DMP title

Project Name FWO DMP - Unraveling the molecular mechanisms underlying the biological control of hairy root disease in tomato by Paenibacillus strains - DMP title
Grant Title 1S27122N
Principal Investigator / Researcher Ilse Goyens
Project Data Contact ilse.goyens@kuleuven.be
Institution KU Leuven

1. General Information

Name applicant

Ilse Goyens

FWO Project Number & Title

1S27122N - Unraveling the molecular mechanisms underlying the biological control of hairy root disease in tomato by Paenibacillus strains

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 / source
1 HPLC-MS data	.baf, .txt, .xlsx, .pptx	Max. 150 GB	Separation and identification of antagonistic compounds by HPLC-MS analysis. .baf and .txt files are generated by the instruments. The results are summarized in .xlsx and/or .pptx files.
2 NMR data	.FID	Max. 5 GB	Elucidation of the molecular structure of the antagonistic compounds.
3 Genomes and DNA sequences of target genes	.fas, .fasta, .fastq, .txt	Max. 20 GB	Use of · publicly available data (genomes acquired from https://www.ncbi.nlm.nih.gov/genome/) · existing data in the research group: genomes of several *Paenibacillus* strains. · data generated through genome sequencing with either Illumina, Nanopore or PacBio sequencing. · data generated through Sanger sequencing of the 16S rRNA gene.

4 RNASeq data	.fasta, .fastq, .gtf, .gff, .tabular, .bam, .bai, .html, .R, .py, .jpg, .xlsx, .txt, .tsv	Max. 100 GB	Use of · existing data within the research group: genomes of strains ST15.15/027 and ST15.15/006 (.fasta) and their annotations (.gtf and .gff) · Generated data: raw Illumina RNA sequencing data is stored in .fastq format. Files of different formats generated throughout the entire data analysis pipeline (.tabular, .bam, .bai, .html, .xlsx, .txt, .tsv). Programming code to visualize the data is stored in .R files and code to summarize the data in tables is stored in .py files. Figures are stored in .jpg format.
5 Photographs	.jpg, .png, .tif	Max. 5 GB	Photographs of electrophoresis gels, petridishes, … generated in numerous different experiments. Stored in .jpg, .png or .tif format.
6 Flow cytometry data	.fcs	Max 10 GB	Competition experiments between *Agrobacterium* and *Paenibacillus*. Growth of these bacteria is measured over time with a flow cytometer instrument.
7 Greenhouse experiment	.xlsx, .jpg	Max. 1 GB	qPCR data to quantify *Agrobacterium* and *Paenibacillus* during greenhouse trials. Data is exported to .xlsx files and .jpg files.
8 Papers and other articles	.pdf, .docx	Max. 1 GB	Scientific papers and articles in other non-academic journals.
9 Figures in papers and articles	.jpg, .png, .svg	Max. 1 GB	Figures for papers and articles. Stored in .jpg or .png format and when possible in .svg format as well.
10 Physical items	/	/	RNA samples from the RNASeq experiment are stored at -80°C. Bacterial cultures are stored in glycerol, at -80°C, 3 replicates per strain. DNA samples of the greenhouse experiment and genome sequencing are stored at -20°C. Notes in hard copy lab books, all chronologically labeled.

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?

- Yes

Data of HPLC-MS, NMR and RNASeq will contribute to the identification of antimicrobial compounds with potential for valorisation. If the identified compounds are novel, they will be patented in collaboration with KU Leuven Research & Development (LRD).

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?

1. In all excel files a tab will be included containing metadata and a summary, providing enough information to understand the raw and processed data.
2. Additional .docx ReadMe files will be provided whenever
    1. More detailed information regarding an experiment is needed to understand the reasoning behind the work.
    2. Several (re-)analyses have been performed on the same raw data, in order not lose overview

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.

- Yes
- No

The MIAPE-MS metadata standard will be used for MS data, MiFlowCyt for flowcytometrydata, BMRB (Biological Magnetic Resonance data Bank) for NMR data and MIRAGE LC for LC data.

Comments will be provided for programming code (Python and R).

ReadMe files and documentation will be provided as explained in the previous entry.

5. Data storage and backup during the FWO project

Where will the data be stored?

The data will be stored in different locations.

1. Daily work is saved on a personal account on the Onedrive network of KU Leuven (up to 5TB per user). Throughout the entire PhD project, the daily work will be regularly copied from Onedrive to the research group' server (PME&BIM). This way, the daily work is preserved for potential later purposes. The server is managed, secured and backed-up by the ICTS service of KU Leuven. Unauthorized persons do not have access to this server.
2. Large files such as RNASeq data (raw sequencing reads and alignment files, see 2. RNASeq) and LC-MS data (see 1. HPLC-MS) are additionally stored on the research group' server (PME&BIM), the same server as mentioned in point 1.
3. The HPLC-MS data (see 1. HPLC-MS) is stored on an 2TB external hard disk as well (WD Elements). This hard disk will be used for this purpose only.

How is backup of the data provided?

The server (mentioned in point 1 in the previous question) is maintained by the ICTS service of KU Leuven with automatic daily back-up and mirror procedures.

Access to the public Onedrive service is regulated by the central KULeuven login. Onedrive

provides backups with mirror 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

KU Leuven provides sufficient storage and back-up capacity of the server during and after the project. The personal Onedrive account has sufficient storage and backup capacity as well.

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

Costs for this are covered by the project or the research group. The costs for Onedrive specifically are centrally financed by KULeuven.

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

The server is secured by the ICTS service of KU Leuven. Unauthorized persons do not have access to the folder containing the research data. Access can only be granted by the PI.

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, ...).

Digital data: KU Leuven policy on data management will be followed which entails a preservation term of 5 years.

Physical data: Compound extracts and DNA and RNA samples will be stored at -20°C for 5 years after the end of the project. Transformed bacterial strains will be stored at -80°C for 5 years after the end of the project.

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

The 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. A copy of the HPLC-MS data will exist on the external hard disk, as mentioned in section 5.

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

Expected costs are the same as in section 5, and are covered by the project or the research group.

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)?

- Yes. Specify:

The antimicrobial compounds might hold commercial value. If they do, we will collaborate with LRD to safeguard IP rights. This limits sharing of HPLC-MS, NMR and RNASeq data before ensuring IP rights.

Which data will be made available after the end of the project?

The aim is to publish all data or make it at least available for relevant requests after the end of the project. Data with valuable IP will be protected prior to publication.

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

- In an Open Access repository

Open access data repositories and research papers. Published data will be available for everybody with access to the publication as per publisher's rules.

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?

All data and metadata will be available only to the researches participating in the project before publishing. Published data will be available for everybody with access to the publication as per publisher's rules.

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

Publishing costs will be covered by the FWO project, or the lab. There are no expected costs for public data repositories.

8. Responsibilities

Who will be responsible for data documentation & metadata?

Ilse Goyens – the FWO fellow

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

Ilse Goyens – the FWO fellow

Together with promotor (Hans Rediers), co-promotor (Joleen Masschelein) and ICTS.

Who will be responsible for ensuring data preservation and reuse ?

Promotor (Hans Rediers), co-promotor (Joleen Masschelein)

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

Ilse Goyens – the FWO fellow