

---

## FWO project

*A Data Management Plan created using DMPonline.be*

**Creator:** n.n. n.n.

**Affiliation:** KU Leuven (KUL)

**Funder:** Fonds voor Wetenschappelijk Onderzoek - Research Foundation Flanders (FWO)

**Template:** FWO DMP (Flemish Standard DMP)

**Principal Investigator:** n.n. n.n.

**Data Manager:** n.n. n.n.

**Grant number / URL:** G061824N

**ID:** 209105

**Start date:** 01-01-2024

**End date:** 01-01-2029

### Project abstract:

In eco-evolutionary studies, the phenotype has a central place, but phenotypic changes driven by the MICRObiome associated metaboLITEs have so far not been considered. The gut microbiota is seen as a source of metabolic flexibility for the host, but it is not clear how it structures fast acclimatisation to stress. We hypothesise that this is possible through host phenotypic changes. We will use the Daphnia - microbiome experimental model, in which host genotype x microbiome interactions have been unraveled through controlled gut microbiome transplants leading to modified host phenotypes. We will link the phenotypes with microbial communities and expressions through metagenomics and metabolomics. We expect that the MICROLITE expression is stressor and host genotype specific and reflects microbial adaptation in geographic mosaics of environmental stress (exposure to toxic cyanobacteria or viral infections). Through mono-association experiments, the role of the detected microbial strains and gene functions will be validated. The role of the microbiome associated metabolites in eco-evolutionary processes will be further tested in semi-realistic settings of mesocosms studying community and ecosystem features. Deciphering the mechanisms underlying these tolerances is essential to predict how zooplankton communities can respond to cyanobacterial blooms and viral epidemics to be able to control them in natural ecosystems.

**Last modified:** 13-07-2024

## FWO project

### FWO DMP (Flemish Standard DMP)

#### 1. 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.

				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
WP1. Transplant experiment	Data on host phenotypes following gut microbiome transplants in spreadsheet	Generate new data	Digital	Experimental	csv	<100MB	NA
WP2. Metabolomics data		generate new data	digital	metabolome		<5GB	NA
WP3. Metagenomics data		generate new data	digital	metagenome	FASTQ	<5GB	NA
WP4. Gene expression analysis data		generate new data	digital	gene expression		<100MB	NA
In all WPs: samples of gut microbiome isolates		generate new isolates	Physical				1000 aliquots

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)? Describe these issues in the comment section. Please refer to specific datasets or data types when appropriate.

- No

Will you process personal data? If so, briefly describe the kind of personal data you will use in the comment section. Please refer to specific datasets or data types when appropriate.

- No

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.

- No

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 in the comment section to what data they relate and what restrictions are in place.

- No

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 in the comment section to what data they relate and which restrictions will be asserted.

- No

## 2. 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).

Digital data:

We will maintain a record of the following for every WP (where applicable):

- Experimental design and protocol (.doc file)
- Sampling protocol (.doc file)
- Abbreviations used (.doc file)
- Structure of the data (.doc file)
- Steps involved in data analysis and relevant analysis scripts (R scripts, MOTHUR/QIIME scripts)
- Raw data (specific file format according to data type)
- Analysed data (specific file format according to data type)
- Index file/read me file (.txt file) for every WP, linking the name, location (folder and subfolder on KULAK server/hard disk) and description of above-mentioned files. These index files will be named according to the WP number.

In the concluding stage of the project, a master index file containing the combined information for all the WPs will be compiled which will be archived and also stored on the personal harddrives/PC of each PI.

Physical data:

Samples taken from natural systems or from experiments will be documented and stored for up to three years after the end of the project. Storage will be in fixative or in freezers depending on the kind of sample.

Microbial strains will be preserved as freezer stocks and a file (excel file) containing description of storage (storage medium/revival guide/location) will be maintained.

**Will a metadata standard be used to make it easier to find and reuse the data? If so, please specify (where appropriate per dataset or data type) which metadata standard will be used. If not, please specify (where appropriate per dataset or data type) which metadata will be created to make the data easier to find and reuse.**

The experiments are unique, but the data will be standardized according to data-type across experiments to make it easier to interpret the structure. Below, we list the metadata standards applicable to this project:

- 1) Ecological Metadata Language (EML), a metadata standard specific for the discipline of ecology will be used as a standard for community metadata for this project. We will use the application Morpho to create, edit and search metadata files following the EML specifications. Such files can be used to exchange information within and among research groups.

