FWO DMP Template - Flemish Standard Data Management Plan

Version KU Leuven

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	**Rosa-Maria Määttälä** (0000-0003-0564-4406)
Contributor name(s) (+ ORCID) & roles	**Vitor Bernardes Pinheiro** (0000-0003-2491-0028) *Supervisor* **Lien De Wannemaeker** (0000-0003-4073-8897) *Co-supervisor*
Project number [1] & title	Developing cross-species genetic tools and a standardized genetic circuit for non-model microorganism engineering
Funder(s) GrantID [2]	**1S50625N**
Affiliation(s)	☒ KU Leuven ☐ Universiteit Antwerpen ☐ Universiteit Gent ☐ Universiteit Hasselt ☐ Vrije Universiteit Brussel ☐ Other: ROR identifier KU Leuven: 05f950310

[1] "Project number" refers to the institutional project number. This question is optional. 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.

Please provide a short project description	The field of Synthetic Biology is rapidly advancing, and the use of genetic circuits in non-model microorganisms is becoming increasingly important for large-scale industrial applications. However, the absence of well-characterized genetic parts and, often, the lack of molecular tools greatly impede progress. This project aims to establish cross-species transcription and translation regulators as well as design and validate a standard for genomic engineering and traceability via barcoding. Synthetic RNA thermometers and terminators developed in this project will provide a collection of novel, modular, cross-species genetic parts to accelerate the development of non-model organisms in key industrial applications. Standardization has been a strategy successfully implemented by Synthetic Biology to optimize processes, enabling systematic developments and gains in productivity. Thus, a standardized context for genome engineering will be designed to facilitate the development of evolution-resistant genetic circuits in non-model organisms, while also introducing traceability. By providing a standardized platform for the development and implementation of genetic circuits in non-model microorganisms, this research is expected to have a major impact on the field of Synthetic Biology as it will enable researchers to efficiently integrate new genetic parts into diverse microorganisms, which will accelerate their development as Synthetic Biology chassis.

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 [3].

				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
Protocols & SOPs		☒ Generate new data ☒ Reuse existing data	☒ Digital ☒ Physical	☐ Audiovisual ☐ Images ☐ Sound ☐ Numerical ☒ Textual ☐ Model ☐ Software ☐ Other:	.pdf .docx	☒ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	< 50 pages
DNA	Purchased oligos & purified plasmids	☒ Generate new data ☐ Reuse existing data	☐ Digital ☒ Physical	☐ Audiovisual ☐ Images ☐ Sound ☐ Numerical ☐ Textual ☐ Model ☐ Software ☐ Other:		☐ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	1000 1.5 mL & 2mL tubes
Oligo specification sheets	Specification sheets provided by DNA synthesis	☒ Generate new data ☐ Reuse existing data	☒ Digital ☒ Physical	☐ Audiovisual ☐ Images ☐ Sound ☐ Numerical	.pdf	☒ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB	4 binders (~800 A4 paper sheets)

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

	companies			☒ Textual ☐ Model ☐ Software ☐ Other:		☐ > 5 TB ☐ NA	
Glycerol stocks		☒ Generate new data ☐ Reuse existing data	☐ Digital ☒ Physical	☐ Audiovisual ☐ Images ☐ Sound ☐ Numerical ☐ Textual ☐ Model ☐ Software ☐ Other:		☐ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	500 1.8 mL tubes (~ boxes) and 100 96-well plates
DNA sequences		☒ Generate new data ☒ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☒ Images ☐ Sound ☐ Numerical ☒ Textual ☒ Model ☐ Software ☐ Other:	.fasta .dna .txt .gb .csv	☒ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	
PDB files and protein amino acid sequences		☒ Generate new data ☒ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☒ Images ☐ Sound ☐ Numerical ☒ Textual ☒ Model ☐ Software ☐ Other:	.pdb	☒ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	
Agarose, SDS-PAGE, Urea-PAGE gel		☒ Generate new data ☐ Reuse existing	☒ Digital ☐ Physical	☐ Audiovisual ☒ Images ☐ Sound	.gel .jpg .png	☐ < 1 GB ☒ < 100 GB ☐ < 1 TB	

