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	Agten Lieze; 0000-0002-8296-3964					
Contributor name(s) (+ ORCID) & roles	Michiels Jan; 0000-0001-5829-0897; Promotor					
	Verstraeten Natalie; 0000-0002-9548-4647; Co-promotor					
Project number ¹ & title	3E220848: Evolutionary trajectories of antibiotic susceptibility in longitudinal bacterial isolates					
Funder(s) GrantID ²	1102323N					
Affiliation(s)	✓ KU Leuven					
	☐ Universiteit Antwerpen					
	☐ Universiteit Gent					
	☐ Universiteit Hasselt					
	☐ Vrije Universiteit Brussel					
	✓ Other: VIB					
	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/

Antibiotics have revolutionized the treatment of infectious diseases worldwide. However, to cope with the detrimental effects of antibiotics, bacteria have at their disposal a wide range of defense strategies. In addition to the well-known example of resistance, therapy failure can result from the presence of a subset of cells that have transiently switched to an antibiotic-tolerant "persister" state. We and others have demonstrated that persister levels in vitro can be tuned to the frequency of antibiotic encounters by genetic adaptation. We hypothesize that a similar evolutionary process may occur in vivo. To lend support to this hypothesis, we will study persistence in longitudinal bacterial isolates obtained from patients with suprapubic catheters. Evolutionary trajectories of persistence will be charted and a postulated role for exposure to antibiotics as well as the composition of the microbiome on the evolution of persistence will be investigated. Since persistence is a known driver of resistance development, evolution of resistance will also be studied. Finally, mechanistic insights will be obtained into genes, regulatory networks and physiological conditions that drive persistence and its evolution. Results from this project will be integrated in a comprehensive model for persistence and its evolution in vivo. The latter is expected to contribute to the rational development of future therapeutic approaches to clear bacterial infections.

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 data⁴.

				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
WP1: Building a	collection of longitud	dinal isolates					
Collection of longitudinal urine samples	Urine samples of patients 1 – 10; frozen at - 80°C (±500); ethical clearance: S65007	☑ 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	2 ml cryotubes ±500 samples
Images of plated urine samples Plate counts urine samples	Images of the urine samples plated out on CHROMagar™ orientation medium Plate counts of a dilution series of	☑ Generate new data☑ Generate new data	☑ Digital☑ Digital	☑ Experimental☑ Experimental	☑ . jpeg☑ . xslx☑ . docx	⊠ < 100 GB ⊠ < 100 MB	

⁴ Add rows for each dataset you want to describe.

	the urine samples on CHROMagar™ orientation medium						
Collection of longitudinal bacterial isolates	Bacterial isolates from the longitudinal urine samples; ± 10 colonies per species; glycerol stocks frozen at -80°C	⊠ Generate new data	⊠ Physical				2 ml cryotubes ±3 000 isolates Backup in 96- wellplates
16S rRNA sequencing results	16S rRNA sequencing results of the bacterial isolates obtained by Sanger sequencing	⊠ Generate new data	⊠ Digital	⊠ Experimental	⊠ .fasta	⊠ < 1 GB	
Gels	Images of DNA gels	⊠ Generate new data	⊠ Digital		⊠ .Tif	⊠ < 1 GB	
WP2: In vitro anal	WP2: In vitro analysis of longitudinal isolates to reveal evolutionary trajectories and underlying mechanisms						
Quantification of persistence levels	Quantification of persistence levels of the longitudinal bacterial	⊠ Generate new data	⊠ Digital	⊠ Experimental	⊠ .xlsx	⊠ < 100 GB	

