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	**Chantal Mathieu & 0000-0002-4055-5233**
Contributor name(s) (+ ORCID) & roles	**Conny Gysemans (0000-0003-3559-6089) & co-promotor**
Project number [1] & title	G059123N & Neutrophils: neglected as potential players in autoimmune type 1 diabetes initiation, progression, and perpetuation
Funder(s) GrantID [2]	
Affiliation(s)	☒ KU Leuven ☐ Universiteit Antwerpen ☐ Universiteit Gent ☐ Universiteit Hasselt ☐ Vrije Universiteit Brussel ☐ Other: ROR identifier KU Leuven: 05f950310
Please provide a short project description	While the role of innate immune cells such macrophages and dendritic cells in the initiation and perpetuation of autoimmune type 1 diabetes (T1D) have been well documented, the particular actions of neutrophils remain unfairly overlooked in the aetiology of this chronic disease. As neutrophils can influence cells of the innate and adaptive immune system, a better characterization of the neutrophil population during disease development is needed and could provide novel insights into whether neutrophils have divergent developmental branches or different maturation sequences influenced by tissue environmental cues. Here, we aim to identify distinct neutrophil populations during T1D development in the non-obese diabetic (NOD) mouse model using state-of-the-art technologies (i.e., multiparameter flow cytometry, single-cell CITE-sequencing, and spatial transcriptomics) and exploit perceptions, obtained via data integration, in the design of neutrophil-related immunotherapeutics to delay or prevent T1D onset. As emerging data indicate that neutrophil extracellular traps are implicated in T1D development, we aim to study the therapeutic potential of small-molecule isozyme-specific PAD4 and TGM2 inhibitors, which catalyse protein modifications known to be involved in certain types of autoimmunity, in delaying or preventing T1D onset in mice. If this strategy proves successful, these PAD4 and TGM2 blockers could be powerful therapies in human individuals at-risk of developing T1D.

[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
WP1: Study the phenotypic heterogeneity and functional versatility of neutrophils during T1D development in the NOD mouse model	droplet-based CITE-sequencing	☒ Generate new data	☒ Digital	☐ Audiovisual ☐ Images ☐ Sound ☒ Numerical ☒ Textual ☐ Model ☒ Software ☐ Other:	.R .RDS .GMT	☐ < 1 GB ☐ < 100 GB ☐ < 1 TB ☒ < 5 TB ☐ > 5 TB ☐ NA	
	Spectral multi-color flow cytometry	☒ Generate new data	☒ Digital	☐ Audiovisual ☐ Images ☐ Sound ☒ Numerical ☐ Textual ☐ Model ☒ Software ☐ Other:	.FCS .WSP .EXDAT .prism	☐ < 1 GB ☐ < 100 GB ☒ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	

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

	spatial transcriptomics	☒ Generate new data	☒ Digital	□ Audiovisual ☒ Images □ Sound ☒ Numerical □ Textual □ Model ☒ Software □ Other:	.R	□ < 1 GB □ < 100 GB □ < 1 TB ☒ < 5 TB □ > 5 TB □ NA	
	INNODIA humans at-risk of developing T1D subjects, and T1D patients	☒ Reuse existing data	☒ Digital	□ Audiovisual □ Images □ Sound ☒ Numerical □ Textual □ Model □ Software □ Other:		□ < 1 GB □ < 100 GB ☒ < 1 TB □ < 5 TB □ > 5 TB □ NA	
WP2: Study the impact of PAD4 and TGM2 inhibition on neutrophil characteristics and T1D development in the NOD mouse model	Mouse data	☒ Generate new data	☒ Physical	□ Audiovisual □ Images □ Sound ☒ Numerical □ Textual □ Model □ Software □ Other:	.doc .excel .prism	☒ < 1 GB □ < 100 GB □ < 1 TB □ < 5 TB □ > 5 TB □ NA	Storage of serum, plasma and/or tissue in ultrafreezers and liquid nitrogen tank (cryotheek)
	Spectral multi-color flow cytometry	☒ Generate new data	☒ Digital	□ Audiovisual □ Images □ Sound	.FCS	☐ < 1 GB	

				☒ Numerical ☐ Textual ☐ Model ☒ Software ☐ Other:	.WSP .EXDAT .prism	☐ < 100 GB ☒ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	
	IncuCyte	☒ Generate new data	☒ Digital	☒ Audiovisual ☒ Images ☐ Sound ☒ Numerical ☐ Textual ☐ Model ☒ Software ☐ Other:	.ZGF .MP4 .prism	☐ < 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.	Primary data: experimental data (from single-cell omics, from mouse and human samples), and compiled and integrated data (from experimental data generated during this project). Reuse of existing data: human PBMC samples and database of INNODIA consortium on clinical parameters of human samples plus use part of our own previously published data, analyzed in a new different context. These data will be used for publication in high-ranked international peer-reviewed journals and for patent filing in case we find robust neutrophil-related biomarkers for disease initiation, progression, and perpetuation as well as therapy response.

