of the data can/will be preserved, clearly state the reasons for this (legal or contractual restrictions, physical preservation issues, ...).**

All data described under §2 will be retained until at least 5 years after the end of the project.

**Where will the data be archived (= stored for the longer term)?**

All data will be archived on our research groups server/data storage unit (Raid 5) for at least 5 years, conform the KU Leuven RDM policy ('/LTusers/mgote/FWO/').

**What are the expected costs for data preservation during the retention period of 5 years? How will the costs be covered?**

Data server costs are anticipated in and will be covered by other projects or the FWO bench fee. Given that the storage unit/server are self-hosted, the annual costs are mainly regarding the maintenance of these systems.

## **7. Data sharing and reuse**

**Are there any factors restricting or preventing the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)?**

- Yes. Specify:

Raw data from the farms (data types 1 to 5 in table 1, section 2: data description) needs to be anonymized at the end of the project and can only be accessed, used and published under the requirements specified under the 'Individual Agreement Data Collection and Processing' (3th party agreement with the involved dairy farmers, see section 3: legal and ethical issues). Any additional use of the data that is not specified in this agreement or that does not comply to this agreement, needs to be authorised by the involved dairy farmers.

**Which data will be made available after the end of the project?**

1. Preprints of articles containing anonymized research results will be available on 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 ('https://gitlab.kuleuven.be/livestock-technology/mgote/fwo'). Scripts and functions that are not valorizable, will be made publicly available on KU Leuven GitLab after the end of the project.

### **Where/how will the data be made available for reuse?**

- Other (specify):

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 upon and specified by renewal of the 'Individual Agreement Data Collection and Processing' (3th party agreement with the 15 involved dairy farmers).
2. Preprints of articles containing anonymized research results will be available on zenodo upon submission.
3. Published articles containing anonymized research results will be made available by the publishing authority conform to their restrictions (open-access or not).
4. All source code is stored on GitLab and access is shared with supervisors and authorized colleagues ('https://gitlab.kuleuven.be/livestock-technology/mgote/fwo'). Scripts and functions that are not valorizable, will be made publicly available on GitLab after the end of the project.

### **When will the data be made available?**

- After an embargo period. Specify the length of the embargo and why this is necessary

1. (Un)processed data from and about the farms will not be made publicly available, as specified in the 'Individual Agreement Data Collection and Processing'. 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 and backed up on the local server.
2. Preprints of articles will be available on zenodo upon submission.
3. Articles will be made available by the journal or conference upon publication of the anonymized research results under the restrictions of the publishing authority.
4. All source code is stored on GitLab and access is shared with supervisors and authorized colleagues ('https://gitlab.kuleuven.be/livestock-technology/mgote/fwo'). Scripts and functions that are not valorizable, will be made publicly available on GitLab after the end of the project.

### **Who will be able to access the data and under what conditions?**

All members of Livestock Technology (KU Leuven) and other researchers/collaborators that comply with the 3rd party agreement 'Individual Agreement Data Collection and Processing' will be able to access the raw and processed data (1-6) 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.

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

1. (Un)processed data from and about the farms are stored and backed up on the local server ('farmBackups/'). Costs for setting up and maintenance of this server will be covered by other projects of the research group which use the same datasets. In the case of maintenance costs, those will be taking from the provided FWO strategic basic bench fee.
2. Preprints of articles will be available on zenodo upon submission, free of charge.
3. Articles will be available upon publication of the anonymized research results. Publication costs will be covered by the FWO strategic basic bench fee.
4. All source code is stored on GitLab and access is shared with supervisors and authorized colleagues ('https://gitlab.kuleuven.be/livestock-technology/mgote/fwo'). Scripts and functions that are not valorizable, will be made publicly available on GitLab after the end of the project. This service is free of charge.
5. Data storage units hosted by KU Leuven (not by the group) are made available free of charge. In case of arising costs, these will be covered by the FWO strategic basic bench fee.

## **8. Responsibilities**

### **Who will be responsible for data documentation & metadata?**

The FWO PhD fellow (Martin Julius Gote) will be responsible for the data documentation & metadata.

### **Who will be responsible for data storage & back up during the project?**

The FWO PhD fellow (Martin Julius Gote) will be responsible for data storage & backup during the project.

### **Who will be responsible for ensuring data preservation and reuse ?**

The FWO PhD fellow (Martin Julius Gote) will be responsible for the data preservation and sharing during the project, but will cooperate closely with his promotor 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.

### **Who bears the end responsibility for updating & implementing this DMP?**

The FWO PhD fellow (Martin Julius Gote) bears the overall responsibility for updating & implementing this DMP.