FWO DMP Template - Flemish Standard Data Management Plan

Project supervisors (from application round 2018 onwards) and fellows (from application round 2020 onwards) will, upon being awarded their project or fellowship, be invited to develop their answers to the data management related questions into a DMP. The FWO expects a **completed DMP no later than 6 months after the official start date** of the project or fellowship. The DMP should not be submitted to FWO but to the research co-ordination office of the host institute; FWO may request the DMP in a random check.

At the end of the project, the **final version of the DMP** has to be added to the final report of the project; this should be submitted to FWO by the supervisor-spokesperson through FWO's e-portal. This DMP may of course have been updated since its first version. The DMP is an element in the final evaluation of the project by the relevant expert panel. Both the DMP submitted within the first 6 months after the start date and the final DMP may use this template.

The DMP template used by the Research Foundation Flanders (FWO) corresponds with the Flemish Standard Data Management Plan. This Flemish Standard DMP was developed by the Flemish Research Data Network (FRDN) Task Force DMP which comprises representatives of all Flemish funders and research institutions. This is a standardized DMP template based on the previous FWO template that contains the core requirements for data management planning. To increase understanding and facilitate completion of the DMP, a standardized **glossary** of definitions and abbreviations is available via the following link.

1. General Project Information	
Name Grant Holder & ORCID	**Dominique Holtappels https://orcid.org/0000-0003-4263-3407**
Contributor name(s) (+ ORCID) & roles	**Prof. B. De Coninck - Supervisor https://orcid.org/0000-0002-9349-5086** **Prof. R. Lavigne - Co-supervisor https://orcid.org/0000-0001-7377-1314**
Project number[1] & title	**1225223N** **Mapping the plant-phage-bacterium interaction and impact of viral communities on plant health**
Funder(s) GrantID[2]	FWO junior Postdoctoral fellowship
Affiliation(s)	√ KU Leuven ☐ Universiteit Antwerpen ☐ Universiteit Gent ☐ Universiteit Hasselt ☐ Vrije Universiteit Brussel ☐ Other: Provide ROR[3] identifier when possible:

[1] "Project number" refers to the institutional project number. This question is optional since not every institution has an internal project number different from the GrantID. Applicants can only provide one project number.

[2] Funder(s) GrantID refers to the number of the DMP at the funder(s), here one can specify multiple GrantIDs if multiple funding sources were used.

[3] Research Organization Registry Community. https://ror.org/

Please provide a short project description	With a continuously growing human population and limited resources, an intensification of agriculture is key to meet the growing demand for food. Pests and pathogens, including bacterial phytopathogens such as Xanthomonas campestris pv. campestris in brassica crops, contribute to approximately 30% of production losses worldwide and local outbreaks of diseases significantly impair food production. Bacteriophages are proposed as a valid strategy to control bacterial diseases. However, the interaction between phages and the plant, as well as the effect of viruses on the plant's bacterial communities remains elusive. In this project, we will look into the plant-phage interaction by assessing the distribution of phages in plant tissues. Using a reporter collection of Arabidopsis thaliana, we will evaluate the plant responses to phages in general as this is to date unclear. Furthermore, we zoom in on the natural phageome of A. thaliana and its interaction with the bacterial community to give us fundamental insights. We will also assess the effects of well characterized phages that are taken up by the plant on its endo- and rhizosphere. Finally, effects of bacteriophages on beneficial microorganisms and their relation to biotic stress relief and plant growth will be evaluated. As such, this project serves as a benchmark for further research to look into the interaction between plants and their microbial community and the effects of viruses on this intimate interaction.

2. 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[4].