electrophoresis		data		☐ Numerical ☐ Textual ☐ Model ☐ Software ☐ Other:		☐ < 5 TB ☐ > 5 TB ☐ NA	
Sanger and Next-generation DNA sequencing data	Analysis of plasmids/genomic DNA	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☐ Images ☐ Sound ☐ Numerical ☐ Textual ☐ Model ☐ Software ☐ Other:	.ab .txt .fastq	☐ < 1 GB ☒ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	
Flow cytometry data	Analysis of DNA libraries using fluorescent proteins	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☒ Images ☐ Sound ☒ Numerical ☒ Textual ☐ Model ☐ Software ☐ Other:	.fcs .pptx	☐ < 1 GB ☐ < 100 GB ☒ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	
Plate reader data		☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☐ Images ☐ Sound ☒ Numerical ☒ Textual ☐ Model ☐ Software ☐ Other:	.xlsx	☒ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	
qPCR data		☒ Generate new data	☒ Digital ☐ Physical	☐ Audiovisual ☐ Images	.xlsx .csv	☒ < 1 GB ☐ < 100 GB	

		☐ Reuse existing data		☐ Sound ☐ Numerical ☒ Textual ☐ Model ☐ Software ☐ Other:		☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	
Scripts	Code written for analysis pipelines	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☐ Images ☐ Sound ☐ Numerical ☐ Textual ☐ Model ☒ Software ☐ Other:	R script .py	☒ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	

GUIDANCE:
The data description forms the basis of your entire DMP, so make sure it is detailed and complete. It includes digital and physical data and encompasses the whole spectrum ranging from raw data to processed and analysed data including analysis scripts and code. Physical data are all materials that need proper management because they are valuable, difficult to replace and/or ethical issues are associated. Materials that are not considered data in an RDM context include your own manuscripts, theses and presentations; documentation is an integral part of your datasets and should described under documentation/metadata.
RDM Guidance on data

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.	**Protocols & SOPs**: Protocols delivered in frequently used kits or existing standard lab protocols. **DNA sequences:** existing DNA sequences from GenBank, Addgene, published literature **Protein amino acid sequences**: fluorescent proteins, selection markers, etc.
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, refer to specific datasets or data types when appropriate and provide the relevant ethical approval number.	☐ Yes, human subject data; provide SMEC or EC approval number: ☐ Yes, animal data; provide ECD reference number: ☐ Yes, dual use; provide approval number: ☒ No Additional information:

Will you process personal data[4]? If so, please refer to specific datasets or data types when appropriate and provide the KU Leuven or UZ Leuven privacy register number (G or S number).	☐ Yes (provide PRET G-number or EC S-number below) ☒ No Additional information:
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: Any RNA thermometer or terminator sequence can be used in industrial biotechnology strains via co-development agreements or licensing of individual parts.
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

