# 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 | **Simon Fiesack & https://orcid.org/0000-0002-1350-772X** |
| Contributor name(s) (+ ORCID) & roles | **Sibylle Vonesch (https://orcid.org/0000-0003-2485-1048) & Supervisor** |
| Project number [1] & title | 3E230769 & Developing a novel highly precise RNA targeting CRISPR tool to dissect genetic networks underlying phenotypic variation. |
| Funder(s) GrantID [2] | 11PT624N |
| Affiliation(s) | x KU Leuven<br>☐ Universiteit Antwerpen<br>☐ Universiteit Gent<br>☐ Universiteit Hasselt<br>☐ Vrije Universiteit Brussel<br>☐ Other:<br>ROR identifier KU Leuven: 05f950310 |
| Please provide a short project description | In this project, I will develop a highly specific, RNA targeting CRISPR based tool that allows reducing gene expression in a predictable way and can perturb multiple genes simultaneously. This approximates natural biological processes much more than knock-outs, and allows to study genetic interactions and essential genes. I will apply this tool to study phenotypic consequences of not only full but also partial loss of gene function, and to dissect the role of higher order gene interactions in shaping phenotype. This will increase our understanding of how genotype governs phenotype and open new possibilities in industrial, biotechnological and medical applications. |

[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.

## 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 |
| BIOLOGICAL MATERIAL | yeast libraries (*S. cerevisiae*) constructed through genome engineering, bacterial plasmids and plasmid libraries, Cas13d protein variant libraries, purified Cas13d protein. | ☒ Generate new data<br>☐ Reuse existing data | ☐ Digital<br>☒ Physical | ☐ Audiovisual<br>☐ Images<br>☐ Sound<br>☐ Numerical<br>☐ Textual<br>☐ Model<br>☐ Software<br>☐ Other: | | ☐ < 1 GB<br>☐ < 100 GB<br>☐ < 1 TB<br>☐ < 5 TB<br>☐ > 5 TB<br>☐ NA | Stored in Eppendorf's at -80°C as yeast and plasmid library pool in the lab |
| EXPERIMENTAL RESULTS | digital images, FACS data, sequencing data raw and processed, analysis scripts, software. | ☒ Generate new data<br>☐ Reuse existing data | ☒ Digital<br>☐ Physical | ☒ Images<br>☒ Numerical<br>☒ Textual<br>☒ Software | - gel scans, colony plate pictures, plots<br>- sorting/ analysis plots<br>- FASTQ, BAM, VCF, textfile | ☒ < 1 GB<br>☒ < 100 GB<br>☒ < 1 TB<br>☐ < 5 TB<br>☐ > 5 TB<br>☐ NA | |
| DATA REUSE | For analysis purposes we will use data from published datasets | ☒ Reuse existing data | ☒ Digital<br>☐ Physical | ☒ Numerical<br>☒ Textual<br>☐ Model<br>☒ Software<br>☐ Other: | - published datasets | ☒ < 1 GB<br>☒ < 100 GB<br>☒ < 1 TB | |

| | |
|---|---|
| *GUIDANCE:*<br>*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.*<br>*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. | Not yet applicable. |
| 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:<br>☐ Yes, animal data; provide ECD reference number:<br>☐ Yes, dual use; provide approval number:<br>☒ No<br>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)<br>☒ No<br>Additional information: |

[3] Add rows for each dataset you want to describe.
[4] See Glossary Flemish Standard Data Management Plan

| | |
|---|---|
| Does your work have potential for commercial valorization (e.g. tech transfer, for example spin-offs, commercial exploitation, …)?<br>If so, please comment per dataset or data type where appropriate. | ☒ Yes<br>☐ No<br>If yes, please comment:<br>Novel optimized Cas13d variants and the subsequent use of these variants as a highly specific RNA targeting CRISPR tool can result in potential intellectual properties. This will be decided in cooperation with VIB's IP Management team. |
| Do existing 3rd party agreements restrict exploitation or dissemination of the data you (re)use (e.g. Material/Data transfer agreements, research collaboration agreements)?<br>If so, please explain to what data they relate and what restrictions are in place. | ☐ Yes<br>☒ No<br>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?<br>If so, please explain to what data they relate and which restrictions will be asserted. | ☒ Yes<br>☐ No<br>If yes, please explain:<br>Novel optimized Cas13d variants and the subsequent use of these variants as a highly specific RNA targeting CRISPR tool can result in potential intellectual properties. This will be decided in cooperation with VIB's IP Management team. |