				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
Metagenome data	Metagenome data of the viral communities that are associated with diseased and healthy brassicas	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Observational ☒ Experimental ☐ Compiled/ aggregated data ☐ Simulation data ☐ Software ☐ Other ☐ NA	☐ .por ☐ .xml ☐ .tab ☐ .csv ☐ .pdf ☐ .txt ☐ .rtf ☐ .dwg ☐ .tab ☐ .gml ☒ other: fastq.gz ☐ NA	☐ < 100 MB ☐ < 1 GB ☒ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ < 10 TB ☐ < 50 TB ☐ > 50 TB ☐ NA	
16S sequencing data	16S community sequencing data of bacterial communities associated with brassicas	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Observational ☒ Experimental ☐ Compiled/ aggregated data ☐ Simulation data ☐ Software ☐ Other ☐ NA	☐ .por ☐ .xml ☐ .tab ☐ .csv ☐ .pdf ☐ .txt ☐ .rtf ☐ .dwg ☐ .tab ☐ .gml ☒ other: fastq.gz ☐ NA	☐ < 100 MB ☐ < 1 GB ☒ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ < 10 TB ☐ < 50 TB ☐ > 50 TB ☐ NA	
Fluorescence microscopy data	Microscopy data of Arabidopsis reporter collections	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Observational ☒ Experimental ☐ Compiled/ aggregated data ☐ Simulation data ☐ Software ☐ Other ☐ NA	☐ .por ☐ .xml ☐ .tab ☐ .csv ☐ .pdf ☐ .txt ☐ .rtf ☐ .dwg ☐ .tab	☐ < 100 MB ☒ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ < 10 TB ☐ < 50 TB ☐ > 50 TB ☐ NA	

[4]

					☐ .gml ☒ other: .tiff ☐ NA		
Bacterial and viral count data	CFU and PFU data for enumeration of viral and bacterial concentration	☒ Generate new data ☐ Reuse existing data	☒ Digital ☒ Physical	☐ Observational ☒ Experimental ☐ Compiled/ aggregated data ☐ Simulation data ☐ Software ☐ Other ☐ NA	☐ .por ☐ .xml ☐ .tab ☒ .csv ☒ .pdf ☐ .txt ☐ .rtf ☐ .dwg ☐ .tab ☐ .gml ☒ other: .tiff ☐ NA	☒ < 100 MB ☐ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ < 10 TB ☐ < 50 TB ☐ > 50 TB ☐ NA	Count data will be stored in physical notebooks as well as digitally in cvs files

GUIDANCE:

DATA CAN BE DIGITAL OR PHYSICAL (FOR EXAMPLE BIOBANK, BIOLOGICAL SAMPLES, …). DATA TYPE: DATA ARE OFTEN GROUPED BY TYPE (OBSERVATIONAL, EXPERIMENTAL ETC.), FORMAT AND/OR COLLECTION/GENERATION METHOD.

EXAMPLES OF DATA TYPES: OBSERVATIONAL (E.G. SURVEY RESULTS, SENSOR READINGS, SENSORY OBSERVATIONS); EXPERIMENTAL (E.G. MICROSCOPY, SPECTROSCOPY, CHROMATOGRAMS, GENE SEQUENCES); COMPILED/AGGREGATED DATA[5] (E.G. TEXT & DATA MINING, DERIVED VARIABLES, 3D MODELLING); SIMULATION DATA (E.G. CLIMATE MODELS); SOFTWARE, ETC.

EXAMPLES OF DATA FORMATS: TABULAR DATA (.POR,. SPSS, STRUCTURED TEXT OR MARK-UP FILE XML, .TAB, .CSV), TEXTUAL DATA (.RTF, .XML, .TXT), GEOSPATIAL DATA (.DWG,. GML, ..), IMAGE DATA, AUDIO DATA, VIDEO DATA, DOCUMENTATION & COMPUTATIONAL SCRIPT.

DIGITAL DATA VOLUME: PLEASE ESTIMATE THE UPPER LIMIT OF THE VOLUME OF THE DATA PER DATASET OR DATA TYPE.

PHYSICAL VOLUME: PLEASE ESTIMATE THE PHYSICAL VOLUME OF THE RESEARCH MATERIALS (FOR EXAMPLE THE NUMBER OF RELEVANT BIOLOGICAL SAMPLES THAT NEED TO BE STORED AND PRESERVED DURING THE PROJECT AND/OR AFTER).

[5] These data are generated by combining multiple existing datasets.