2) For metagenetics and metagenomics data, metadata will be organized and reported according to standard developed by the Genomic Standards Consortium (GSC) for reporting marker gene sequences— MInS – Minimum information about any (x) sequence

### 3. Data storage & back-up during the research project

#### Where will the data be stored?

On the Server space at Department of Biology. For sequencing data we will use extra server space available at KU Leuven KULAK (ICT service, contact Brecht Samyn, Brecht Monkerhey).

#### How will the data be backed up?

server space is centrally backed up

#### 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.

We expect that the available server space will be enough. We will also buy additional external hard disks to have back-up copies safely stored in an easy to access way. For sequencing data, we will use the server space available at KU Leuven KULAK (ICT service, contact Brecht Samyn, Brecht Monkerhey).

#### How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

The data will be protected by a password. For extra safety we will for the smaller data sets also have back-ups on external hard disks that are kept at the office of the PI Ellen Decaestecker.

#### What are the expected costs for data storage and backup during the research project? How will these costs be covered?

Costs foreseen in the consumable money on the project application should be enough. KULAK campus offers good deals for campus members and researchers.

### 4. 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...).

Digital data: We will retain all data for the expected 5-year period. For most publications we will make the data publicly available on data repositories.

Sequencing data will be submitted to public databases (EBI-ENA/NCBI-SRA), where they will be permanently archived to preserve access to the public. For community composition and phenotype data we will consider data repositories like DRYAD, GBIF and Zenodo.

Physical data: Freezer stocks of microbial strains will be maintained in the lab and will be available upon request. After the conclusion of the project, we will deposit the microbial strains at BCCM (Belgian Coordinated Collections of Microorganisms), where they will be preserved free of cost. Samples taken from natural systems or from experiments will be stored for up to three years after the end of the project. Storage will be in fixative or in freezers depending on the kind of sample.

#### Where will these data be archived (stored and curated for the long-term)?

Digital data will be archived on the K-drive of the server of KU Leuven KULAK.

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

For external hard disks and most data repositories there are no extra costs. When using DRYAD, one-time submission cost will be covered with the FWO-project fund.

## 5. Data sharing and reuse

**Will the data (or part of the data) be made available for reuse after/during the project? In the comment section please explain per dataset or data type which data will be made available.**

- Yes, in an Open Access repository

Published data will be made available. Unpublished results will be made available after an embargo period (3 years; exceptionally 5 years after the project).

**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 in the comment section per dataset or data type where appropriate.**

- No

**Where will the data be made available? If already known, please provide a repository per dataset or data type.**

Upon publication everybody can access (EBI-ENA/NCBI-SRA, GBIF, Zenodo, DRYAD); conditions may apply e.g. information on what will be done with the data.

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

Upon publication, or 3 years (exceptionally 5 years) after the end of the project to enable us to publish our results. In most cases whenever data are used for a publication they will become available through DRYAD and sequence databases.

**Which data usage licenses are you going to provide? If none, please explain why.**

We will attach Creative Commons Licences (CC BY) to the data. All sequenced data will be uploaded to the public data repository SRA from NCBI. While it will be made confidential and not accessible for the general public until publication, it ensures that the data is safely maintained in a standard and accessible format. The processing of molecular data will occur on the high-performance computing infrastructure available through the Flemish supercomputing center and KU Leuven.

Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, you have the option to provide it in the comment section.

- Yes

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

No expected costs for public data repositories except for DRYAD (if we decide to use it, one-time submission cost will be covered by the project fund)

## **6. Responsibilities**

**Who will manage data documentation and metadata during the research project?**

PI (prof. dr. Ellen Decaestecker), and the day-to-day managers of the FWO-project (Lab manager Isabel Vanoverberghe)

**Who will manage data storage and backup during the research project?**

PI (prof. dr. Ellen Decaestecker), and the day-to-day managers of the FWO-project (Lab manager Isabel Vanoverberghe)

**Who will manage data preservation and sharing?**

PI (prof. dr. Ellen Decaestecker), and the day-to-day managers of the FWO-project (Lab manager Isabel Vanoverberghe)

**Who will update and implement this DMP?**

PI (prof. dr. Ellen Decaestecker), and the day-to-day managers of the FWO-project (Lab manager Isabel Vanoverberghe)