	isolates via persistence assays						
Quantification of resistance levels	Quantification of resistance levels of the longitudinal bacterial isolates via MIC assays	⊠ Generate new data	⊠ Digital	⊠ Experimental	⊠ .xlsx	⊠ < 100 GB	
Whole genome sequences	Whole genome sequencing results of the interesting strains isolated prior to or following a jump in persistence or resistance	⊠ Generate new data	⊠ Digital	⊠ Experimental	⊠ .fastq	⊠ < 100 GB	
DNA oligonucleotide sequences	Digitally preserved sequences of the DNA oligonucleotides needed to construct the mutants	⊠ Generate new data	⊠ Digital	⊠ Experimental	⊠ .fasta	⊠ < 1 GB	
DNA oligonucleotides	A few (± 50) Eppendorf tubes	□ Generate new data	⊠ Physical				1.5 ml Eppendorf tubes

	with the DNA oligonucleotides needed to construct the mutants; frozen at -20°C				
Lab strain mutants	Mutants from lab strains (corresponding to the natural strains) containing the interesting mutations from the WGS results; glycerol stocks frozen at -80°C	⊠ Generate new data	⊠ Physical		2 ml cryotubes 5 mutations will be selected → 5 mutants
Natural strain mutants	Mutants in which the interesting mutations from the WGS results are reverted; glycerol stocks frozen at -80°C	⊠ Generate new data	⊠ Physical		2 ml cryotubes 5 mutations will be selected → 5 mutants
Characterization of the mutations	Unknown at this point. Depends on the nature of the specific mutations.	☑ Generate new data☑ Reuse existing data	☑ Physical☑ DigitalExactdetails		

							
	Likely genetic,		unknown				
	biochemical and		at this				
	expression		point				
	analyses (at						
	population and						
	single cell level),						
	microscopy, the						
	identification of						
	interaction						
	partners, and a						
	molecular						
	characterization						
WP3: Validation of	f in vitro results in (conditions that mimic	the in vivo si	tuation			
Biofilm model	Infection model	⊠ Generate new	□ Digital		⊠ .xlsx	⊠ < 1 GB	
	to assess	data					
	persistence						
	levels in biofilm						
	settings						
	(hanging						
	polystyrene						
	pegs or flat well						
	microplates)						
Cathether model	Infection model	⊠ Generate new	□ Digital		⊠ .xlsx	⊠ < 1 GB	
	to assess	data					
	persistence						
	levels in urinary						
	cathethers						
Intracellular	Description of	⊠ Generate new	□ Digital		⊠ .docx	⊠ < 1 GB	
model of	intracellular	data					
infection in	model of						

human bladder cells	infection to assess persistence levels in human bladder cells						
Evolution of intracellular bacterial counts over time	Monitoring growth resumption of persisters by registering the evolution of intracellular bacterial counts over time	⊠ Generate new data	⊠ Digital	⊠ Experimental	⊠ .xlsx	⊠ < 1 GB	
Luminescence mutants	The interesting lab strain mutants and natural strain mutants (see WP2) are modified to express the luminescence operon luxCDABE	⊠ Generate new data	□ Physical				2 ml cryotubes
Mouse model	Adapt an existing in vivo mouse model for bioluminescenc	☑ Generate new data	⊠ Digital	⊠ Experimental	⊠ .docx	⊠ < 1 GB	

	e imaging to					
	monitor					
	persistence of					
	the					
	luminescence					
	mutants in vivo					
Bioluminescence	Bioluminescenc	⊠ Generate new	□ Digital	⊠ .tif	⊠ < 1 TB	
imaging mouse	e imaging data	data				
model	of the mouse					
	model					
WP4: Microbiome	analysis of longitu	dinal urine samples				
Microbiome	Microbiome	⊠ Generate new	□ Digital	☑ .fastq	⊠ < 1 TB	
analysis	analysis data of	data				
	the longitudinal					
	urine samples					

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.	While unknown at this point in time, MTAs might be needed for materials requested from other labs, for example for the characterization of the mutations in WP2 (e.g. reporters). Before signing, all MTAs and other legal documents will be proofread and approved by VIB Innovation & Business officers.
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: Human subject data: collection of urine samples from patients with a suprapubic catheter for which ethical clearance has been granted by the Ethics Committee UZ/KU Leuven (reference number s65007) to comply with applicable personal data protection and the processing of personal data. (dataset: Collection of longitudinal urine samples) Animal data: an <i>in vivo</i> mouse infection model will be used to assess persistence using bioluminescence imaging. Ethical clearance will be obtained by an ECD application when sufficient <i>in vitro</i> data is generated. (dataset: Mouse model)
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: pseudonymized patient data i.e. antibiotic use, age, sex, date of placement of catheter Privacy Registry Reference: approved by the Ethics Committee Research UZ/KU Leuven on March 4 2021 (reference number S65007).