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.	N/A
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, please describe these issues further and refer to specific datasets or data types when appropriate.	☐ Yes, human subject data ☐ Yes, animal data ☐ Yes, dual use ☒ No If yes, please describe:
Will you process personal data[6]? If so, briefly describe the kind of personal data you will use. Please refer to specific datasets or data types when appropriate. If available, add the reference to your file in your host institution's privacy register.	☐ Yes ☒ No If yes: - Short description of the kind of personal data that will be used: - Privacy Registry Reference:
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.	☐ Yes ☒ No If yes, please comment:

[6] See Glossary Flemish Standard Data Management Plan

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 to what data they relate and what restrictions are in place.	☐ Yes ☒ No If yes, please explain:
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 to what data they relate and which restrictions will be asserted.	☐ Yes ☒ No If yes, please explain:

3. 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).	**Sequencing data will be accompanied by README.txt files that clearly describe the origin of the data, the sequencing date and facility, and metadata necessary for reanalysis of the read data. Data analysis including the version of the bioinformatical tools will be noted in digital notebooks and made accessible within the research group upon request.** **Experimental procedures will be catalogued in duplicate (one physical and one digital) notebooks and exported as .pdf files to prevent data manipulations. Within these notebooks, the experimental procedure, manipulation of samples, as well as the dates of data collection will be noted meticulously. Bacterial and viral count data will be stored in digital and physical notebooks and certified on a monthly basis by the supervisors of the project to prevent data manipulation.** **Microscopy data will be accompanied with README.txt files to catalogue metadata needed to interpret the images.**

Will a metadata standard be used to make it easier to **find and reuse the data**? If so, please specify which metadata standard will be used. If not, please specify which metadata will be created to make the data easier to find and reuse. *Repositories could ask to deliver metadata in a certain format, with specified ontologies and vocabularies, i.e. standard lists with unique identifiers.*	☒ Yes ☐ No If yes, please specify (where appropriate per dataset or data type) which metadata standard will be used: **Folders containing sequencing data will be named by the fellow's initials, followed by the experiment number and data of data collection: DH_exp001_DDMMYYYY** **Each folder will be provided with raw sequencing reads (folder: DH_exp00X_rawreads) along with a README.txt including information necessary metadata including details on the sequencing. Processed reads will be included in a separate folder as follows: DH_exp00X_assembledreads which includes a README.txt describing information on the processing of the reads, the assembler used (including version) and data necessary for the reanalysis of the data. Additional files generated will be included in a third folder: DH_exp00X_downstream.** **A similar set-up will be established for microscopy data** If no, please specify (where appropriate per dataset or data type) which metadata will be created:

4. Data Storage & Back-up during the Research Project	
Where will the data be stored?	**Data will be stored short term on a personal computer and backed up to the drives of KU Leuven. These drives are accessible lab members upon request.**

How will the data be backed up? *What storage and backup procedures will be in place to prevent data loss? Describe the locations, storage media and procedures that will be used for storing and backing up digital and non-digital data during research.*[7] *Refer to institution-specific policies regarding backup procedures when appropriate.*	**Digital data will be backed-up on a physical hard drive (1 TB data storage) that is securely stored within the lab and is accessible upon request. These physical back-ups are generated after raw sequencing/microscopy data is available to prevent data manipulation and accompanied by README.txt files describing the necessary metadata.** **Non digital data generated during the project will be signed and dated by the supervisor monthly during one-to-one meetings. Afterwards, pictures or scans will be taken from the notebooks and stored digitally on a hard drive. This hard drive contains a folder with the fellow's initials including a folder for each month throughout the fellowship with pictures of the notebook from every page of that month. Similarly, digital notebooks will be converted to .pdf files, dated and stored in a similar way.**
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 ☐ No If yes, please specify concisely: **A hard drive (1 TB) is available for the fellow to store all the data generated during the project.** If no, please specify:
How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons? *Clearly describe the measures (in terms of physical security, network security, and security of computer systems and files) that will be taken to ensure that stored and transferred data are safe.* [7]	**Temporary digital data will be stored on personal drives to ensure data manipulations from third parties. The external hard drive will be stored in a secured place (locked cabinet) with a key for the supervisor and fellow. This includes the notebook of the fellow and any other data that might be manipulated by others.**

