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	**Anthony Sezirahiga, https://orcid.org/0000-0002-4319-7873**
Contributor name(s) (+ ORCID) & roles	**Peter Carmeliet, https://orcid.org/0000-0001-7961-1821, supervisor**
Project number [1] & title	11Q0K24N, Immunosuppressive mystery genes in immunomodulatory endothelial cells as overlooked targets for alternative immunotherapy for cancer
Funder(s) GrantID [2]	FWO (11Q0K24N)
Affiliation(s)	KU Leuven ROR identifier KU Leuven: 05f950310
Please provide a short project description	Anti-cancer Immunotherapy (IT) has huge potential but faces challenges of insufficient efficacy and resistance, in part due to the immunosuppressive (ImmuSup) nature of the endothelial cells (ECs) lining the tumor vasculature. I aim to tune these ECs from ImmuSup to immunostimulatory as potential alternative IT (alterIT). I fully realize that (too) many basic research results fail to translate into drugs/clinical treatments due to selectivity and toxicity issues following suboptimal initial target selection and validation. Therefore, I will 1) use the in-house developed artificial intelligence (AI)-based scMysterYdentifier tool to predict truly novel ImmuSup genes in the 'mystery genome' (lacking functional annotation, 1/3 of the genome), a goldmine for innovative target discovery; 2) Prioritize the predicted candidates using a text-mining-based approach to interrogate their known features; 3) directly validate the prioritized ImmuSup mystery genes in vivo by generating EC-selective knockdown mice rapidly (days)/inexpensively using a new in-house developed gene knockdown REVOLT technology; 4) perform elaborate mechanistic studies to delineate their molecular mode-of-action. All approaches are operational. Expected outcome and perspectives: thoroughly in vivo validated innovative and EC-selective ImmuSup mystery targets for alterIT, including insight in their mechanisms of action.

[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
		☐ 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	
TEXT	Protocols, description of research results, literature studies, output of predicted mystery genes	New	Digital	Textual	.csv, .docx	< 1 GB	
Gene prioritization soft-	Gene prioritization tool + pre-	- Reuse (public databases; software)	Digital	- Software (UniApp)	.jpg, .pdf	< 1 GB	

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

ware (UniApp, Unicle Biomedical Datascience) & Public databases	liminary inquiry into mechanistic understanding of candidate targets	- New (gene prioritization outcome & predictions on mechanistic actions)		- Textual - Experimental			
Biological samples	Cells from mice and patient samples	New	Physical	Experimental			Storage in N2 tanks
Fluorescent and confocal microscopy images	Confocal microscopy images of tumor samples	New	Digital	Images, Experimental	.czi, .tif, .jpg	< 1 TB	
RNAseq	RNAseq on tumor samples and cultured cell lines upon target silencing	New	Digital	Other (transcriptomic)	.bam, .fastq, .count, .txt, .csv	< 1 TB	
Flow cytometry data	Flow Cytometry and FACS sort files (FlowJo and equipment specific files)	New	Digital	Other (experimental)	.fcs, .pdf	< 1 GB	
Observational numerical data	Characterization of molecular mode of action	New	Digital	Numerical	.xls, .csv, .pzfx	< 100 GB	

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.	Public databases will be consulted during this project. These include: • https://www.genecards.org/ • https://www.proteinatlas.org/ • https://maayanlab.cloud/Harmonizome/dataset/Biocarta+Pathways • …
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: S57123 ☒ Yes, animal data; provide ECD reference number: P115/2023 ☐ Yes, dual use; provide approval number: ☐ No Additional information: Yes, animal experiments will be done (WP2, ECD: P115/2023), as well as immunostainings on human samples, (WP3, S57123). In case human tissue will be used, relevant data will be handled according to the principles of the General Data Protection Regulation (GDPR) 2016/679 and the Belgian privacy law, and approval by the Ethics Committee Research UZ/KU Leuven will be applied for new type of experiments.
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:

[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: Deep mechanistic studies will be performed on the top mystery target to uncover its immunosuppressive properties on a molecular level. Not only will this provide us fundamental knowledge to publish in a top journal, but it also provides the most promising mystery target with potential to be translated to clinical applications. Submission of patents will be evaluated in collaboration with VIB.
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: Human tumor samples are provided by UZ Leuven to the host laboratory under an MTA. The MTA stipulates that provider and recipient agree to negotiate a joint ownership agreement which shall specify rights and obligations of each Party related to the use, exploitation, and protection of patentable or non-patentable joint results.
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: Human tumor samples are provided by UZ Leuven to the host laboratory under an MTA. The MTA stipulates that provider and recipient agree to negotiate a joint ownership agreement which shall specify rights and obligations of each Party related to the use, exploitation, and protection of patentable or non-patentable joint results.

