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	Helena Mendes Ferreira 0000-0003-2090-4528
Contributor name(s) (+ ORCID) & roles	Tom Wenseleers 0000-0002-1434-861X Promotor
Project number[1] & title	3E230554 PDMT2/23/031 Testing the hologenome theory of evolution in social insects
Funder(s) GrantID[2]	BOF
Affiliation(s)	x KU Leuven ☐ Universiteit Antwerpen ☐ Universiteit Gent ☐ Universiteit Hasselt ☐ Vrije Universiteit Brussel ☐ Other: Provide ROR[3] identifier when possible:
Please provide a short project description	This project aims to test the hologenome theory of evolution in social insects. The hologenome theory posits that if microbiomes are vertically inherited, host and microbiome will co-adapt and could then potentially act as a single unit of selection. To test this theory, our project will use wholegenome shotgun metagenomic microbiome and mitogenome sequencing of solitary and social bees to quantify changes in mode of transmission and degree of phylosymbiosis of gut microbiome symbionts across different levels of sociality. Subsequently, microbiome transplant experiments will be conducted to study the impacts of microbiome variation on host fitness. Finally, the project will test if intercolony variation in microbiome composition and fitness consequences are transmitted across generations and are heritable. Overall, this project will provide a unique test of the hologenome theory of evolution and allow for a better understanding of how hosts and microbiomes co-evolve.

---

[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.

[3] Research Organization Registry Community. https://ror.org/

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[4].

Dataset Name	Description	New or Reused	Digital or Physical	*ONLY FOR DIGITAL DATA* Digital Data Type	*ONLY FOR DIGITAL DATA* Digital Data Format	*ONLY FOR DIGITAL DATA* Digital Data Volume (MB, GB, TB)	*ONLY FOR PHYSICAL DATA* Physical Volume
Phylosimbiosis solitary species (Biological samples and description)	Biological samples *of Osmia cornuta* mason bees and *Xylocopa violacea* carpenter bees and the datasets with the description on the data collected	☒ Generate new data ☐ Reuse existing data	☒ Digital ☒ Physical	☒ Observational ☒ Experimental ☐ Compiled/ aggregated data ☐ Simulation data ☐ Software ☐ Other ☐ NA	☐ .por ☐ .xml ☒ .tab ☒ .csv ☐ .pdf ☒ .txt ☐ .rtf ☐ .dwg ☐ .tab ☐ .gml ☒ other: Biological data ☐ NA	☒ < 100 MB ☐ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ < 10 TB ☐ < 50 TB ☐ > 50 TB ☒ NA	-90 samples *Osmia Cornuta* -90 samples *Xylocopa violacea*
Phylosimbiosis social species (Biological samples and description)	Biological samples of Buff tailed bumblebee *Bombus terrestris* and	☒ Generate new data ☐ Reuse existing data	☒ Digital ☒ Physical	☒ Observational ☒ Experimental ☐ Compiled/ aggregated data ☐ Simulation	☐ .por ☐ .xml ☒ .tab ☒ .csv ☐ .pdf	☒ < 100 MB ☐ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB	-180 samples (*Bombus terrestris*) -130 samples (*Apis mellifera* subspecies)

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

| | the Western honeybee *Apis mellifera* from two different subspecies, *A. m. mellifera* and *A. m. carnica* the datasets with the description on the data collected | | | data
☐ Software
☐ Other
☐ NA | ☒ .txt
☐ .rtf
☐ .dwg
☐ .tab
☐ .gml
☐ other:
☐ NA | ☐ < 10 TB
☐ < 50 TB
☐ > 50 TB
☐ NA | |

GUIDANCE:

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.

EXAMPLES OF DATA TYPES: OBSERVATIONAL (E.G. SURVEY RESULTS, SENSOR READINGS, SENSORY OBSERVATIONS); EXPERIMENTAL (E.G. MICROSCOPY, SPECTROSCOPY, CHROMATOGRAMS, GENE SEQUENCES); COMPILED/AGGREGATED DATA[5] (E.G. TEXT & DATA MINING, DERIVED VARIABLES, 3D MODELLING); SIMULATION DATA (E.G. CLIMATE MODELS); SOFTWARE, ETC.

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.

DIGITAL DATA VOLUME: PLEASE ESTIMATE THE UPPER LIMIT OF THE VOLUME OF THE DATA PER DATASET OR DATA TYPE.

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).

---

[5] These data are generated by combining multiple existing datasets.

