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	Lauren Van de Vliet 0000-0003-0723-5446
Contributor name(s) (+ ORCID) & roles	PI – Hans Steenackers (https://orcid.org/0000-0001-5021-2069)
Project number[1] & title	1SHBO24N Expanding the horizon of Bacillus environmental probiotics using experimental evolution
Funder(s) GrantID [2]	206532
Affiliation(s)	☒ KU Leuven ☐ Universiteit Antwerpen ☐ Universiteit Gent ☐ Universiteit Hasselt ☐ Vrije Universiteit Brussel ☐ Other: ROR identifier KU Leuven: 05f950310
Please provide a short project description	In this project we will explore the potential of experimental evolution to improve the antipathogenic activity of Bacillus probiotics with the specific focus on application in a hospital environment. We will evaluate 4 different set-ups (co-culture in static/shaking liquid, agar plates & swarming set-up) in which we will use competition as a selective pressure. We will validate the improved strains in vitro on abiotic surfaces and characterize the genetic basis behind the evolved traits. We will test the stability of the evolved traits and possible trade-offs with traits important for probiotic function. Finally, we will explore sequential evolution, where 2 different selection pressures from 2 different set-ups will be imposed sequentially to increase the overall activity of the probiotic.

[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
Bacterial strain collection	**WP1-8:** Wild-type probiotic strains of *Bacillus* (*B. amyloliquefaciens, B. subtilis, B. velezensis, B. pumilus, B. safensis, B. licheniformis, B. claussii, B. megaterium, B. paralicheniformis*) will be evolved, either alone or with the pathogenic strains of *S. aureus* Newman, *E. coli* UTI89 and *P. aeruginosa* PAO1, all available in the MiCA lab. Both probiotic and pathogenic strains will be evolved and genetically modified. All newly obtained evolved and constructed strains will be deposited in the bacterial collection of the Centre for Microbial and Plant Genetics (CMPG).	☒ Generate new data ☒ Reuse existing data	☐ Digital ☒ Physical	/	/	/	2 mL cryotubes and 96-well plates (300 µL)
Bacterial cell counts	**WP1-4,7,8**: Quantification of bacterial cell counts using flow cytometry and CytExpert software (Beckman Coulter) during competition assays and evolution experiments and other evaluations of antipathogenic activity	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☐ Images ☐ Sound ☒ Numerical ☐ Textual	.fcs, .xit, .csv, .xlsx	☒ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB	/

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

				☐ Model ☐ Software ☐ Other:		☐ NA	
Characterisation of (evolved) probiotic strains	**WP1-4:** Emulsion/inhibition zone diameters, OD measurements for surfactant quantification and growth rate assays.	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☐ Images ☐ Sound ☒ Numerical ☐ Textual ☐ Model ☐ Software ☐ Other:	.xlsx,	☒ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	/
Time-lapse swarming experiments	**WP1-2:** Evaluation of swarming behaviour before, during and after evolution by capturing images every 30 mins using a Epson scanner	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☒ Images ☐ Sound ☐ Numerical ☐ Textual ☐ Model ☐ Software ☐ Other:	.jpeg, .tiff	☒ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	/
Sequencing data	**WP5:** Whole genome sequencing reads of the (evolved) bacterial strains	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☐ Images ☐ Sound ☒ Numerical ☐ Textual ☐ Model ☐ Software ☐ Other:	.fastq	☐ < 1 GB ☒ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	/

Image J Swarming analysis	**WP1,2:** ImageJ Macro scripts for the quantification of swarming areas before, during and after evolution	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☐ Images ☐ Sound ☐ Numerical ☐ Textual ☐ Model ☒ Software ☐ Other:	.ijim	☒ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	/

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.	/
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:
Will you process personal data[4]? If so, please refer to specific datasets or data types when	☐ Yes (provide PRET G-number or EC S-number below) ☒ No

