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	**Daan Dierickx (0000-0002-8917-022X)**
Contributor name(s) (+ ORCID) & roles	**Jan Cools (0000-0001-6626-5843), head Laboratory for Molecular Biology of Leukemia** **Marlies Vanden Bempt (0000-0003-0111-0263), postdoc Laboratory Experimental Hematology** **Flore Sneyers (0000-0003-0329-7973), postdoc Laboratory Experimental Hematology**
Project number [1] & title	Molecular characterization of rare lymphomas to uncover new biomarkers and treatment targets
Funder(s) GrantID [2]	18B5824N
Affiliation(s)	x 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	**Posttransplant lymphoproliferative disorders (PTLD)** have high mortality and morbidity rates, implying a need for improved preventive tools, identifying patients at risk. We aim to investigate biomarkers that would identify patient at risk and also predict response to treatment. In parallel, as existing therapies are often toxic or have increased risk for graft rejection, we plan to investigate new therapeutic targets to improve outcome. In **WP1**, starting from whole genome sequencing data**,** we will explore mechanisms of both EBV^+ and EBV^- DLBCL-PTLD aiming to provide an integrated insight in the genomic and biological landscape of both EBV^+ and EBV^- PTLD, identifying (1) predictive and prognostic biomarkers and (2) new therapeutic targets. To gain a better understanding of the underlying **Peripheral T cell lymphoma (PTCL)** biology, we have started to fully characterize the genetic and transcriptomic landscape of PTCL using combined whole genome and transcriptome sequencing of a cohort of 28 clinical PTCL cases. This effort led to the identification of the novel FYN-TRAF3IP2 fusion gene and other new genetic aberrations. In line with previous studies, we found that the mutational landscape of PTCL is heterogeneous, with many genetic alterations occurring at low frequencies. Interestingly, approximately half of the patients presented with a genetic abnormality in one or more components of the T cell receptor (TCR) signaling pathway. Therefore, deregulated TCR signaling may act as a driver in PTCL pathogenesis, paving the way for novel targeted therapies applicable to a larger subset of PTCL patients. In addition, mutations in epigenetic factors are frequently found in PTCL, indicating an important driver role for epigenetic deregulation. In **WP2** and **WP3** we aim to investigate how the different newly identified TCR signaling and epigenetic mutations contribute to T cell lymphoma 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 [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
		☐ 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	
WGS PTCL and PTLD	Whole genome sequencing data from PTCL and PTLD cases	Reuse existing data (PTCL) and generate new data (PTLD)	Digital	Textual		>5 TB	
CUT&TAG / CUT&RUN	PTCL patient chromatin profiling on selected PTCL cases	Generate new data	Digital	Textual		<1 TB	
Optical	Optical genome	Generate new data	Digital	Textual		<100 GB	

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

genome mapping	mapping dataset from selected PTCL and PTLD cases						
UMI 4C	4C on engineered cell lines	Generate new data	Digital	Textual		<100 GB	
Cell lines ChIP	Chromatin profiling on engineered cell lines	Generate new data	Digital	Textual		<100 GB	
Mouse survival data + phenotypic information	Data on the mice that developed disease	Generate new data	Digital	Textual + numerical	Xlsx, doc	< 1 GB	
NGS analysis mice	RNA-seq, ChIP-seq, ATAC-seq on mouse models	Generate new data	Digital	Textual			
Treatment data mouse model	Data from treatment experiments ex vivo and in vivo	Generate new data	Digital	Textual + numerical	xlsx, doc, fcs	< 1 GB	
BLI treatment mouse models	Bioluminescent imaging during treatment experiments	Generate new data	Digital	Images	TIFF	< 100 GB	
CRISPR screen	Sequencing data from CRISPR	Generate new data	Digital	Textual		< 100 GB	

	screen						

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.	The existing PTCL patient data we are using is partially described in this publication: https://www.nature.com/articles/s41467-021-24037-4
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: S6120 ☒ Yes, animal data; provide ECD reference number: 031/2023 ☐ 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): S6120 ☐ No Additional information:

[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, …)? If so, please comment per dataset or data type where appropriate.	☒ Yes ☐ No If yes, please comment: we might identify novel drug targets for the treatment of PTCL and PTLD from our sequencing data. We will discuss the data and potential commercial valorization with LRD at KU Leuven to determine the possibilities for tech transfer. We will work with them to determine a publication plan to ensure that publication does not affect the tech transfer possibilities.
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

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 dataset will be accompanied by a detailed excel file and text file explaining how the experiment was performed (samples used, oncogenes used, cell culture conditions, amounts of cells used, RNA/protein isolation methods, purification methods, antibodies used, gRNAs included in the screen, meaning of the different labels used in the dataset).
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: For the next generation sequencing data, we will follow the recommendations of the KU Leuven genomics core facility. MIAME guidelines will be followed: https://www.ncbi.nlm.nih.gov/geo/info/MIAME.html 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) ☒ OneDrive (KU Leuven) ☐ Sharepoint online ☐ Sharepoint on-premis ☒ Large Volume Storage ☐ Digital Vault ☒ Other: VSC
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)
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: We pay yearly for storage space at VSC and KU Leuven. ☐ 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*	secure login (2 factor authorization login)

What are the expected costs for data storage and backup during the research project? How will these costs be covered?	70 Euro per TB per year. We have budgeted the costs for data storage (especially for large sequencing data files) on the fund for lymphoma research, which is managed by prof. Daan Dierickx.

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 (longterm for large volumes) ☐ Shared network drive (J-drive) ☒ Other (specify): next-generation sequencing data is deposited at GEO.
What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	We have budgeted the costs for data storage (especially for large sequencing data files) on the fund for lymphoma research, which is managed by prof. Daan Dierickx. Data storage at GEO is free.

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:
If access is restricted, please specify who will be able to access the data and under what conditions.	For sequencing data from primary human samples, we will work with a data transfer agreement.
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): GEO ☐ 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): the mouse and cell line data will be made available without license. For the human data, we will make the data available under a data transfer agreement.
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?	70 Euro per TB per year. These costs can be covered by our consumable costs. There are no costs for GEO.

7. Responsibilities	
Who will manage data documentation and metadata during the research project?	Marlies Vanden Bempt, Flore Sneyers, Sofie Demeyer
Who will manage data storage and backup during the research project?	Marlies Vanden Bempt, Flore Sneyers, Sofie Demeyer
Who will manage data preservation and sharing?	Marlies Vanden Bempt, Flore Sneyers, Sofie Demeyer, Jan Cools, Daan Dierickx
Who will update and implement this DMP?	Marlies Vanden Bempt, Flore Sneyers, Daan Dierickx