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, please describe these issues further and refer to specific datasets or data types when appropriate.	☐ Yes, human subject data ☐ Yes, animal data ☐ Yes, dual use ☒ No If yes, please describe:
Will you process personal data[6]? 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.	☐ Yes ☒ No If yes: - Short description of the kind of personal data that will be used: - Privacy Registry Reference:
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:

---

[6] See Glossary Flemish Standard Data Management Plan

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

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).	I will use Electronic Lab Notebooks (ELNs) to record sample collection and colony manipulation details, including date, time, location, colony specifics, and treatments. These ELNs will be updated in real-time to minimize errors, providing a comprehensive chronological record. Each biological sample will have a unique identifier with accompanying information, including, species, subspecies, apiary, colony origin, generation, date of collection, and other relevant notes. README.txt files will provide context for datasets, including methodology regarding the sample collection and data structure. Similarly, a tab-separated values codebook will accompany the datasets, providing detailed descriptions of the data fields, codes used, and their respective meanings, ensuring accurate data interpretation. Biological samples will be stored at -80°C in two locations for redundancy, catalogued digitally with ELNs, and .txt files will simplify sample retrieval. The respective documentation will be structured for easy understanding, following open data principles while complying with ethical and legal requirements for biological research.

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 studying the best metadata standards to be used when referring to biological samples. We intend to adhere to similar standards, such as The Access to Biological Collections Data (ABCD) Schema, which serves as a developing and comprehensive standard for accessing and exchanging information about specimens and observations. 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?	Biological samples will be stored in the laboratory facilities, which are equipped with multiple refrigerators, including a -80°C fridge. Additionally, datasets resulting from the curation of this data will be kept on KU Leuven OneDrive as well as on the lab's two QNAP NAS systems.

| How will the data be backed up?

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.[7]

REFER TO INSTITUTION-SPECIFIC POLICIES REGARDING BACKUP PROCEDURES WHEN APPROPRIATE.	Our lab is equipped with 2 QNAP NAS (Network Attached Storage) serves as a centralized and secure storage solution, facilitating the storage, backup, and sharing of data across a network. All the data generated will be saved on both servers, to ensure that the data is kept safe and I will perform backups on a weekly basis.
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 ☐ No If yes, please specify concisely: Our server, QNAP NAS, currently has over 50 TB of available space, which exceeds the requirements of the current data management plan. "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. [7] | Access to the physical storage of the data is restricted solely to authorized personnel. Physical data will be maintained within our laboratory facilities, which requires a badge to access that is only provided to authorized staff. Concerning network facilities and the security of computer systems and files, data will be stored in QNAP NAS systems. These systems employ encryption technology and operate through a secure connection, ensuring the safety of stored data and preventing unauthorized access. |

---

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

What are the expected costs for data storage and backup during the research project? How will these costs be covered?	Since all data storage systems, including the aforementioned QNAP NAS systems, are already established in the lab, no anticipated costs are foreseen. The same applies to physical data, as the lab is already equipped with the facilities that ensure the storage of the data.

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...).	The data (both physical data such as biological samples and digital data) will be retained for at least ten years, with the possibility of indefinite retention.
Where will these data be archived (stored and curated for the long-term)?	All data generated in this project, whether physical or digital, will be preserved in the lab facilities. The lab is equipped with several fridges, including a -80°C unit for the physical biological samples, and two servers (QNAP NAS) for the digital data. For the long-term archiving of data, digital repositories will be used to store various types of data, including genomic, raw, and processed data. The selected repositories include NCBI (https://www.ncbi.nlm.nih.gov/), Dryad (https://datadryad.org/stash), Zenodo (https://zenodo.org/), and the KU Leuven Research Data Repository (RDR) (https://www.kuleuven.be/rdm/en/rdr). The selected platform for data repository aligns with the established protocol procedures in our lab, which are consistently employed for data storage.
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, in an Open Access repository ☐ Yes, in a restricted access repository (after approval, institutional access only, …) ☐ No (closed access) ☐ Other, please specify:
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.	As the current data management plan only involves the collection of biological samples, the current question does not apply. However, possible repositories include the KU Leuven Research Data Repository (RDR), Dryad, or alternative platforms based on the specific publishing guidelines of the journal, such as Dryad or Zenodo.

When will the data be made available? *THIS COULD BE A SPECIFIC DATE (DD/MM/YYYY) OR AN INDICATION SUCH AS 'UPON PUBLICATION OF RESEARCH RESULTS'.*	Data will be made publicly available upon the dissemination/publication of the research findings.
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.* *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." [8]*	All data from the project is expected to be accessible, and it will be made available under a Creative Commons Attribution License (CC-BY 4.0), requiring users to give credit to the original data creators.
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 ☐ No If yes: Upon publication of the results/findings.
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.

---

[8] Source: Ghent University Generic DMP Evaluation Rubric: https://osf.io/2z5g3/

7. Responsibilities	
Who will manage data documentation and metadata during the research project?	The applicant (Helena Mendes Ferreira)
Who will manage data storage and backup during the research project?	The applicant (Helena Mendes Ferreira)
Who will manage data preservation and sharing?	The applicant (Helena Mendes Ferreira) will be responsible for ensuring data preservation and reuse during the duration of the project. After the end of the [project, responsibility for a long-term data preservation and reuse will be assigned to the promotor Tom Wenseleers
Who will update and implement this DMP?	The applicant (Helena Mendes Ferreira) bears the end responsibility for updating & implementing this DMP