| 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).<br><br>*RDM guidance on documentation and metadata*. | BIOLOGICAL MATERIAL: Yeast strains are stored in a -80°C freezer and as yeast and plasmid library pool in the lab, for at least 10 years after the project ends. Costs are covered by general lab expenses. Unauthorized people do not have access to strains.<br><br>EXPERIMENTAL RESULTS: (Meta)data will be documented in lab notebooks and digital files will be stored in a Dropbox Business account, 256-bit AES and SSL/TLS encryption. Raw and processed sequencing data and any end values derived from these data will be stored on a server in an ordered structure, and a separate hard drive as third backup. All data will be stored for at least 10 years, conform KU Leuven RDM policy. |
| Will a metadata standard be used to make it easier to **find and reuse the data**?<br><br>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.<br><br>*REPOSITORIES COULD ASK TO DELIVER METADATA IN A CERTAIN FORMAT, WITH SPECIFIED ONTOLOGIES AND VOCABULARIES, I.E. STANDARD LISTS WITH UNIQUE IDENTIFIERS.* | ☐ Yes<br>☒ No<br>If yes, please specify (where appropriate per dataset or data type) which metadata standard will be used:<br><br>If no, please specify (where appropriate per dataset or data type) which metadata will be created:<br>Text documents and Excel files stored within each experiment folder will respectively contain guidelines describing data collection/analysis methods and all relevant metadata (including experimental conditions, quality control metrics, computational analysis pipelines and their parameters) to ensure the reusability of the data and the reproducibility of any further data generation. |

| 4. Data Storage & Back-up during the Research Project | |
|---|---|
| Where will the data be stored?<br><br>*Consult the interactive KU Leuven storage guide to find the most suitable storage solution for your data.* | ☐ Shared network drive (J-drive)<br>☐ Personal network drive (I-drive)<br>☒ OneDrive (KU Leuven)<br>☐ Sharepoint online<br>☐ Sharepoint on-premis<br>☐ Large Volume Storage<br>☐ Digital Vault<br>☒ Other:<br>- Dropbox Business account, 256-bit AES and SSL/TLS encryption. Raw and processed sequencing data and any end values derived from these data will be stored on a server in an ordered structure, and a separate hard drive as third backup. All data will be stored for at least 10 years, conform KU Leuven RDM policy.<br>- Yeast strains are stored in a -80°C freezer and as yeast and plasmid library pool in the lab, for at least 10 years after the project ends. Costs are covered by general lab expenses. Unauthorized people do not have access to strains. |
| How will the data be backed up?<br><br>*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<br>☐ Personal back-ups I make (specify)<br>☒ Other (specify)<br>- Dropbox Business account, 256-bit AES and SSL/TLS encryption. Raw and processed sequencing data and any end values derived from these data will be stored on a server in an ordered structure, and a separate hard drive as third backup. All data will be stored for at least 10 years, conform KU Leuven RDM policy.<br>- Yeast strains are stored in a -80°C freezer and as yeast and plasmid library pool in the lab, for at least 10 years after the project ends. Costs are covered by general lab expenses. Unauthorized people do not have access to strains. |

| | |
|---|---|
| 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<br>☐ No<br>- Dropbox Business account, 256-bit AES and SSL/TLS encryption. Raw and processed sequencing data and any end values derived from these data will be stored on a server in an ordered structure, and a separate hard drive as third backup. All data will be stored for at least 10 years, conform KU Leuven RDM policy.<br>- Yeast strains are stored in a -80°C freezer and as yeast and plasmid library pool in the lab, for at least 10 years after the project ends. Costs are covered by general lab expenses. Unauthorized people do not have access to strains. |
| How will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?<br><br>*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.*<br>*Guidance on security for research data* | - Dropbox Business account, 256-bit AES and SSL/TLS encryption. Raw and processed sequencing data and any end values derived from these data will be stored on a server in an ordered structure, and a separate hard drive as third backup. All data will be stored for at least 10 years, conform KU Leuven RDM policy.<br>- Yeast strains are stored in a -80°C freezer and as yeast and plasmid library pool in the lab, for at least 10 years after the project ends. Costs are covered by general lab expenses. Unauthorized people do not have access to strains. |
| What are the expected costs for data storage and backup during the research project? How will these costs be covered? | Costs are covered by general lab expenses. |