⁶ See Glossary Flemish Standard Data Management Plan

Does your work have potential for commercial valorization (e.g. tech transfer, for example spinoffs, commercial exploitation,)? If so, please comment per dataset or data type where appropriate.	☐ Yes ☑ No If yes, please comment:
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: while unknown at this point in time, MTAs might be needed for materials requested from other labs, for example for the characterization of the mutations in WP2 (e.g. reporters). Before signing, all MTAs and other legal documents will be proofread and approved by VIB Innovation & Business officers.
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: Copyright of the generated data belongs to Jan Michiels, ownership of the data belongs to VIB and KU Leuven in accordance with the framework agreement between both institutes.

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).

EXPERIMENTAL RESULTS: Data files are named using "yearmonthday_titleordescription". The data are generated following standardized protocols that are saved in a OneNote notebook. To keep the data organised, day-by-day activities are registered in a detailed digital lab notebook in OneNote in which a page is made for each month and within such page the experiments are noted per day.

BIOLOGICAL MATERIAL: The cryotubes of the collection of urine samples are labelled according to the accompanying database. Cryotubes containing the bacterial isolates and mutants are labelled with the date of storage and an identifier (e.g. PatientNr_VisitNr_StrainIdentifier).

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.

If yes, please specify (where appropriate per dataset or data type) which metadata standard will be used: NGS files (.fastq) or microscopy files (.nd2) contain standard metadata on respectively sequencing technique or the Nikon metadata standards

REPOSITORIES COULD ASK TO DELIVER METADATA IN A CERTAIN FORMAT, WITH SPECIFIED ONTOLOGIES AND VOCABULARIES, I.E. STANDARD LISTS WITH UNIQUE IDENTIFIERS.

If no, please specify (where appropriate per dataset or data type) which metadata will be created: For datasets containing .xlsx files, extra tabs will be generated to explain the data (including the date, experimental protocol, species and strain name, mutations ...).

For remaining data sets for which there is no formally acknowledged metadata standard specific to our discipline, Dublin Core Metadata will be used. Additionally, MIBBI (Minimum Information for Biological and Biomedical Investigations) will be closely monitored for metadata standards that are more specific to our data.

4. Data Storage & Back-up during the Research Project

Where will the data be stored?	BIOLOGICAL MATERIAL: Urine samples in cryotubes will be temporarily stored at -80° in the Michiels lab and later transferred to the UZ/KU Leuven Biobank. Bacterial samples in tubes or 96-well plates will be stored at -80° in the Michiels lab (costs covered by general lab expenses). EXPERIMENTAL RESULTS: Data will be generally stored on two or more locations, one of which a secure university device on which the data is generated and the other a Onedrive/Sharepoint location.
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.	BIOLOGICAL MATERIAL: A backup of selected strains will be stored in -80° freezers present in a different location (Kevin Verstrepen lab, Heverlee). EXPERIMENTAL RESULTS Password-protected hard drives equipped with anti-virus programs will be used as backup. Also, the use of Onedrive/Sharepoint is backed up via a local copy and a back-up copy and the network drives of KU Leuven follow daily back-up procedures and offer version control via the built-in features of Windows.
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: BIOLOGICAL MATERIAL: Sufficient storage space is present in the -80°C freezers available in the Michiels lab. EXPERIMENTAL RESULTS: OneDrive at KU Leuven alone already offers 5 TB of data per user. Network storage is purchased on a group level an increased whenever needed. 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	BIOLOGICAL MATERIAL: Unauthorized people do not have access to the collection of strains. EXPERIMENTAL DATA: The datasets are stored in the university's secure environment. OneDrive storage is linked to my personal secured KU Leuven account that is secured by a two factor authorization and frequently changed passwords.
What are the expected costs for data storage and backup during the research project? How will these costs be covered?	BIOLOGICAL MATERIAL: Storage in the -80°C freezers in the Michiels lab of which the costs are covered by general lab expenses.
	EXPERIMENTAL DATA: The costs for large volume storage will be covered by general lab financing (€503.66 per TB per year).

