

## FWO DMP Template

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.

1. General Information	
Name applicant	Sarah Meyers
FWO Project Number & Title	11L9122N  Combining CRISPR screening and single-cell RNA sequencing to unravel the mechanisms of transcriptional deregulation in T-cell acute lymphoblastic leukemia
Affiliation	<input checked="" type="checkbox"/> KU Leuven <input type="checkbox"/> Universiteit Antwerpen <input type="checkbox"/> Universiteit Gent <input type="checkbox"/> Universiteit Hasselt <input type="checkbox"/> Vrije Universiteit Brussel <input checked="" type="checkbox"/> Other: VIB
2. Data description	
Will you generate/collect new data and/or make use of existing data?	<input checked="" type="checkbox"/> Generate new data <input type="checkbox"/> Reuse existing data

<p>Describe the origin, type and format of the data (per dataset) and its (estimated) volume</p> <p><i>If you <b>reuse</b> existing data, specify the <b>source</b> of these data.</i></p> <p><i>Distinguish data <b>types</b> (the kind of content) from data <b>formats</b> (the technical format).</i></p>	<p>This project will generate datasets of single-cell sequencing data of primary mouse T cells. As the experiments make use of CRISPR technology, we will generate both single-cell RNA-seq data and single-cell gRNA data. We expect to generate 3 to 4 datasets with an expected size of around 400GB per dataset (with 3 timepoints per experiment).</p> <p>Raw data consists of sequencing reads in .fastq format which will be stored in the secure archives of the Vlaams Supercomputer Centrum (VSC).</p> <p>Raw sequencing reads will be mapped using 10X Genomics CellRanger algorithm to generate .bam files together with count matrices (.mtx.gz and .tsv.gz format).</p> <p>Besides single-cell data, the project will also generate bulk sequencing data. For this bulk data, we will also store fastq and bam files.</p>
---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

### 3. Ethical and legal issues

<p>Will you use personal data? If so, shortly describe the kind of personal data you will use AND add the reference to your file in your host institution's privacy register.</p> <p><i>In case your host institution does not (yet) have a privacy register, a reference is not yet required of course; please add the reference once the privacy register is in place in your host institution.</i></p>	<p><input type="checkbox"/> Yes</p> <p><input checked="" type="checkbox"/> No</p> <p>If yes:</p> <ul style="list-style-type: none"> <li>- Privacy Registry Reference:</li> <li>- Short description of the kind of personal data that will be used:</li> </ul>
<p>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, add the reference to the formal approval by the relevant ethical review committee(s).</p>	<p><input checked="" type="checkbox"/> Yes</p> <p><input type="checkbox"/> No</p> <p>If yes:</p> <ul style="list-style-type: none"> <li>- Reference to ethical committee approval:</li> </ul> <p>We have approval of the Ethical Committee for Animal Experimentation (ECD) to use primary mouse cells from transgenic Bl/6 mice.</p> <p>ECD P-number: 023/2018</p>

<p>Does your work possibly result in research data with potential for tech transfer and valorisation? Will IP restrictions be claimed for the data you created? If so, for what data and which restrictions will be asserted?</p>	<p><input checked="" type="checkbox"/> Yes  <input type="checkbox"/> No  If yes, please comment:</p> <p>This is fundamental research executed on a cell model, which could result in patentable or non-patentable inventions with valorization potential. VIB maintains a dedicated technology transfer team to help translate research results into tangible applications. Results will be discussed with the tech transfer office of VIB and KU Leuven (LRD). In case tech transfer is possible, we will apply the necessary restrictions to the data to allow for a patent filing and subsequent valorization track, while ensuring that the publication of results and release of data is not unreasonably delayed.</p>
<p>Do existing 3<sup>rd</sup> party agreements restrict dissemination or exploitation of the data you (re)use? If so, to what data do they relate and what restrictions are in place?</p>	<p><input type="checkbox"/> Yes  <input checked="" type="checkbox"/> No  If yes, please comment:</p>

#### 4. Documentation and metadata

<p>What documentation will be provided to enable understanding and reuse of the data collected/generated in this project?</p>	<p>All data that will be published will be made publicly available with a clear description in the respective publication.</p> <p>Furthermore, we store the annotated scripts that were used during data analysis, with relevant Readme text files.</p>
-------------------------------------------------------------------------------------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Will a metadata standard be used? If so, describe in detail which standard will be used. If not, state in detail which metadata will be created to make the data easy/easier to find and reuse.	<input checked="" type="checkbox"/> Yes <input type="checkbox"/> No If yes, please specify:  The data will be annotated in concordance with the standards used in the public database to which it will be submitted, most likely either GEO from NCBI or the European EGA database.
-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

#### 5. Data storage & backup during the FWO project

Where will the data be stored?	<p>Data will be analysed at the Vlaams Supercomputer Centrum (VSC), and will be stored in the secure VSC archives for at least 5 years after the end of the research.</p> <p>Published data will be stored in a public database (e.g.: EGA or GEO) and will be publicly available after publication.</p>
How will the data be backed up?	<p>Data will be archived on the HPC of the VSC with the necessary security and redundancy measures.</p> <p>Furthermore, regular backups of the computers are created for safe storage of any additional data.</p>

