FWO DMP Template - Flemish Standard Data Management Plan

Version KU Leuven

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	**Viviana Di Pietro (0000-0003-1116-9795)**
Contributor name(s) (+ ORCID) & roles	**Tom Wenseleers (0000-0002-1434-861X) – Promotor**
Project number [1] & title	PDMT2/24/042 – From genes to societies: evolutionary conflicts over caste development in bee societies.
Funder(s) GrantID [2]	BOF
Affiliation(s)	X KU Leuven ☐ Universiteit Antwerpen ☐ Universiteit Gent ☐ Universiteit Hasselt ☐ Vrije Universiteit Brussel ☐ Other: ROR identifier KU Leuven: 05f950310
Please provide a short project description	This project aims to unravel the genetic and behavioural mechanisms controlling caste development in bee societies. To do so, it will employ a unique blend of cutting-edge genomic techniques and sophisticated behavioural experiments. In a first part of the project, I will conduct a Genome-Wide Association Study (GWAS) to identify the genetic determinants of queen development in *Melipona* stingless bees, where individual control over caste fate leads to intense social conflict. An innovative reciprocal cross design will facilitate this investigation and will also allow me to test if caste determination is subject to intragenomic conflict, as is predicted by David Haig's kin conflict theory of genomic imprinting. In a second part of the project, I will explore how honeybee societies resolve caste conflicts, and investigate whether behavioural policing mechanisms can deter larvae from cheating on their intended caste fate. This project's unique integration of advanced genomic and behavioural studies promises to significantly advance our understanding of the evolution of eusociality in insects and will greatly enhance our knowledge on how caste development is regulated.

[1] "Project number" refers to the institutional project number. This question is optional. 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 data [3].

				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
Melipona beecheii reciprocal crosses	Dataset containing information about reciprocal crosses between the two population and the percentage of queen overproduction before and after the manipulation	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☐ Images ☐ Sound ☒ Numerical ☒ Textual ☐ Model ☐ Software ☐ Other:	.csv; .txt	☒ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	
Melipona beecheii wing clips	Dataset with description of the data for Genome Wide Association study (GWAS) and genomic imprinting	☒ Generate new data ☐ Reuse existing data	☒ Digital ☒ Physical	☐ Audiovisual ☐ Images ☐ Sound ☒ Numerical ☒ Textual ☐ Model ☐ Software ☒ Other: Biological data	.csv;	☒ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	Wing clips from each mother queen (50 samples) + pool of 20 female offspring (antennae from 20 pupae) per each colony

Apis mellifera caste conflict pilot experiment	Dataset with experimental data from caste conflict pilot experiments in honeybees, behavioural observations	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☒ Audiovisual ☐ Images ☐ Sound ☒ Numerical ☒ Textual ☐ Model ☐ Software ☐ Other:	. MTS, .csv, .txt	☐ < 1 GB ☒ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☒ NA	Biological samples may include larvae used in experimental grafting and their developmental outcomes

GUIDANCE:
The data description forms the basis of your entire DMP, so make sure it is detailed and complete. It includes digital and physical data and encompasses the whole spectrum ranging from raw data to processed and analysed data including analysis scripts and code. Physical data are all materials that need proper management because they are valuable, difficult to replace and/or ethical issues are associated. Materials that are not considered data in an RDM context include your own manuscripts, theses and presentations; documentation is an integral part of your datasets and should described under documentation/metadata.
RDM Guidance on data

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.	NA
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, refer to specific datasets or data types when appropriate and provide the relevant ethical approval number.	☐ Yes, human subject data; provide SMEC or EC approval number: ☐ Yes, animal data; provide ECD reference number: ☐ Yes, dual use; provide approval number: ☒ No Additional information:

[3] Add rows for each dataset you want to describe.

