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 X	
Name Grant Holder & ORCID	**Yvan Carhel Njifon Ngapout; https://orcid.org/0000-0002-1719-138X; PhD researcher**
Contributor name(s) (+ ORCID) & roles	**Hervé Vanderschuren; https://orcid.org/0000-0003-2102-9737; Promoter**
Project number[1] & title	1SHEQ24N; Banana Beyond Bunchy Top Disease: identifying and applying natural virus resistance to mitigate losses in an important commodity
Funder(s) GrantID[2]	Fonds voor Wetenschappelijk Onderzoek (FWO)
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	Bananas are a vital crop for both food security and trade, providing a significant source of income for many countries, including non-producing nations such as Belgium. However, the crop is susceptible to pathogens, particularly the Banana Bunchy Top Virus (BBTV), which causes various symptoms, including stunted growth, chlorosis, deformation of leaf edges, and bunching of new leaves, leading to yield losses of up to 100%. BBTV is the most significant threat to banana production, and prevention of infection is the primary approach currently used. Unfortunately, due to the homogeneity of banana crops in the field, they are particularly vulnerable to pathogen attacks. As such, sustainable solutions such as natural resistance are necessary to combat BBTV pandemics. While resistance to BBTV has been reported in some accessions, no attempts have been made to map the resistance architecture of BBTV resistance in Musa spp. Recognizing the need for sustainable solutions and the availability of plant material through collaborations with IITA, ITC, and Meise Botanic Garden, we have decided to screen parental lines and mapping populations, conduct QTL mapping and RNA seq analysis. Our goal is to provide genomic tools, such as markers and candidate genes, to expedite the breeding of BBTV resistance. By doing so, we hope to protect this vital crop and ensure a stable supply of bananas for years to come.

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
Genotypic data	- Sequence reads for a selected set of banana accessions - The data will be provided by IITA or ITC or Meise Botanic garden or from public repository such as NCBI	☒ 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: .fasta; .fastq ☐ NA	☐ < 100 MB ☐ < 1 GB ☒ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ < 10 TB ☐ < 50 TB ☐ > 50 TB ☐ NA	
Gene expression data	- RNA sequence reads for a selection of banana accessions	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Observational ☒ Experimental ☐ Compiled/ aggregated data ☐ Simulation	☐ .por ☐ .xml ☐ .tab ☐ .csv ☐ .pdf	☐ < 100 MB ☐ < 1 GB ☒ < 100 GB ☐ < 1 TB ☐ < 5 TB	

[4] Add rows for each dataset you want to describe.

	- Expression of a targeted subset of genes			data ☐ Software ☐ Other ☐ NA	☒ .txt ☐ .rtf ☐ .dwg ☐ .tab ☐ .gml ☒ other: .fasta; .fastq ☐ NA	☐ < 10 TB ☐ < 50 TB ☐ > 50 TB ☒ NA	
Dry lab protocols	Protocols, R, linux and python scripts generated during the project to process available data	☒ 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: ☐ NA	☐ < 100 MB ☒ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ < 10 TB ☐ < 50 TB ☐ > 50 TB ☐ NA	
Wet lab protocols	Protocols used when conducting experiments	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Observational ☒ Experimental ☐ Compiled/ aggregated data ☐ Simulation data ☐ Software ☐ Other	☐ .por ☐ .xml ☐ .tab ☐ .csv ☐ .pdf ☒ .txt	☐ < 100 MB ☒ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ < 10 TB	

				☐ NA	☐ .rtf ☐ .dwg ☐ .tab ☐ .gml ☒ other: .Docx ☐ NA	☐ < 50 TB ☐ > 50 TB ☐ NA	
BBTV status table	List of screened biological material and their BBTV status	☒ 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: .Docx ☐ NA	☐ < 100 MB ☒ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ < 10 TB ☐ < 50 TB ☐ > 50 TB ☐ NA	

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).*	
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.	Data from iita: - https://data.iita.org/ - https://cassavabase.org/ftp//
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:

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

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: - BBTV status table -> Enhance genetic resource value by providing the BBTV status of commercially important banana breeding parental lines and that of a subset of strategically selected banana accessions. - Genotypic and gene expression data -> Map and narrow down genetic loci underlying BBTV resistance/tolerance. Any candidate marker/gene/locus is a potential breeding goal for the banana industry
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:

[6] See Glossary Flemish Standard Data Management Plan

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 sets will include a README.txt file explaining the data format, units, labels, and field-specific information. - Dry lab protocols will be provided with a header section with information on the date of creation and last modification, authors, and a brief description of the analyses performed by the script. The script will also be commented. - Wet lab protocols will be provided with a header section with information on the date of creation and last modification, authors and a brief descriptive of what the script does. The script will also be commented.
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: - For sequencing data uploaded on NCBI, these will follow NCBI's metadata standards (For BioSample, BioProject, SRA, GenBank). - Alternatively, metadata generated by the sequencing platform will be provided. 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?	Large data will be stored on my DATA and SCRATCH partitions on the HPC during analyses. A “working” copy will be stored on external hard drive, fast J-drive, or work PC.
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.*	Data will be backed up, archived, and remain modifiable on the secured L-drive of the KU Leuven Additional copies can be made on external hard drives or K-drives.
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: VSC_SCRATCH -> 10TB VSC_DATA -> 80 GB J-DRVIE -> 1TB K-Drive -> 1TB L-Drive -> 1TB KUL-Work computer -> 1 TB Western Digital Elements Desktop - External Hard Drive -> 2TB Please note that all the storage points mentioned above are expandable by contacting KUL's IT services or by buying an extra unit/copy/SD card. If no, please specify:

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

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]	ALL KU Leuven network drives (including HPC partitions and excluding OneDrive) are suitable for strictly confidential data. These drives enable several ways to modulate access rights.
What are the expected costs for data storage and backup during the research project? How will these costs be covered?	Cost for data storage on KU Leuven network drives are covered by my lab, the Laboratory of Tropical Crop Improvement

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...).	Upon completion of the project, data will be archived on the K-drive
Where will these data be archived (stored and curated for the long-term)?	Upon completion of the project, data will be archived on the K-drive.

What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	The Cost for data storage and preservation will be covered by the Laboratory of Tropical Crop Improvement

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:
If access is restricted, please specify who will be able to access the data and under what conditions.	
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.	NA

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'.*	Genetic data - > on NCBI BBTV status - > on MGIS
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]	CC-BY; Will be updated if necessary
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: NA

[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?	The cost of sharing data will be covered by the requesting parties

7. Responsibilities	
Who will manage data documentation and metadata during the research project?	The PhD researcher, under the supervision of the PI
Who will manage data storage and backup during the research project?	The PhD researcher, under the supervision of the PI
Who will manage data preservation and sharing?	The PI
Who will update and implement this DMP?	The PI and the PhD researcher