DMP title

Project Name Data management plan 12V5222N - DMP title

Grant Title 12V5222N

Principal Investigator / Researcher Jeroen De Smet

Description During my senior postdoctoral fellowship, I will explore the virome in the gut of the black soldier fly. The research will include phage isolation and characterisation, as well as the mapping of the whole virome from the gut (parts) of larvae reared on different diets.

Institution KU Leuven

1. General Information

Name applicant

Jeroen De Smet

FWO Project Number & Title

12V5222N - Unraveling the gut virome in black soldier fly larvae (*Hermetia illucens*) and its interaction with the microbial gut community

Affiliation

- KU Leuven

2. Data description

Will you generate/collect new data and/or make use of existing data?

- Generate new data
- Reuse existing 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).

Type of data	Format	Volume	How created
Observational data	.tiff, .txt or .csv	5 GB	Digitzed version of observations during experiments (eg rearing data, phage counts)
Sequencing data	.FASTQ	50 GB	Results of illumina sequencing of all extracted gut samples to map the virome
Existing datasets	.FASTQ	10 GB	Existing datasets can be used to compare composition of virome in different studies
Papers	.doc & .pdf	2 GB	Final publications will be stored
Larval gut samples	physical	1000 guts	Extracted from larvae during feeding experiments
Phages	physical	20	Isolated from black soldier fly gut in WP2.

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

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)

- No

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

Both the sequencing data generated to map the entire virome and the phages isolated against key members of the black soldier fly microbiome have no direct potential for valorisation. The goal of the project is rather to generate and distribute fundamental knowledge on the role of bacterial viruses in the gut of this insect.

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?

- Yes
- No

For the comparison of the generated virome data, existing virome analyses will be used. However these are published datasets to which the correct acknowledgements will always be included in the research. Hence the use of this data does not lead to any resistrictions.

4. Documentation and metadata

What documentation will be provided to enable reuse of the data collected/generated in this project?

The necessary documentation (standard operating procedure (SOP)) will always be digitally documented for every executed experiment and attached to the results file (e.g. phage counts of phage stability assessment (WP2)). If the results file are .csv format, care will be taken to include a first tabsheet with a thorough explanation of which data can be found where in the document and how it has to be interpreted.

For the sequencing data (viromics), all needed metadata (as described by my collaborators at Quadram Institute in Roux et al, 2019, Nature Biotechnology), will be deposited alongside the actual data.

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

For the results of the sequencing (WP1), which will clearly be the largest dataset, all data will be deposited using Bioproject organisation in the metadata.

For all other digitized results from various experiments, the fellow will keep track of the executer, date, time of incubation and so one. This data will be added in the first tab of the excel (.csv) document .

5. Data storage and backup during the FWO project

Where will the data be stored?

The time-stamped master copy of the data will be kept on our research unit central storage facility. Copies can be made and kept on personal devices.

How is backup of the data provided?

The data will be stored on the university's central servers with automatic daily back-up procedures.

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

After the project, data will be stored on a server back-end storage. One block of storage entails 100 GB, which is sufficient for the expected output (around 50-60 GB).

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

All costs associated to data storage and back up are part of a general agreement at campus Geel between KU Leuven and Thomas More and hence are funded by the campus's financial means.

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

The access to the folder containing all project related data is restricted to researchers of our group and access can only be granted by the project coordinator.

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, ...).

KU Leuven policy on data management will be followed which entails a preservation term of 5 years. This policy will only be adjusted for the physical 'retain' samples and this is due to two reasons: (i) the slow deterioration of such samples over time and (ii) the physical limit on storage capacity at -20°C. Hence, an overturn of samples of about two years is estimated to be more relevant for this data type.

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

The digitalized data will be stored on the university's central servers (with automatic back-up procedures) for at least 5 years, conform the KU Leuven RDM policy.

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

All costs associated to data storage and back up are part of a general agreement at our campus between KU Leuven and Thomas More and hence are funded by the campus's financial means. At the moment no costs will have to be foreseen by the research group itself.

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?

Data for which added value can be seen in sharing (e.g. sequencing data) will be made available.

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

- In an Open Access repository

A dataverse repository will be selected that adheres to the FAIR principle.

When will the data be made available?

- Upon publication of the research results

Data will be made available either upon the publication of an article including the research results.

Who will be able to access the data and under what conditions?

The data will be put in an open access repository and will be made available only upon request to the fellow.

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

Open repositories will be selected that do not ask additional funds, hence no costs will be made.

8. Responsibilities

Who will be responsible for data documentation & metadata?

Jeroen De Smet (jeroen.desmet@kuleuven.be)

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

Jeroen De Smet (jeroen.desmet@kuleuven.be)

Who will be responsible for ensuring data preservation and reuse ?

Jeroen De Smet (jeroen.desmet@kuleuven.be)

Who bears the end responsibility for updating & implementing this DMP?

The FWO postdoctoral fellow bears the end responsibility of updating & implementing this DMP.