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*.	Generated sequencing data for this project will be uploaded to EGA and GEO in combination with related metadata (e.g. age, gender, case/control status, sequencing platform/library... etc.) to be accessible to the public upon submission. A physical sample inventory will be stored in freezers (plasmids, vectors, RNA and protein extracts) and liquid nitrogen tank (cells) and a file with sample details will be saved on the shared server. Flow cytometry and sorting: information on gating strategy for cell identification and sorting will be saved in electronic files with details on antibody concentrations and protocols for cell preparation and staining will be described in detail in lab books. Imaging (confocal and cytation5 imager): images and settings will be saved in electronic files. Details on staining techniques and antibody or dye concentrations and protocols for cell preparation will be described in detail in lab books. Datafiles and the imaging protocols will be stored on KULeuven servers. Excel documents will always be saved in the csv format, so that it can be read as American csv/tab- delimited text or European csv/tab-delimited text. For publication, the standards of the journal in which the data will be published, will be used. For all stored data, a readme-file is provided, which includes a short description of the filename, definitions of column headings and row labels, data processing steps, storage information and contact information.

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: The metadata standards of EGA will be used for submission of sequencing data, as can be consulted on https://egaarchive.org/submission/sequence/unaligned

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: L-drive (lvs.inetloc) KU Leuven LUNA, centrally managed by the central computer, IT department of KU Leuven: ICTS (Informatie en Communicatie: Technologie en Systemen), Archive.inetloc, Synology Imaging server, data can also be stored at the institutional research data repository: RDR from KU Leuven (after completion of the project).

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) The data will be backed up in a double way. Automatic back-up (every 24 hours) of the network L-drive is controlled by the ICTS KU Leuven department. In addition, every researcher's computer has installed the Druva Cloud Platform. Druva Cloud protects and manages data across all devices, and allows to perform the backup operations even every 5 minutes (managed individually - depends on the user).
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 KU Leuven provides sufficient storage and back-up capacity during and after the project. A dedicated folder will be made for the project on which the collaborators will work jointly and store data files.
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*	The network drive for the FWO project folder and the large volume storage folder are secured by the ICTS service of KU Leuven with a mirror copy. Only other lab members, will have access to the shared folder. Unauthorized persons do not have access to this system.
What are the expected costs for data storage and backup during the research project? How will these costs be covered?	Yearly storage costs of 1TB data on large storage servers of the host lab are estimated at 130 €/year. Costs will be covered by internal lab fundings.

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 (specifiy): Archive.inetloc All the generated data will be stored and archived on the "large storage network L-drive" - KU Leuven LUNA, centrally managed by the central computer, IT department of KU Leuven. All data is backed-up daily to the cloud-storage to ensure safe storage.
What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	Yearly storage costs of 1TB data on large storage servers of the host lab are estimated at 130 €/year. Costs will be covered by internal lab fundings.

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: Embargoed data: data with temporary access restriction until publication (e.g.RNA-sequencing datasets) Restricted data: data/material under MTA agreement (physical data, e.g. newly constructed plasmids, mouse samples)
If access is restricted, please specify who will be able to access the data and under what conditions.	All relevant data will made publicly available upon publication. However, before publication, the data will be accessible only by the researchers working on the project. The identity of the person who accesses the data will be verified using institutional account system. If for any reason of accession is needed at an earlier time-point, this can be arranged through collaborations and in cooperation with the host institution's guidance. All third parties will be able to access data under restriction, under MTA with VIB.
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: Some data will become available when published. Further, some mystery targets may be considered for patent applications. However, this will be decided internally upon discussion with VIB and KU Leuven/LRD, to decide which candidate targets will be considered for this purpose.

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) ☒ Other (specify) VIB
When will the data be made available?	☒ Upon publication of research results ☐ Specific date (specify) ☒ Other (specify): For Restricted data: upon finalization of an MTA agreement with VIB.
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): Material Transfer Agreement (restricted data)
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?	We do not expect any costs for data sharing to publicly available repositories.

7. Responsibilities	
Who will manage data documentation and metadata during the research project?	Anthony Sezirahiga, Peter Carmeliet (promotor)
Who will manage data storage and backup during the research project?	Anthony Sezirahiga and departmental IT staff (Urbain Scherpereel, Pieter Joris) and the ICTS KU Leuven department
Who will manage data preservation and sharing?	While the project is ongoing, the doctoral researcher (Anthony Sezirahiga) will manage the data preservation. Prof. Peter Carmeliet, the promoter, will take care of the preservation after the completion of the doctoral dissertation, together with the departmental IT staff (Urbain Scherpereel, Pieter Joris). The researcher will manage the sharing of the data.
Who will update and implement this DMP?	Anthony Sezirahiga