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...).

BIOLOGICAL MATERIAL: All bacterial strains will be stored for at least 5 more years after the end of the project (and longer, complying with the 10 year data preservation rule of KU Leuven). For this purpose, -80°C freezers are present in the Michiels lab. Urine samples will be stored at the UZ/KU Leuven Biobank for 15 years following collection.

EXPERIMENTAL DATA: After the project, all data will be stored on the university's central servers with automatic back-up procedures for at least 5 years (and longer, complying with the 10 year data preservation rule of KU Leuven), conform the KU Leuven RDM policy. The costs will be covered by KU Leuven overhead budgets.

Where will these data be archived (stored and curated for the long-term)?	BIOLOGICAL MATERIAL: For the purpose of storage after the project, -80°C freezers are present in the Michiels lab. Urine samples will be stored at the UZ/KU Leuven Biobank for 15 years following collection.
	EXPERIMENTAL DATA: Data will be stored on the university's central servers for 10 years. Published results will be deposited conform the journal's policy.
What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	BIOLOGICAL MATERIAL: Storage in the -80°C freezers in the Michiels lab of which the costs are covered by general lab expenses.
	EXPERIMENTAL DATA: Estimated cost of storage on the university's central servers is 100 euro per year.

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.	 ✓ Yes, in an Open Access repository ☐ Yes, in a restricted access repository (after approval, institutional access only,) ☐ No (closed access) ☐ Other, please specify: 	
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	The aim is to communicate the results in top journals that require full disclosure upon publication of all included data. However depending on the journal, accessibility restrictions may apply. Proper links to these data sets will be provided in the corresponding publications.	
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.	The aim is to communicate our results in top journals that require full disclosure of all included data in an Open Access repository. After an embargo period of 5 years, final datasets that have not been published will be made publicly available using the Research Data Repository (RaDaR), managed by the KU Leuven.	

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'.	Upon publication of the research results. After an embargo period of 5 years for unpublished data to allow the research group to publish further research findings. The biological material generated in this project (e.g. microbial strains) will be shared upon simple request following publication.
Which data usage licenses are you going to provide? If none, please explain why.	Open Access publications in scientific, peer-reviewed journals are typically covered by a Creative Commons Attribution License (CC-BY). Also the RaDaR repository allows to share data using the CC-BY license.
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	
Do you intend to add a PID/DOI/accession	⊠ Yes
number to your dataset(s)? If already available, please provide it here.	□ No
INDICATE WHETHER YOU INTEND TO ADD A PERSISTENT AND UNIQUE IDENTIFIER IN ORDER TO IDENTIFY AND RETRIEVE THE DATA.	If yes: Manuscripts will be assigned a unique DOI upon publication. Likewise, datasets will receive a PID upon deposit in an online platform.

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

What are the expected costs for data sharing?	BIOLOGICAL MATERIAL: Shipment is generally paid by the requesting parties
How will these costs be covered?	EXPERIMENTAL RESULTS: Publication costs will be paid by general lab expenses. An online repository is
	free of charge. Network storage at KU Leuven comes with a cost of 100 euro per TB per year, covered by
	the host lab after termination of the project.

7. Responsibilities		
Who will manage data documentation and metadata during the research project?	Lieze Agten	
Who will manage data storage and backup during the research project?	Lieze Agten	
Who will manage data preservation and sharing?	Jan Michiels (Centre of Microbial and Plant Genetics, KU Leuven)	
Who will update and implement this DMP?	Jan Michiels (Centre of Microbial and Plant Genetics, KU Leuven)	