

FWO DMP Template - Flemish Standard Data Management Plan

Project supervisors (from application round 2018 onwards) and fellows (from application round 2020 onwards) will, upon being awarded their project or fellowship, be invited to develop their answers to the data management related questions into a DMP. The FWO expects a **completed DMP no later than 6 months after the official start date** of the project or fellowship. The DMP should not be submitted to FWO but to the research co-ordination office of the host institute; FWO may request the DMP in a random check.

At the end of the project, the **final version of the DMP** has to be added to the final report of the project; this should be submitted to FWO by the supervisor-spokesperson through FWO's e-portal. This DMP may of course have been updated since its first version. The DMP is an element in the final evaluation of the project by the relevant expert panel. Both the DMP submitted within the first 6 months after the start date and the final DMP may use this template.

The DMP template used by the Research Foundation Flanders (FWO) corresponds with the Flemish Standard Data Management Plan. This Flemish Standard DMP was developed by the Flemish Research Data Network (FRDN) Task Force DMP which comprises representatives of all Flemish funders and research institutions. This is a standardized DMP template based on the previous FWO template that contains the core requirements for data management planning. To increase understanding and facilitate completion of the DMP, a standardized **glossary** of definitions and abbreviations is available via the following [link](#).

1. General Project Information

Name Grant Holder & ORCID	Dyan Meuwissen (0000-0002-4846-3595)
Contributor name(s) (+ ORCID) & roles	Ben Aernouts (0000-0001-6266-3019) Ines Adriaens (0000-0001-9768-2308)
Project number ¹ & title	1SB2923N – Bridging gaps: Integration of on-farm phenotypic and genotypic data to improve monitoring and decision support in dairy cattle – application on fertility.
Funder(s) GrantID ²	1SB2923N
Affiliation(s)	KU Leuven
Please provide a short project description	In this PhD project, I aim to develop novel methodologies and algorithms to improve monitoring and decision support tools for dairy cows by integrating on-farm available phenotypic and genotypic data. This research goal builds on the hypothesis that combining these data can reveal the complex interactions between health, production performance and fertility in dairy cattle. These insights will result in more accurate and robust predictive tools that help the farmer to improve animal health, performance and welfare.

¹ “Project number” refers to the institutional project number. This question is optional since not every institution has an internal project number different from the GrantID. Applicants can only provide one project number.

² Funder(s) GrantID refers to the number of the DMP at the funder(s), here one can specify multiple GrantIDs if multiple funding sources were used.

2. 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 data³.

Dataset Name	Description	New or Reused	Digital or Physical	ONLY FOR DIGITAL DATA	ONLY FOR DIGITAL DATA	ONLY FOR DIGITAL DATA	ONLY FOR PHYSICAL DATA
				Digital Data Type	Digital Data Format	Digital Data Volume (MB, GB, TB)	Physical Volume
Farm management software backups	Daily/weekly automated backup from DelPro farm management software available on the research group's data server	Reuse existing data	Digital	Software	.bak.zip	350 GB	
DeLaval DelPro backup data	Daily/weekly automated backup from DelPro management software, accessed via LogMeln remote software.	Ongoing – generated data	Digital	Software	.bak.zip	100-200 GB	
Genetic data	pedigree information, genomic	Historical data	Digital	Software	.txt/.csv/.bam	10-20 GB	

³ Add rows for each dataset you want to describe.

	breeding values and single nucleotide polymorphism (SNP) genotypes obtained from database of CRV (Cooperatie Rundvee Verbetering)						
Treatment and disease registers	Administrative records of diseases and treatment generated by farmers and veterinarians in either a management software or a logbook.	Historical and ongoing – generated data	Digital	Software and compiled data	.xlsx	Max 1 GB	
Generated/processed data	Data tables containing output of data processing algorithms generated during the project	Generated data	Digital	Software	.txt/.csv/.png/.jpg	100 GB	

Scripts for data processing and developed software algorithms	Software code in python to analyse the abovementioned data.	Generated	Digital	Software - computational	.py	0.5 – 1 GB	
Manuscripts, articles, presentations and visualisations	Peer-review articles, scientific presentations and further dissemination of the project by the researchers involved in the project.	Generated	Digital	Documentation	.pdf/.pptx	Max 1 GB	

