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	**Patrick Van Dijck, https://orcid.org/0000-0002-1542-897X,**
Contributor name(s) (+ ORCID) & roles	**Katrien Lagrou, https://orcid.org/0000-0001-8668-1350, copromoter** **Johan Maertens, https://orcid.org/0000-0003-4257-5980, copromoter**
Project number[1] & title	C3/22/007,
Funder(s) GrantID[2]	KU Leuven Industrial Research Fund
Affiliation(s)	X 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	Candida auris is the first pathogenic fungus to be officially considered an urgent antimicrobial resistance threat by CDC, while resistant Candida glabrata makes up 30% of all systemic Candida infections. Antifungal resistance is on the rise, while the antifungal drug market comprises only four classes, significantly reducing survival rates of systemic candidiasis. Nevertheless, insights into the evolutionary dynamics of antifungal resistance are scarce, although they can revolutionize treatment. By using high-throughput experimental evolution, we have mapped collateral sensitivity and cross resistance dynamics under different selective pressures from various antifungal drugs in Candida auris. Collaral sensitivity and cross resistance have been extensively studied in tumours and bacteria, but remain unexplored in fungi. Based on these evolutionary dynamics, we have designed drug cycling and combinatorial treatment schemes that prevent antifungal drug resistance development. In this project our main aim is to prove that certain drug cycling and combinatorial treatment schemes reduce the development of resistance and increase treatment success of C. auris and C. glabrata infections. Moreover, we will generate a vast library of resistant strains, on which a library composed of selected repurposed drugs with antifungal activity will be screened, to identify clinically approved compounds that show collateral sensitivity. This might lead to the discovery of adjuvant drugs that further reduce resistance development in the clinic. At last, resistance mechanisms will be investigated by both genome and targetted sequencing. Our preliminary study shows evidence of several novel mechanisms of (multi)drug resistance, which present a valuable resource for future drug target characterization and drug development.

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
Experimentally evolved strains	Resistant strains obtained from exposure to drugs in vitro or in vivo	☒ 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	One drawer in the -80 °C freezer, Max. 500 samples
Strain characterization data	Drug susceptibility profiling, in vivo colonization data, stress tolerance, virulence and competitive fitness output	☒ 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	

					☐ .gml ☒ other: .tiff, .png; .pzfx ☐ NA		
Sequencing	Both targeted and whole genome sequencing to unravel drug resistance mechanisms	☒ 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, .fasta., .bam, .vcf ☐ NA	☐ < 100 MB ☐ < 1 GB ☐ < 100 GB ☒ < 1 TB ☐ < 5 TB ☐ < 10 TB ☐ < 50 TB ☐ > 50 TB ☐ NA	All sequenced strains are stored at -80°C in the general lab stock. Sequenced DNA is stored at -20°C.

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

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.	We do not reuse previously published or public data.
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: The principle of collateral sensitivity will be validated in mice. Therefore mice will be infected with strains resistant to a certain drug and will then be treated with the drug towards they show collateral sensitivity to see whether alternative CS based treatment decreases the infective colonization of internal organs. Ethical approval has been granted to perform these experiments.

[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: A patent application was already filed by LRD. Our work may result in novel treatment options to treat infections caused by Candida auris
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: Some of the compounds will be received from pharmaceutical companies. Agreements between these companies and LRD are being prepared.
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).	We have a lab guide will standardized protocols that are being done in the lab. In addition, the results will be published and materials and methods will be described in detail so that the experienced researcher can repeat the analysis. All generated data is stored on the KU Leuven server, which contains protected project directories to which only researchers involved have access. In addition, every researcher has a personal directory on the KU Leuven server for safe data storage and a OneDrive and Teams directory. Data is never stored on personal or work devices to prevent data loss upon technical failure.
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: If no, please specify (where appropriate per dataset or data type) which metadata will be created: All raw data set are deposited in the online directories as described above in a special directory for raw data files, next to the processed data and final results (e.g. figures).

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

Where will the data be stored?	**All data is stored on the KU Leuven shared server and additional online directories (e.g. personal directory on KU Leuven server, Teams, OneDrive) as described before**
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.*	**The data will be stored on the university's central servers with automatic daily back-up 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 ☐ No If yes, please specify concisely: There is 1,1 TB available for the project folder in which data of this project belongs. 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]	**The data will be stored in the university's secure environment**
What are the expected costs for data storage and backup during the research project? How will these costs be covered?	**The expected cost is 500 euro/TB/year. We do not expect to go over that cost. This is part of the consumables on the project.**

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...).	**Both all raw data as well as final processed data presented in papers will be stored on the J drive (shared folder of KU Leuven server). All physical data (e.g. strains in -80°C freezers) will be stored for at least 5 years.**
Where will these data be archived (stored and curated for the long-term)?	**The data will be stored on the university's central servers (with automatic back-up procedures) for at least 10 years, conform the KU Leuven RDM policy.**

What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	**In view of the expected size of the database (less than 1 TB), estimated cost will be below 500 euro to set up the database and an annual fee of 50 euro for support.**

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.	**All sequencing data will be made available via the SRA (sequence read archive) of NCBI.**

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'.*	**As soon as the publication has been published as preprint and/or accepted for publication in a peer reviewed journal.**
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 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**
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:

[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?	**Sharing data via NCBI is free of charge. Data present in manuscripts will cost the cost of such publications which may go up to 6500 euro for a paper in Nature Communications.**

7. Responsibilities	
Who will manage data documentation and metadata during the research project?	**Hans Carolus (PhD student), Dimitrios Sofras (PhD student), Celia Lobo Romero (laboratory technician), Patrick Van Dijck (PI)**
Who will manage data storage and backup during the research project?	**Luc Grauwels (ICT support),**
Who will manage data preservation and sharing?	**Patrick Van Dijck**
Who will update and implement this DMP?	**Patrick Van Dijck**