[4] See Glossary Flemish Standard Data Management Plan

appropriate and provide the KU Leuven or UZ Leuven privacy register number (G or S number).	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: The MiCA lab already has experience with valorization, through contacts with the Leuven Research and Development office. If the evolved strains show promising activity, these strains can be patented after consultation with the LRD office.
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: 11 of the 17 *Bacillus* strains in this project are strains commercialized in a probiotic cleaner. A material transfer agreement is under construction with the LRD office. In this MTA, specifics for valorization and publication will be discussed. The usage of these commercialized strains requires an agreement on what can be published and patented.
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). *RDM guidance on documentation and metadata.*	The MiCA lab uses a sharepoint-based electronic lab notebook (ELN) to document the experimental data. For each project, all data is organized in a standardized manner with a fixed set of metadata attached to it (such as the coordinating researcher etc.). All different types of data (raw, processed and final) are stored in different, specific folders.
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 no, please specify (where appropriate per dataset or data type) which metadata will be created: The ELN incorporates a fixed set of metadata, such as the coordinating researcher, the researcher performing the experiment, the data and used protocol for the experiment.

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 (Sharepoint based ELN hosted by KUL) ☐ 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) The ELN and the internal KU Leuven work with an automatic back-up for the used drive capacity, meaning the following: • A back-up every few hours (at 9h, 12h30, and 17h), of which the latest 7 versions are saved • A daily back-up (at 21h), of which the latest 10 version are saved • A weekly back-up (Sunday at 11h), of which the latest 6 version are saved
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 The drive system of the KU Leuven provides expandable storage space based on their internal server , which is maintained by the IT service of KU Leuven (SET-IT). The MiCA lab uses two drives on the KU Leuven internal server: the J-drive & the K-drive. The J-driver is meant for daily use, while the K-drive serves as a long-term storage space.

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*	The access to the ELN only works via KU Leuven Single Sign On, where each user should have a valid KU Leuven intranet user ID and password. Permission for access can be defined in detail by the local admin of the SharePoint site. For the biological samples, storage will occur in a secured -80°C freezer at the facility.
What are the expected costs for data storage and backup during the research project? How will these costs be covered?	An estimation for the costs of the ELN can be made based on other SharePoint services (sites) with a similar storage capacity, as accounting the costs for the new system has not been done yet: • €344.80 for the first year • €274.80 for the following years If necessary, additional storage space on the KU Leuven internal server can be bought. The MiCA lab uses two drives on the internal server: the J-drive & the K-drive: • Costs for the J-drive: €519/TB per year • Costs for the K-drive: €100/TB/year All costs will be covered by the obtained project budget.

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) All research data will be retained for at least five years after the end of the project.
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): KU Leuven internal server via the electronic lab book
What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	Similar to the costs of the storage during the project the costs for the Sharepoint ELN services will include: • €344.80 for the first year • €274.80 for the following years And costs for expansion of the internal KUL drives: • Costs for the J-drive: €519/TB per year • Costs for the K-drive: €100/TB/year The costs will be covered by the obtained project budget.

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: Published, digital data (including scripts, cell counts, characterization data etc.) will be deposited in an Open Access Repository, while unpublished data will be made available on the KUL research data repository (RDR).
If access is restricted, please specify who will be able to access the data and under what conditions.	The unpublished data on the KUL RDR will be restricted to institutional access.
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: If certain datasets include information on the commercialized *Bacillus* strains, the MTA will include what can be shared and what must be kept restricted.
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) ☐ Other (specify)

When will the data be made available?	☒ Upon publication of research results ☐ Specific date (specify) ☒ Other (specify) The published data will be made available upon publication of research results, with links to the datasets in the corresponding publications. Unpublished data will be made available on RDR at the end of this project.
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) The shared research data will be protected using a Creative Commons license (CC-BY-NC-SA-4.0).
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 When the data are deposited in the KUL RDR, a DOI is automatically assigned to the data.
What are the expected costs for data sharing? How will these costs be covered?	KU Leuven staff has free access to the Research Data Repository (RDR), managed by the KU Leuven. Costs of other data repositories will be covered by the project budget.

7. Responsibilities	
Who will manage data documentation and metadata during the research project?	Lauren Van de Vliet
Who will manage data storage and backup during the research project?	Lauren Van de Vliet and ELN IT Team
Who will manage data preservation and sharing?	Lauren Van de Vliet, ELN Technicians and Hans Steenackers (PI)
Who will update and implement this DMP?	Lauren Van de Vliet