<p>GUIDANCE:</p> <p>DATA CAN BE DIGITAL OR PHYSICAL (FOR EXAMPLE BIOBANK, BIOLOGICAL SAMPLES, ...). DATA TYPE: DATA ARE OFTEN GROUPED BY TYPE (OBSERVATIONAL, EXPERIMENTAL ETC.), FORMAT AND/OR COLLECTION/GENERATION METHOD.</p> <p>EXAMPLES OF DATA TYPES: OBSERVATIONAL (E.G. SURVEY RESULTS, SENSOR READINGS, SENSORY OBSERVATIONS); EXPERIMENTAL (E.G. MICROSCOPY, SPECTROSCOPY, CHROMATOGRAMS, GENE SEQUENCES); COMPILED/AGGREGATED DATA⁴ (E.G. TEXT & DATA MINING, DERIVED VARIABLES, 3D MODELLING); SIMULATION DATA (E.G. CLIMATE MODELS); SOFTWARE, ETC.</p> <p>EXAMPLES OF DATA FORMATS: TABULAR DATA (.POR, .SPSS, STRUCTURED TEXT OR MARK-UP FILE XML, .TAB, .CSV), TEXTUAL DATA (.RTF, .XML, .TXT), GEOSPATIAL DATA (.DWG, .GML, ..), IMAGE DATA, AUDIO DATA, VIDEO DATA, DOCUMENTATION & COMPUTATIONAL SCRIPT.</p> <p>DIGITAL DATA VOLUME: PLEASE ESTIMATE THE UPPER LIMIT OF THE VOLUME OF THE DATA PER DATASET OR DATA TYPE.</p> <p>PHYSICAL VOLUME: PLEASE ESTIMATE THE PHYSICAL VOLUME OF THE RESEARCH MATERIALS (FOR EXAMPLE THE NUMBER OF RELEVANT BIOLOGICAL SAMPLES THAT NEED TO BE STORED AND PRESERVED DURING THE PROJECT AND/OR AFTER).</p>	
<p>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.</p>	
<p>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, please describe these issues further and refer to specific datasets or data types when appropriate.</p>	<p>No ethical issues.</p>

⁴ These data are generated by combining multiple existing datasets.

Will you process personal data ⁵ ? If so, briefly describe the kind of personal data you will use. Please refer to specific datasets or data types when appropriate. If available, add the reference to your file in your host institution's privacy register.	No issues. The animal data will be anonymized so it cannot be linked to the farm and farmer.
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, 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.
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 to what data they relate and what restrictions are in place.	Yes, a 3 rd 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 FWO project and further research in the Livestock Technology group after the data collection period. This concerns data derived from the farm management software (DeLaval Delpro) back-ups and treatment and disease registers. In addition, in this agreement, the farmers declare that they give permission to the authorized representative of the Livestock Technology group to collect their genetic and genomic data that is stored in the CRV database.
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 to what data they relate and which restrictions will be asserted.	No

⁵ See Glossary Flemish Standard Data Management Plan

3. 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).

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.

Will a metadata standard be used to make it easier to **find and reuse the data**?

If so, please specify which metadata standard will be used. If not, please specify which metadata will be created to make the data easier to find and reuse.

REPOSITORIES COULD ASK TO DELIVER METADATA IN A CERTAIN FORMAT, WITH SPECIFIED ONTOLOGIES AND VOCABULARIES, I.E. STANDARD LISTS WITH UNIQUE IDENTIFIERS.

Yes, 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.

4. Data Storage & Back-up during the Research Project

Where will the data be stored?	<ol style="list-style-type: none"> 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.
<p>How will the data be backed up?</p> <p><i>WHAT STORAGE AND BACKUP PROCEDURES WILL BE IN PLACE TO PREVENT DATA LOSS? DESCRIBE THE LOCATIONS, STORAGE MEDIA AND PROCEDURES THAT WILL BE USED FOR STORING AND BACKING UP DIGITAL AND NON-DIGITAL DATA DURING RESEARCH.⁶</i></p> <p><i>REFER TO INSTITUTION-SPECIFIC POLICIES REGARDING BACKUP PROCEDURES WHEN APPROPRIATE.</i></p>	<ol style="list-style-type: none"> 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.

⁶ Source: Ghent University Generic DMP Evaluation Rubric: <https://osf.io/2z5g3/>

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 (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.
How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons? <i>CLEARLY DESCRIBE THE MEASURES (IN TERMS OF PHYSICAL SECURITY, NETWORK SECURITY, AND SECURITY OF COMPUTER SYSTEMS AND FILES) THAT WILL BE TAKEN TO ENSURE THAT STORED AND TRANSFERRED DATA ARE SAFE. ⁶</i>	Yes, 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. 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 §2).
What are the expected costs for data storage and backup during the research project? How will these costs be covered?	Given the current data-volume estimates, the data storage facilities that are currently available will suffice and therefore, data storage facilities won't add extra costs to the project.