## 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...).<br><br>*Guidance on data preservation* | ☒ All data will be preserved for 10 years according to KU Leuven RDM policy<br>☐ 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<br>☐ Certain data cannot be kept for 10 years (explain) |
| Where will these data be archived (stored and curated for the long-term)?<br><br>*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<br>☐ Large Volume Storage (longterm for large volumes)<br>☐ Shared network drive (J-drive)<br>☒ Other (specifiy):<br>- Dropbox Business account, 256-bit AES and SSL/TLS encryption. Raw and processed sequencing data and any end values derived from these data will be stored on a server in an ordered structure, and a separate hard drive as third backup. All data will be stored for at least 10 years, conform KU Leuven RDM policy.<br>- Yeast strains are stored in a -80°C freezer and as yeast and plasmid library pool in the lab, for at least 10 years after the project ends. Costs are covered by general lab expenses. Unauthorized people do not have access to strains. |
| What are the expected costs for data preservation during the expected retention period? How will these costs be covered? | Costs are covered by general lab expenses. |

## 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.<br><br>*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<br>☒ Yes, as embargoed data (temporary restriction)<br>☐ Yes, as restricted data (upon approval, or institutional access only)<br>☒ No (closed access)<br>☒ Other, please specify:<br>I aim for 3 high impact manuscripts with the data generated in this project.<br>1) Sequence-function landscape of Cas13d.<br>2) Novel highly specific RNA targeting CRISPR tool.<br>3) Effect of higher order genetic interactions using pooled screens.<br><br>Conform the Open Access publication requirement for FWO, data used in published manuscripts will be openly available. |
| If access is restricted, please specify who will be able to access the data and under what conditions. | Only lab members can access the data. |
| 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<br>☒ Yes, intellectual property rights<br>☐ Yes, ethical aspects<br>☐ Yes, aspects of dual use<br>☐ Yes, other<br>☐ No<br><br>If yes, please specify:<br>Novel optimized Cas13d variants and the subsequent use of these variants as a highly specific RNA targeting CRISPR tool can result in potential intellectual properties. This will be decided in cooperation with VIB's IP Management team. |

| | |
|---|---|
| Where will the data be made available?<br>If already known, please provide a repository per dataset or data type. | ☐ KU Leuven RDR<br>☐ Other data repository (specify)<br>☐ Other (specify) |
| When will the data be made available? | ☒ Upon publication of research results<br>☐ Specific date (specify)<br>☐ Other (specify) |
| Which data usage licenses are you going to provide? If none, please explain why.<br><br>*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.*<br>*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)<br>☐ Data Transfer Agreement (restricted data)<br>☐ MIT licence (code)<br>☐ GNU GPL-3.0 (code)<br>☐ Other (specify) |
| Do you intend to add a PID/DOI/accession number to your dataset(s)? If already available, please provide it here.<br><br>*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<br>☐ My dataset already has a PID<br>☐ No |
| What are the expected costs for data sharing? How will these costs be covered? | **No extra costs** |

| 7. Responsibilities | |
|---|---|
| Who will manage data documentation and metadata during the research project? | **Simon Fiesack & Sibylle Vonesch, assisted by Célie Cokelaere (labmanager)** |
| Who will manage data storage and backup during the research project? | **Simon Fiesack & Sibylle Vonesch, assisted by Célie Cokelaere (labmanager)** |
| Who will manage data preservation and sharing? | **Simon Fiesack & Sibylle Vonesch, assisted by Célie Cokelaere (labmanager)** |
| Who will update and implement this DMP? | **Simon Fiesack** |