

DMP_11H1721N

Project Name My plan (FWO DMP) - DMP_11H1721N

Principal Investigator / Researcher Bart Geens

Institution KU Leuven

1. General Information

Name applicant

Bart Geens

FWO Project Number & Title

11H1721N: The role of DNA methylation in swarming behaviour of desert locusts

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).

Spanning the several work packages, I expect to generate several broad data types, which will be described below. An approximate volume will be given for each dataset.

1. RNA and methylome sequencing datasets (*g.* deep sequencing data) will be generated. These will mainly consist of fasta/fastq files and are between 0.1 and 1 TB in size.
2. Various physical data such as DNA, RNA, tissue samples,... These will be stored in appropriate conditions in the lab (*g.* -80°C freezers).
3. Microscopy images will be stored as .TIF files of 1-100 MB.
4. Manuscripts detailing observations and results will be stored in physical or digital formats (.docx, PDF, .txt files of several MB).

The total size of the generated data is estimated to be around 3 TB, which can be accommodated on both the Large Volume Storage (LVS) drives of KU Leuven, as well as on the lab's Network Accessible Storage (NAS) system. All biological samples generated can be stored in the lab's -20°C and -80°C freezers, fridges and storage shelves.

Additionally, I may make use of several publically available datasets, such as the genome assembly of *Schistocerca gregaria*. These datasets will be obtained from their respective repositories (NCBI, Genbank, PDB, ...) and referenced to appropriately when needed.

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)

- 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

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?

In terms of documentation, for each data type information will be provided on the software (version) and hardware used to generate the data, as well as the means needed to open/process it. Additionally, any metadata provided from the instrumentation will also be included (e.g. automatically generated capture data of microscopy images).

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

While we will not follow any published metadata standard, the lab will follow a strict set of rules aimed at optimizing data maintenance and storage. We will provide a text file with detailed explanation regarding the organization and naming rules of the data. Also, an index file will be generated and provided along with the data that will allow readers to quickly locate and identify the files of interest. The metadata (including name, date, project etc.) will be contained in the filenames and folder structure. More detailed information such as experimental design, sampling and research methods employed will be provided via additional info files located within the folders. All researchers in the lab are also obliged to take detailed written notes about their work (chronologically ordered) in their individual lab notebook, which is handed-in to the supervisor at the end of their job term.

5. Data storage and backup during the FWO project

Where will the data be stored?

The host lab will be responsible for storing all wet-lab data. The partnering research group of the UGent will be responsible for the storage of all the sequencing and bioinformatics data. The host lab will store the raw sequencing data and the principal bioinformatic analyses as well. Prior to publication or the end of the project, the wet-lab will be stored on a Network Accessible Storage solution (NAS) in the host lab. This NAS has both physical as well as locational redundancy, which ensures the security. Critical data can also be backed up on the secure cloud storage provided by the KU Leuven. The large sequencing datasets will be stored on a failsafe cloud-based storage system of the UGent.

How is backup of the data provided?

Data stored by the host lab is regularly backed up on a physically separated storage device in the host lab using Synology's backup software, which allows for incremental backups as well as periodic complete backups. The sequencing data stored in the UGent facilities are automatically backed up to both on-site as well as off-site storage.

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 host lab has acquired specialized hardware dedicated to large data storage. The UGent group has access to large volume storage servers on which to store all bioinformatic and sequencing data.

What are the expected costs for data storage and back up during the project? How

will these costs be covered?

We do not expect additional costs for storage besides operating costs for the servers.

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

Data in the host lab is both physically as well as digitally secured: The storage devices are kept in a restricted access area in a locked storage cabinet. Digitally, the servers are only accessible through a lab-restricted LAN using a Secure Shell (SSH) protocol. The UGent storage is located in a dedicated facility, only accessible using valid credentials.

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, ...).**

All data will be stored in the respective institutions for at least 5 years and uploaded to recognized data repositories upon publication.

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

After the initial storage of 5 years, the data will be archived. This will be done by transferring the data to archiving-drives. All papers generated from the project will be published as open access. The necessary datasets and supplementary information will be uploaded to international data repositories.

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

The costs for longer-term storage of all relevant project data on the KU Leuven's Large Volume Servers (LVS) are estimated at €4000 for a five year-period.

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?

All data will be made publically available upon publication.

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

- In an Open Access repository

Upon publication, the relevant data will be made available by uploading them to an open access repository according to the journal's policy. Alternatively, the data can be shared upon request by mail.

When will the data be made available?

- Upon publication of the research results

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

Data will be securely shared between the project's partners (and their lab members involved in the work) by granting joint access to eachothers' servers, in consultation with the respective ICT departments.

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

We do not expect extra costs for data sharing between partners. Nevertheless, most open access journals in our research field ask for publication processing fees, which seem to become increasingly expensive. Since the costs for the experimental work in the project are high too, an alternative that may help us reduce these growing publication costs is to make use of the new legislation that would allow us to make an accepted manuscript available to the public free-of-charge (after an embargo period of six months).

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

The PI of the host lab will be responsible for the documentation and metadata regarding the wet-lab. The PI of the UGent group is responsible for this documentation of the sequencing and bioinformatics data.

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

In consultation with the ICT departments at their respective institutions, the PI of the host lab will be responsible for storing and backing up all wet-lab and behavioural data, and the PI of the partnering institution in UGent for storing and backing up all sequencing and bioinformatics data.

Who will be responsible for ensuring data preservation and reuse ?

idem as above

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

The PI bears the end responsibility of updating & implementing this DMP.