<p>Is there currently sufficient storage &amp; 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.</p>	<p><input checked="" type="checkbox"/> Yes  <input type="checkbox"/> No  If no, please specify:</p> <p>The lab has sufficient resources and funding to acquire the necessary storage capacity on the HPC. The estimated volume of data is 400GB per single-cell RNA-seq experiment. We expect to generate 3 to 4 of such datasets, resulting in a total of 1,5 to 2TB.</p> <p>Furthermore, there is sufficient storage and back-up capacity on all KU Leuven servers:</p> <ul style="list-style-type: none"> <li>- the “L-drive” is an easily scalable system, built from General Parallel File System (GPFS) cluster with NetApp eseries storage systems, and a CTDB samba cluster in the front-end.</li> <li>- the “J-drive” is based on a cluster of NetApp FAS8040 controlers with an Ontap 9.1P9 operating system.</li> </ul>
<p>What are the expected costs for data storage and backup during the project? How will these costs be covered?</p> <p><i>Although FWO has no earmarked budget at its disposal to support correct research data management, FWO allows for part of <b>the allocated project budget</b> to be used to cover the cost incurred.</i></p>	<p>The cost of data storage on the HPC of the VSC is €70 per TB per year, and this cost can be covered by the lab.</p>
<p>Data security: how will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?</p>	<p>The HPC has all necessary security measures to ensure safe data storage. Only the persons involved in the project will have access to the data (Sarah Meyers, Sofie Demeyer, Jan Cools).</p>

## 6. Data preservation after the end of the FWO project

FWO expects that data generated during the project are retained for a period of minimally 5 years after the end of the project, in as far as legal and contractual agreements allow.

Which data will be retained for the expected 5 year period after the end of the project? In case only a selection of the data can/will be preserved, clearly state the reasons for this (legal or contractual restrictions, physical preservation issues, ...).	<p>All RNA sequencing data that was created during this project will be retained for at least 5 years after the end of the project. The data will be stored in the form of raw reads as fastq files and the mapped bam files. In case there would not be sufficient storage capacity, only the raw fastq files will be stored and bam files will be deleted, as those can be regenerated by mapping the raw data again.</p> <p>Plasmids that were created will be stored securely in the -80°C storage in the Jan Cools lab, and a description of each plasmid and the vector map (.clc file) will be kept in a database on KU Leuven and VIB servers. Cell lines will be cryopreserved in a liquid nitrogen biobank, with the correct annotations and any additional information stored on KULEuven servers.</p>
Where will these data be archived (= stored for the long term)?	All data will be archived on the HPC of the VSC with the necessary security and redundancy measures.
<p>What are the expected costs for data preservation during these 5 years? How will the costs be covered?</p> <p><i>Although FWO has no earmarked budget at its disposal to support correct research data management, FWO allows for part of <b>the allocated project budget</b> to be used to cover the cost incurred.</i></p>	The cost of data preservation is € 70 per TB per year, which will be covered by the host lab.

## 7. Data sharing and reuse

<p>Are there any factors restricting or preventing the sharing of (some of) the data (e.g. as defined in an agreement with a 3<sup>rd</sup> party, legal restrictions)?</p>	<p><input type="checkbox"/> Yes  <input checked="" type="checkbox"/> No          If yes, please specify:</p> <p>There are no restrictions on data sharing, as we do not have any agreements with 3<sup>rd</sup> party organisations and the project does not involve sensitive personal information.</p>
<p>Which data will be made available after the end of the project?</p>	<p>Published data will be made freely available.</p> <p>Unpublished data will be stored securely after the end of the project but will not be made publicly available.</p>
<p>Where/how will the data be made available for reuse?</p>	<p><input checked="" type="checkbox"/> In an Open Access repository  <input type="checkbox"/> In a restricted access repository  <input checked="" type="checkbox"/> Upon request by mail  <input type="checkbox"/> Other (specify):</p> <p>Data will be shared through publication, and stored in public data repositories such as GEO and EGA database.</p>
<p>When will the data be made available?</p>	<p>Data will be made available upon publication.</p>
<p>Who will be able to access the data and under what conditions?</p>	<p>Published data will be freely available for research purposes. For commercial purposes, a data access agreement should be signed.</p>
<p>What are the expected costs for data sharing?          How will these costs be covered?</p> <p><i>Although FWO has no earmarked budget at its disposal to support correct research data management, FWO allows for part of <b>the allocated project budget</b> to be used to cover the cost incurred.</i></p>	<p>There are no additional resources needed for data sharing in public databases.</p>

## 8. Responsibilities

Who will be responsible for the data documentation & metadata?	Sarah Meyers and the co-supervisor Sofie Demeyer are responsible for documentation during the course of the project.
Who will be responsible for data storage & back up during the project?	Sarah Meyers and the co-supervisor Sofie Demeyer are responsible for safe storage of data during the course of the project.
Who will be responsible for ensuring data preservation and sharing?	Jan Cools and Sofie Demeyer are responsible for data sharing and preservation.
Who bears the end responsibility for updating & implementing this DMP?  <i>Default response: The PI bears the overall responsibility for updating &amp; implementing this DMP</i>	The PI, Jan Cools, bears the overall responsibility for updating & implementing this DMP.