Will you process personal data[4]? If so, please refer to specific datasets or data types when appropriate and provide the KU Leuven or UZ Leuven privacy register number (G or S number).	☐ Yes (provide PRET G-number or EC S-number below) ☒ No Additional information:
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 If yes, please comment:
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 ☒ No If yes, please explain:
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.	☐ Yes ☒ No If yes, please explain:

3. Documentation and Metadata

[4] See Glossary Flemish Standard Data Management Plan

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). *RDM guidance on documentation and metadata*.	To ensure the data generated during the project remains understandable and usable for myself and others, the following documentation and data management practices will be employed: 1. Electronic Lab Notebooks (ELNs): I will use ELNs to systematically record experimental procedures, including the setup of reciprocal crosses, genome sequencing protocols, and the handling of biological samples. Detailed entries will include dates, locations, colony identifiers, and experimental conditions. These notebooks will be updated in real-time to ensure accuracy and will serve as a chronological and searchable repository of project activities. 2. Unique Identifiers and Metadata: Each biological sample (e.g., bee specimens, extracted DNA, or genome sequencing data) will be assigned a unique identifier. Metadata for each sample will include species, colony origin, genotype, experimental condition, and date of collection, facilitating cross-referencing and interpretation. 3. README.txt Files: Each dataset will be accompanied by a README.txt file, which will detail the methodologies used for data generation, such as genome sequencing, SNP identification, and behavioural observations. These files will describe the structure, variable definitions, and units of the data to ensure proper reuse and understanding. 4. Codebooks (e.g., Codebook.csv): For quantitative datasets, a comma-separated values (CSV) codebook will be prepared. This will contain detailed descriptions of variables, including SNPs, phenotypic traits, and caste allocation metrics, along with any codes or abbreviations used. 5. Data Storage and Accessibility: • All electronic data, including ELNs, raw sequencing data, and processed results, will be securely stored on institutional servers with regular backups. • Biological samples will be stored at -80°C in duplicate storage facilities to prevent loss. • Open-access repositories such as Dryad, Zenodo or Mendeley Data will be used to share datasets and documentation, upon publication of the data.

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 ☐ No If yes, please specify (where appropriate per dataset or data type) which metadata standard will be used: Our laboratory is currently evaluating the most suitable metadata standards for biological data in the context of this project. We aim to ensure that all datasets generated are easily accessible and reusable in compliance with established best practices. We intend to adhere to standards such as the Dublin Core Metadata Standard for general dataset documentation and the Access to Biological Collections Data (ABCD) Schema, which provides a comprehensive framework for managing metadata related to biological samples and observations. This approach will facilitate interoperability and ensure the datasets can be integrated with broader ecological and genomic research efforts. For genomic data, repository-specific metadata guidelines (e.g., NCBI GenBank or EBI's European Nucleotide Archive) will be followed, including detailed annotations for genomic sequences and ontologies such as the Gene Ontology (GO) for functional data. By combining these standards with custom metadata where required—covering experimental methods, sample identifiers, and variable definitions—we ensure that the project aligns with FAIR principles, enhancing data discoverability and reusability. If no, please specify (where appropriate per dataset or data type) which metadata will be created:

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

Where will the data be stored? *Consult the interactive KU Leuven storage guide to find the most suitable storage solution for your data.*	☒ Shared network drive (J-drive) ☒ Personal network drive (I-drive) ☒ OneDrive (KU Leuven) ☐ Sharepoint online ☐ Sharepoint on-premis ☐ Large Volume Storage ☐ Digital Vault ☐ Other:
How will the data be backed up? *WHAT STORAGE AND BACKUP PROCEDURES WILL BE IN PLACE TO PREVENT DATA LOSS?*	☒ Standard back-up provided by KU Leuven ICTS for my storage solution ☐ Personal back-ups I make (specify) ☒ Other (specify) QNAP NAS servers owned by Prof. Wenseleers
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 Prof. Wenseleers currently owns two QNAP NAS servers with approximately 20TB of available storage. Therefore, there will be no issue with storing the data in multiple copies to ensure redundancy and prevent data loss. ☐ No If no, please specify:

How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons? *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.* *Guidance on security for research data*	To ensure that the data are securely stored and protected from unauthorized access or modification, the following measures will be implemented: Physical Security: Biological samples or backup drives will be housed within the laboratory facilities of Prof. Wenseleers. Access to these facilities is restricted to authorized personnel and requires badge access, ensuring that only designated staff members can enter. Network Security: Data will be stored on Prof. Wenseleers's NAS devices, which support encryption protocols to secure the stored files. Data transfers will occur over secure, encrypted connections (e.g., SFTP or HTTPS) to prevent interception or unauthorized access during transmission. Computer Systems and File Security: Access to the NAS systems is restricted to specific user accounts Role-based access controls (RBAC) will be employed, ensuring users can only access data relevant to their role. Files are regularly backed up to multiple NAS devices to prevent data loss in case of hardware failure By implementing these physical, network, and computer system security measures, we ensure the data generated during the project remains secure, protected from unauthorized access, and resilient against potential threats.
What are the expected costs for data storage and backup during the research project? How will these costs be covered?	There are no costs expected

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...). *Guidance on data preservation*	☒ All data will be preserved for 10 years according to KU Leuven RDM policy ☐ All data will be preserved for 25 years according to CTC recommendations for clinical trials with medicinal products for human use and for clinical experiments on humans ☐ Certain data cannot be kept for 10 years (explain)
Where will these data be archived (stored and curated for the long-term)? *Dedicated data repositories are often the best place to preserve your data. Data not suitable for preservation in a repository can be stored using a KU Leuven storage solution, consult the interactive KU Leuven storage guide.*	☐ KU Leuven RDR ☐ Large Volume Storage (longterm for large volumes) ☒ Shared network drive (J-drive) ☐ Other (specifiy):
What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	There are no costs expected.

6. Data Sharing and Reuse

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. *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*	☒ Yes, as open data ☐ Yes, as embargoed data (temporary restriction) ☐ Yes, as restricted data (upon approval, or institutional access only) ☐ No (closed access) ☐ Other, please specify: All types of data, including genomic data and scripts will be made publicly available upon the publication of the research findings.
If access is restricted, please specify who will be able to access the data and under what conditions.	NA
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.	☐ Yes, privacy aspects ☐ Yes, intellectual property rights ☐ Yes, ethical aspects ☐ Yes, aspects of dual use ☐ Yes, other ☒ No If yes, please specify:
Where will the data be made available? If already known, please provide a repository per dataset or data type.	☒ KU Leuven RDR ☒ Other data repository (specify): Dryad, Zenodo, Mendeley Data ☐ Other (specify)

When will the data be made available?	☒ Upon publication of research results ☐ Specific date (specify) ☐ Other (specify)
Which data usage licenses are you going to provide? If none, please explain why. *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.* *Check the RDR guidance on licences for data and software sources code or consult the License selector tool to help you choose.*	☒ CC-BY 4.0 (data) ☐ Data Transfer Agreement (restricted data) ☐ MIT licence (code) ☐ GNU GPL-3.0 (code) ☐ Other (specify)
Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, please provide it here. *INDICATE WHETHER YOU INTEND TO ADD A PERSISTENT AND UNIQUE IDENTIFIER IN ORDER TO IDENTIFY AND RETRIEVE THE DATA.*	☒ Yes, a PID will be added upon deposit in a data repository ☐ My dataset already has a PID ☐ No
What are the expected costs for data sharing? How will these costs be covered?	Currently, no costs are expected for the existing data management plan.

7. Responsibilities	
Who will manage data documentation and metadata during the research project?	The applicant, Viviana Di Pietro

Who will manage data storage and backup during the research project?	The applicant, Viviana Di Pietro
Who will manage data preservation and sharing?	The applicant, Viviana Di Pietro, will take care of data preservation during the duration of the project. After the end of the project, the promotor, Tom Wenseleers, will be responsible for the long-term data preservation and sharing.
Who will update and implement this DMP?	The applicant, Viviana Di Pietro