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: S62381 ☒ Yes, animal data; provide ECD reference number: 125/2018; 139/2018; 229:2018 ☐ 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: S62381
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:
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: The research agreement with GSK on the use of the TGase2i is still under construction with KU Leuven. As backup, we have access to TGase2 ko NOD mice and rederived PAD4 ko NO mice to accompany the PAD4i studies.
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:

[4] See Glossary Flemish Standard Data Management Plan

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*.	Raw experimental data will be collected per experimental test and will include a README.txt file with a clear description of what the data represent and how they were generated. Each individual file will contain information on the study design, the origin of the samples, and all necessary information for an independent analyst to use or reuse the data. This description will be documented in page-numbered lab notebooks as well as in electronic format. The lab also uses SOP (.pdf) accompanying the raw experimental data. The lab has document (.pdf) with overview of all SOP (different versions and updates). Analysed data (e.g.: graphs, tables, texts, power point presentations etc.) will be stored in folders containing the raw and processed data files they are referring to. These folders are organized per project. File formats will be .docx, .pdf, .RData, .jpg, .tiff, .png, .csv, etc. Inclusion of dates will indicate the different version of specific file. Programming languages and code are text-based format and will provide an overview of the necessary packages and libraries in the datasets. Data analysis methods and particularities (including metadata) will be described in text documents and Excel files included in these folders. All files will be stored in the KU Leuven J- or L-drives with sharing possibilities via One Drive (managed by the KU Leuven IT department). Several students working with single-cell omics have followed the VIB course on GitLab in which changes in files are trackable and managed automatically especially code reviews, sharing code snippets etc. are possible.

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: If no, please specify (where appropriate per dataset or data type) which metadata will be created: Text documents and Excel files stored within each experiment folder in the J- and L-drives will respectively contain guidelines describing data collection/analysis methods and all relevant metadata (including experimental conditions, sample keys, computational analysis pipelines and their parameters) to ensure the reusability of the data and the reproducibility of any further data generation. For data on human blood samples the clinical study number will be included; for data on pre-clinical mouse experiments the type of mice will be included.

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:

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 ☐ No We have sufficient storage and backup capacity both on J (55,296 Gb of which 6,216 free space) and L (5,000 GB of which 2,160 free space) drive. We can easily request for addition storage capacity. 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*	For paper notebooks: Office doors are always locked when researchers are out of the office. For digital files: all data on J- and L-drives are stored in password protected drives that are only accessible by people from the PIs laboratories. dr. Gysemans is responsible for allowing people access to these drives.
What are the expected costs for data storage and backup during the research project? How will these costs be covered?	Costs for data storage are incorporated in the requested FWO funding. Our J- fand L-drive have a current capacity of 54 TB and of 5 TB respectively. The annual cost of L-drive storage is 569 € per 5 TB of storage space per year. This cost and capacity include the performance of mirror copies of the stored data, for safety backup purposes. We expect that 1 TB will be sufficient to store all data generated as part of the project. These costs will be covered by the budget of the project lead (Prof. dr. Chantal Mathieu).

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) Our university's data management policy expects that relevant research data generated are retained for a period of minimally 10 years after the end of the project, in a safe, secure & sustainable way for purposes of reproducibility, verification, and potential reuse. However, for biological samples it is not always possible to keep them for 10 years since the long-term stability of some biological samples has not been established. Publication data will be further organized and catalogued on a figure-by-figure basis for future reference to raw datasets used for figure generation.
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 (specifiy):
What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	Our J- and L-drive have a current capacity of 54 Tb and of 5 Tb respectively. The annual cost of L-drive storage is 569 € per 5TB of storage space per year. This cost and capacity include the performance of mirror copies of the stored data, for safety backup purposes. We expect that 1 TB will be sufficient to store all data generated as part of the project. These costs will be covered by the budget of the project lead (Prof. Chantal Mathieu).

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.	All data will be generated and collected within the Mathieu lab. Data may be shared externally upon reasonable requests from collaborating scientists, which will be reviewed and approved on a case-by-case basis by the project lead. Single cell omics data are mostly deposited in open access repositories upon publication.
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) ☐ Other (specify) Relevant raw data will at that same moment be made available in well-established open-access data repositories.
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?	Costs for data sharing will be discussed with collaborators on a case-by-case basis.

7. Responsibilities	
Who will manage data documentation and metadata during the research project?	Students and technicians involved in the project will be responsible for data documentation.
Who will manage data storage and backup during the research project?	Students and technicians will have the daily responsibility of recording all data (i.e., digital, paper and biological samples). They will also be responsible for the correct and accurate data entry and recording of the metadata.
Who will manage data preservation and sharing?	dr. Conny Gysemans is responsible for the storage (J- and L-) drives of the Mathieu lab. She will ensure data preservation and reuse.
Who will update and implement this DMP?	The PIs bear the end responsibility of updating & implementing this DMP.