DMP title

Project Name My plan (FWO DMP) - DMP title
Principal Investigator / Researcher Charlotte Theys
Institution KU Leuven

1. General Information
Name applicant

Charlotte Theys

FWO Project Number & Title

1138722N
An eco-evolutionary study of the role of the gut microbiome in driving the pace-of-life and consequences for pathogen tolerance within and across life stages in a damselfly

Affiliation

- KU Leuven

2. Data description
Will you generate/collect new data and/or make use of existing data?

- Generate new data

Describe in detail the origin, type and format of the data (per dataset) and its (estimated) volume. This may be easiest in a table (see example) or as a data flow and per WP or objective of the project. If you reuse existing data, specify the source of these data. Distinguish data types (the kind of content) from data formats (the technical format).

Digital data	Format	Volume	How created
Numerical data of measured response variables from ecological experiments: life history traits (mortality, growth rate, development time), behavioural traits (activity, boldness), physiological traits (metabolic rate, phenoloxidase activity), water quality parameters, metal concentration.	.xls & .csv	6 files, 1 GB	Two field experiments and 4 common garden laboratory experiments in temperature-controlled incubators
Sequencing data of gut microbiome and bacterioplankton samples from ecological experiments: 16S rRNA sequences of V3-V4 region	.fastq.gz	6 files 2.5 GB each	Two field experiments and 4 common garden laboratory experiments in temperature-controlled incubators
Abbreviations	.doc	10 MB	Explanation of used abbreviations
Data structure	.doc	10 MB	Explanation of structure of data
Scanned files of collected data	.pdf	500 MB	Scanned files of collected, handwritten data (to serve as back-up)
Materials and methods: information about protocols and experimental design	.doc & .pdf	500 MB	Written protocols and experimental designs
Pictures of experimental setup, study organisms, sampling sites, etc.	.jpeg	5 GB	Pictures taken with camera or smartphone
Script of data-analyses: phenotype (life-history, behaviour, physiology) and gut microbiome / bacterioplankton analyses	.doc & .R	500 MB	Code and extra information about code for analysing phenotypical and sequencing data (gut microbiome / bacterioplankton samples)
Files from analysing sequencing data (metadata, taxonomy per ASV, ASV table per sample, rooted tree)	.txt & .qza	500 MB	Files created after analyses of sequencing data on supercomputer cluster (VSC)
Manuscripts	.doc	100 MB	Written manuscripts, revisions, revised manuscripts
Presentations	.ppt	5 GB	Platform and poster presentations made for conferences, seminars, etc.

Physical data

For each WP, samples of damselfly larvae (heads, bodies, and guts) will be stored (frozen at -80 °C) for further physiological analyses (heads and bodies) and for sequencing (guts).

For gut samples, after library preparations, extracted and purified DNA will be stored (frozen at -20 °C).

3. Legal and ethical issues

Will you use personal data? If so, shortly describe the kind of personal data you will use. Add the reference to your file in KU Leuven's Register of Data Processing for Research and Public Service Purposes (PRET application). Be aware that registering the fact that you process personal data is a legal obligation.

- No

Privacy Registry Reference:
Short description of the kind of personal data that will be used:

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, add the reference to the formal approval by the relevant ethical review committee(s)

- Yes

I will sample very common aquatic invertebrates in Sweden and France. I will locally dissect them or transport these to Belgium for indoor experimental work. I have taken the necessary measures regarding the Nagoya Protocol.

Does your work possibly result in research data with potential for tech transfer and valorisation? Will IP restrictions be claimed for the data you created? If so, for what data and which restrictions will be asserted?

- No

Do existing 3rd party agreements restrict dissemination or exploitation of the data you (re)use? If so, to what data do they relate and what restrictions are in place?

- No

4. Documentation and metadata
What documentation will be provided to enable reuse of the data collected/generated in this project?
Digital data
A record will be maintained of the following for each WP (if applicable):

- Experimental designs and protocols (sampling, culturing, weighing, etc.; .doc)
- Abbreviations (.doc)
- Data structure (.doc)
- Raw data and analysed data (.xls & .csv)
- Raw sequencing data and analysed data (.fastq.gz, .txt & .qza)
- Script of data-analyses with code and extra information about code for analysing phenotypical and sequencing data (.doc & .R)
- Read me file (.txt): name, folder location (OneDrive/local server/hard disk), description of abovementioned files, named according to WP.

Physical data
For each WP, samples of damselfly larvae (heads, bodies, and guts) will be stored (frozen at -80 °C) for further physiological analyses (heads and bodies) and for sequencing (guts).
For gut samples, after library preparations, extracted and purified DNA will be stored (frozen at -20 °C).
Number and location of stored boxes (at -80 or -20 °C) with samples of each WP will be kept and updated in one document (.doc).