[7] Source: Ghent University Generic DMP Evaluation Rubric: https://osf.io/2z5g3/

What are the expected costs for data storage and backup during the research project? How will these costs be covered?	**The expected costs for data storage include:** **100 euro for an external hard drive and will be covered by the fellow's bench fee**

5. 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...).	**All data generated during the project will be kept for at least five years on the drives of KU Leuven within an archived folder. Additionally, the external hard drive will be handed to the supervisor for safe keeping. This includes raw sequencing data, microscopy data, raw count data, as well as processed sequencing data and generated figures and datasheets. Physical notebooks will be kept in a locked closet with limited access by other parties.**
Where will these data be archived (stored and curated for the long-term)?	**The external hard drive and physical notebooks will be kept for long-term storage in secured cabinets.**
What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	**No additional costs will be expected for the retention period.**

6. Data Sharing and Reuse	
Will the data (or part of the data) be made available for reuse after/during the project? Please explain per dataset or data type which data will be made available. *Note that 'available' does not necessarily mean that the data set becomes openly available, conditions for access and use may apply. Availability in this question thus entails both open & restricted access. For more information: https://wiki.surfnet.nl/display/standards/info-eu-repo/#infoeurepo-AccessRights*	☒ Yes, in an Open Access repository ☐ Yes, in a restricted access repository (after approval, institutional access only, …) ☐ No (closed access) ☐ Other, please specify: **Sequencing reads will be made available as bioprojects in NCBI according to the published manuscripts with accession numbers included in the manuscript and final DMP.** **Isolated bacterial and viral strains will be deposited in NCBI accordingly.**
If access is restricted, please specify who will be able to access the data and under what conditions.	N/A
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 per dataset or data type where appropriate.	☐ Yes, privacy aspects ☐ Yes, intellectual property rights ☐ Yes, ethical aspects ☐ Yes, aspects of dual use ☐ Yes, other ☒ No If yes, please specify:

Where will the data be made available? If already known, please provide a repository per dataset or data type.	**Data will be made available on NCBI as a BioProject.**
When will the data be made available? *This could be a specific date (dd/mm/yyyy) or an indication such as 'upon publication of research results'.*	**The data will be made available upon publication with accession numbers included in the manuscript.**
Which data usage licenses are you going to provide? If none, please explain why. *A data usage license indicates whether the data can be reused or not and under what conditions. If no licence is granted, the data are in a grey zone and cannot be legally reused. Do note that you may only release data under a licence chosen by yourself if it does not already fall under another licence that might prohibit that.* *Example Answer: E.g. "Data from the project that can be shared will be made available under a Creative Commons Attribution license (CC-BY 4.0), so that users have to give credit to the original data creators."* [8]	**Data will be released under a Creative Commons Attribution license (CC-BY 4.0) to ensure correct referencing to the original manuscript and data collection.**
Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, please provide it here. *Indicate whether you intend to add a persistent and unique identifier in order to identify and retrieve the data.*	☒ Yes ☐ No If yes: **Data will be available under a BioProject accession number within NCBI.**

[8] Source: Ghent University Generic DMP Evaluation Rubric: https://osf.io/2z5g3/

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

7. Responsibilities	
Who will manage data documentation and metadata during the research project?	**The research fellow (Dominique Holtappels) will be responsible for data documentation and metadata during the research project.**
Who will manage data storage and backup during the research project?	**The research fellow (Dominique Holtappels) will be responsible for the management of data storage and back up of the research project.**
Who will manage data preservation and sharing?	**The research fellow (Dominique Holtappels) will be responsible for initial data preservation and storage. Long-term storage will be the responsibility of the supervisor of the project.**
Who will update and implement this DMP?	**This DMP will be updated and implemented by the research fellow (Dominique Holtappels) under close supervision of the supervisor (Barbara De Coninck).**