5. 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 described under §2 will be retained until at least 5 years after the end of the project.
Where will these data be archived (stored and curated for the long-term)?	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.
What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	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.

6. Data Sharing and Reuse

<p>Will the data (or part of the data) be made available for reuse after/during the project? Please explain per dataset or data type which data will be made available.</p> <p><i>NOTE THAT 'AVAILABLE' DOES NOT NECESSARILY MEAN THAT THE DATA SET BECOMES OPENLY AVAILABLE, CONDITIONS FOR ACCESS AND USE MAY APPLY. AVAILABILITY IN THIS QUESTION THUS ENTAILS BOTH OPEN & RESTRICTED ACCESS. FOR MORE INFORMATION: HTTPS://WIKI.SURFNET.NL/DISPLAY/STANDARDS/INFO-EU-REPO/#INFOEUREPO-ACCESSRIGHTS</i></p>	<p>Yes, the data that will be made available is:</p> <ol style="list-style-type: none"> 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 after the end of the project.
<p>If access is restricted, please specify who will be able to access the data and under what conditions.</p>	<p>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.</p>
<p>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 per dataset or data type where appropriate.</p>	<p>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.</p>

<p>Where will the data be made available? If already known, please provide a repository per dataset or data type.</p>	<ol style="list-style-type: none"> 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.
<p>When will the data be made available?</p> <p><i>THIS COULD BE A SPECIFIC DATE (DD/MM/YYYY) OR AN INDICATION SUCH AS 'UPON PUBLICATION OF RESEARCH RESULTS'.</i></p>	<p>(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.</p>

<p>Which data usage licenses are you going to provide? If none, please explain why.</p> <p><i>A DATA USAGE LICENSE INDICATES WHETHER THE DATA CAN BE REUSED OR NOT AND UNDER WHAT CONDITIONS. IF NO LICENCE IS GRANTED, THE DATA ARE IN A GREY ZONE AND CANNOT BE LEGALLY REUSED. DO NOTE THAT YOU MAY ONLY RELEASE DATA UNDER A LICENCE CHOSEN BY YOURSELF IF IT DOES NOT ALREADY FALL UNDER ANOTHER LICENCE THAT MIGHT PROHIBIT THAT.</i></p> <p><i>EXAMPLE ANSWER: E.G. "DATA FROM THE PROJECT THAT CAN BE SHARED WILL BE MADE AVAILABLE UNDER A CREATIVE COMMONS ATTRIBUTION LICENSE (CC-BY 4.0), SO THAT USERS HAVE TO GIVE CREDIT TO THE ORIGINAL DATA CREATORS." ⁷</i></p>	<p>After the end of the project, new consent will be asked to the farmers (owners of the data) to reuse their data for further research.</p>
<p>Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, please provide it here.</p> <p><i>INDICATE WHETHER YOU INTEND TO ADD A PERSISTENT AND UNIQUE IDENTIFIER IN ORDER TO IDENTIFY AND RETRIEVE THE DATA.</i></p>	<p>No, 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/).</p>

⁷ Source: Ghent University Generic DMP Evaluation Rubric: <https://osf.io/2z5g3/>

What are the expected costs for data sharing? How will these costs be covered?	<ol style="list-style-type: none"> 1. Costs for maintaining the local server and Synology are covered by follow-up projects. 2. Preprints of articles will be available on a preprint server, such as Zenodo, upon submission and free of charge. 3. Articles will be available upon publication of the anonymized research results. Publication costs will be covered by the project budget. 4. All source code is stored on GitLab and access is shared with supervisors and authorized colleagues, free of charge. 5. Making scripts, functions and snips of code publicly available on GitLab after the end of the project is free of charge. 6. Data storage units (SharePoint and OneDrive) hosted by KU Leuven (not by the group) are made available free of charge.
---	---

7. Responsibilities

Who will manage data documentation and metadata during the research project?	The FWO PhD student (Dyan Meuwissen) will be responsible for the data documentation & metadata.
Who will manage data storage and backup during the research project?	The FWO PhD student (Dyan Meuwissen) will be responsible for data storage & backup during the project.
Who will manage data preservation and sharing?	The FWO PhD student (Dyan Meuwissen) will be responsible for the data preservation and sharing during the project, but will cooperate closely with her 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 will update and implement this DMP?	The FWO PhD student (Dyan Meuwissen) bears the overall responsibility for updating & implementing this DMP