Will a metadata standard be used? If so, describe in detail which standard will be used. If no, state in detail which metadata will be created to make the data easy/easier to find and reuse.

- No

5. Data storage and backup during the FWO project
Where will the data be stored?
The collected data will be stored on a personal computer, on a local server of the Department of

Biology, on OneDrive, and on Dryad.

How is backup of the data provided?

Local server of the Department of Biology, OneDrive, hard disk

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 available space on a personal computer/server/OneDrive will be enough to store all data, except for the raw microbiome sequencing data. Yet, these sequencing data will be stored on several hard disks and on the Sequence Read Archive (SRA) (online and free of charge).

What are the expected costs for data storage and back up during the project? How will these costs be covered?

No extra costs need to be made since enough storage is already available for the type of data that will be gathered and used.

An additional hard disk will be bought after 3 years, which can easily be covered by the personal bench fee (ca. € 100 for 5 TB).

Data security: how will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?

The back-ups on the hard disks will be kept in the office of the PI.

The data on the personal computer/server/OneDrive/Dryad are protected by passwords.

Stronger security procedures are not needed as I will not be working with data types that need higher security measures.

6. Data preservation after the FWO project
Which data will be retained for the expected 5 year period after the end of the project? In case only a selection of the data can/will be preserved, clearly state the reasons for this (legal or contractual restrictions, physical preservation issues, ...).

Digitial data:

All data will be retained for 10 years (KU Leuven guidelines) after the end of the project. After publication of the research data, the dataset will be made publicly available on data repositories (Dryad or Mendeley Data). Corresponding R scripts (with specific analyses and explanations) will also be retained for 10 years after the end of the project. After publication, R scripts will be made available online on GitHub.

Physical data:

DNA samples for sequencing will be stored for 10 years. Frozen samples for physiological analyses will not be stored for 10 years. These samples for physiological analyses (measurement of fat, sugar and protein content, etc.) cannot be kept for a long time, because after homogenization, enzyme activities typically decrease rapidly. These samples will be stored upon publication of the research data.

Where will the data be archived (= stored for the longer term)?

Each dataset will be archived on a personal computer, on a local server of the Department of Biology, on OneDrive, on hard disks, and on the Sequence Read Archive (SRA) for at least 10 years, conform the KU Leuven RDM policy.

A Read me file (.txt) will be made containing the name, folder location (OneDrive/local server/hard disk), and description of all files, named according to each WP.

Published data will be made available online on data repositories (Dryad or Mendeley Data), and R script on GitHub.

What are the expected costs for data preservation during the retention period of 5 years? How will the costs be covered?

An additional hard disk will be bought after 3 years, which can easily be covered by the personal bench fee (ca. € 100 for 5 TB). Afterwards, there are no extra costs for these external hard disk(s) and most data repositories (typically one-time submission cost for Dryad which is sometimes covered by the submission journal, Mendeley Data is free to use, as is GitHub).

The expected size of all files (datasets, R scripts) is small (max 10 GB), estimating that additional costs to preserve the data on the K-drive of the server of the Department of Biology for a 10 year period after the project has ended, will be very low (or non existing) for the PI, Robby Stoks.

7. Data sharing and reuse
Are there any factors restricting or preventing the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)?

- No

Which data will be made available after the end of the project?
After publication, each dataset (.csv) will be made available after being deposited in an online data repository. Corresponding R scripts will be made available on GitHub. Unpublished results will be made available after an embargo period (typically 3 years after the project).

Where/how will the data be made available for reuse?

- In an Open Access repository
- Upon request by mail

When will the data be made available?

- Upon publication of the research results

Upon publication of the research results, the dataset and corresponding R script will be made available. Or 3 years after the end of the project to enable publication of the results.

Who will be able to access the data and under what conditions?
Full datasets will be uploaded in a .csv format on Dryad or Mendeley Data, always as an open access dataset. Therefore, they will be available to anyone for any purpose.
Creative Commons Licences (CC BY) will be attached to the deposited data to enable researchers to access and reproduce the data.
The R scripts will be uploaded in a .R format in GitHub, always under open access. Therefore, they will be available to anyone for any purpose.

What are the expected costs for data sharing? How will the costs be covered?
There are no expected costs for public data repositories, except for Dryad (the one-time submission cost will be covered the personal bench fee, if the submission journal does not cover these costs). The data repository 'Mendeley Data' can be used to make datasets online available, and GitHub to make R scripts online available, both free of charge.

8. Responsibilities
Who will be responsible for data documentation & metadata?
Charlotte Theys

Who will be responsible for data storage & back up during the project?
Charlotte Theys

Who will be responsible for ensuring data preservation and reuse ?
Charlotte Theys

Who bears the end responsibility for updating & implementing this DMP?
Both Charlotte Theys and the PI, Robby Stoks, bears the end responsibility of updating and implementing this DMP.