[4] See Glossary Flemish Standard Data Management Plan

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). *RDM guidance on documentation and metadata*.	Each experiment is documented in an electronic lab notebook on Benchling by the performing scientist following standard operating procedures (SOPs), which will be or have been written down. Each notebook has a table of summary containing the background and rational with the objective, protocols and samples used, results and conclusions, linking to other notebooks as needed. The raw data of each experiment will be sorted per experiment type and stored in separate folders. Processed data will be stored in separate folders with the same name containing links to their respective raw data files. Separate documents of non-experimental nature will be sorted and stored in a documents folder. README.txt files will be included explaining the design/protocol, analysis methods, results, labels used and references to the electronic lab notebook, following the relevant reporting guidelines per experiment type (e.g. MIFlowCyt for flow cytometry experiments). Metadata will link the data files, lab samples, and experimental notes (including descriptions of equipment, setting, and used experimental settings). New folders will be created for processed data with links to the raw data included when needed for specific publications.
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: When depositing data in a local or public repository, the final dataset will be accompanied with a README.txt file containing all relevant information, following the Dublin Core Metadata standard (*if no other meta-standard is available yet*). This file will be located in the top-level directory of the dataset and will also list the contents of the other files and outline the file-naming convention used. This will add contextual value to the dataset for future reuse within and outside of the lab. 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? *Consult the interactive KU Leuven storage guide to find the most suitable storage solution for your data.*	☐ Shared network drive (J-drive) ☐ Personal network drive (I-drive) ☒ Teams ☐ Sharepoint online ☐ Sharepoint on-premis ☐ Large Volume Storage ☐ ManGO ☒ Digital vault ☐ Other:
How will the data be backed up? *WHAT STORAGE AND BACKUP PROCEDURES WILL BE IN PLACE TO PREVENT DATA LOSS?*	☐ Standard back-up provided by KU Leuven ICTS for my storage solution ☒ Personal back-ups I make (specify) ☒ Other (specify) Syncthing is being used to create an automated backup service both onsite and offsite, relying on personal machines from Prof. Pinheiro. This creates redundancy with data stored in multiple secure locations.
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 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.* *Guidance on security for research data*	OneDrive for business is secured using the KU Leuven data security systems and authentication. Syncthing creates a peer-to-peer encrypted network that must be authorized by both machines via a randomly-generated 128-bit key. Both ends must authorize the sharing of a folder before syncthing will allow data transfer.The network itself was set up in the Pinheiro lab as a spoke and hub model, with the spoke (individual researchers) only able to deposit data into the hub (onsite). The hub is backed up to an offsite secure server.
What are the expected costs for data storage and backup during the research project? How will these costs be covered?	OneDrive for business is free for researchers based at KU Leuven. Similarly, syncthing is free to use.

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...). *Guidance on data preservation*	☒ All data will be preserved for 10 years according to KU Leuven RDM policy ☐ All data will be preserved for 25 years according to CTC recommendations for clinical trials with medicinal products for human use and for clinical experiments on humans ☐ Certain data cannot be kept for 10 years (explain)
Where will these data be archived (stored and curated for the long-term)? *Dedicated data repositories are often the best place to preserve your data. Data not suitable for preservation in a repository can be stored using a KU Leuven storage solution, consult the interactive KU Leuven storage guide.*	☒ KU Leuven RDR ☐ Large Volume Storage (long-term for large volumes) ☐ Shared network drive (J-drive) ☒ Other (specify): Open Science Foundation
What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	At present, the amount of data generated falls within the free data storage allocation.

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, as open data ☐ Yes, as embargoed data (temporary restriction) ☒ Yes, as restricted data (upon approval, or institutional access only) ☐ No (closed access) ☐ Other, please specify: Where possible, all data will be kept open apart from datasets (NGS, DNA sequences, etc.) that may retain potential commercial value.
If access is restricted, please specify who will be able to access the data and under what conditions.	The only restricted data will be those of commercial value and therefore with suitable MTAs those data may be shared.
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.	☒ KU Leuven RDR ☒ Other data repository (specify): Open Science Foundation ☐ Other (specify)
When will the data be made available?	☒ Upon publication of research results ☐ Specific date (specify) ☐ Other (specify)

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.* *Check the RDR guidance on licences for data and software sources code or consult the License selector tool to help you choose.*	☒ CC-BY 4.0 (data) ☒ Data Transfer Agreement (restricted data) ☒ MIT licence (code) ☐ GNU GPL-3.0 (code) ☐ Other (specify)
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, a PID will be added upon deposit in a data repository ☐ My dataset already has a PID ☐ No
What are the expected costs for data sharing? How will these costs be covered?	Negligible.

7. Responsibilities	
Who will manage data documentation and metadata during the research project?	The researcher will be responsible for data documentation & metadata.
Who will manage data storage and backup during the research project?	Both the researcher (primary) and the supervisor (secondary) will be responsible for data storage & back up during the project.
Who will manage data preservation and sharing?	After completion of the project, the principal investigator / supervisor will bear the responsibility of ensuring data preservation and reuse.
Who will update and implement this DMP?	For the duration of the project, the researcher bears the responsibility of updating and implementing this DMP. The PI bears the end responsibility of updating and implementing the DMP beyond the project duration.