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	**Prof. Dr. Jef Verbeek** http://orcid.org/0000-0002-1549-8003
Contributor name(s) (+ ORCID) & roles	
Project number [1] & title	**G0ADH24N** De rol van B cellen en immunoglobulines in alcohol-gerelateerd leverlijden.
Funder(s) GrantID [2]	**G0ADH24N FWO**
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	Approximately half of all deaths due to end-stage liver disease (cirrhosis) are alcohol-related. Currently, no effective drug is available. Therefore, increased insight in how alcohol leads to liver injury is needed. An important role has been proposed for intestinal bacteria in developing alcohol-associated liver disease (ALD). Due to increased alcohol-induced gut damage, gut bacteria and their products leak into circulation and travel to the liver where they induce inflammation and injury. In order to protect against invading pathogens, our body produces antibodies (also called immunoglobulins) that exist in various types and forms that can recognize, bind and clear foreign material. Interestingly, ALD patients have increased antibody titers in their blood compared to healthy individuals, indicating altered immune response towards specific substances. Yet, their contribution to ALD remains unclear. We hypothesize that increased antibody levels in circulation during ALD reflect altered immunity against gut-derived molecules that propagate liver injury. Hence, we aim to characterize antibodies, their targets and the cells that produce them (B cells) in blood and liver of ALD patients. Moreover, we will test their functional role during ALD in appropriate mouse models. Taken together, these studies will result in novel insights in the disease mechanisms by which alcohol use results in liver disease and in the identification of novel targets for therapy.

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	
Sequencing data	Single cell sequencing data, B-cell receptor sequencing data. PhIP sequencing data	Generate new data	Digital	Numerical	.fastq, .rds, .loom, .py, . bam, .csv	<5TB	
Flow cytometry data	Frequency and number of B cell subsets by flow cytometry	Generate new data	Digital	Numerical	.fcs, .xlsx	<100GB	

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

ELISA data	Quantification of total serum IgM, IgG, IgE, IgA by ELISA	Generate new data	Digital	Numerical	.xlsx	<100GB	
Imaging data	IHC data	Generate new data	Digital	Images	.xlsx, .jpg, .png, .czi	<100GB	
Mass spectrometry data	Hepatic immunoglobulin antigen immune complexes by mass spectrometry	Generate new data	Digital	Numerical	.xlsx	<100GB	
Patient characteristics	Patient characteristics	Generate new data	Digital	Textual and numerical	.cvs (Redcap)	<100GB	
Samples	Tissue samples, blood samples	Generate new data	Physical				Frozen samples: Tubes stored at -80°C. Tissue for histology: fixed and stored at 4 °C

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.	
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: S64744 ☒ Yes, animal data; provide ECD reference number: Will solely be performed in Vienna by collaborator (cfr. FWO project) ☐ 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: Health & genetic data. Data will be used as approved in S64744.
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: There is a potential that the data (transcriptomics & spatial transcriptomics) may generate novel concepts that could be further valorized. However, this is realistically not within the scope of the next 4 years.
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: We do have a MTA agreement with VIB (Prof. Diether Lambrechts) regarding the transcriptomic data, however this is not restrictive to exploitation or dissemination of the data since we have a full collaboration and shared controllership between contributing parties.

[4] See Glossary Flemish Standard Data Management Plan

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*.	1. For the sequencing & transcriptomic data the following information will be noted: sample, patient, tissue, timepoint. Protocol used. Every folder will contain (in its name), the patient, date and timepoint. 2. For the flowcytometry data: sample, patient, tissue, timepoint. Protocol used. A text file will be maintained including all the information as mentioned in the MIFlowCyt. Every folder will contain (in its name), the patient, date and timepoint. 3. For the ELISA, mass spectrometry & IHC analyses data: sample, patient, tissue, timepoint. Protocol used. Per experiment a txt file will be included of what the data represent and how they were generated. Every folder will contain (in its name), the patient, date and timepoint. 4. For the patient information: all the patient data will be ordered by patient and by timepoint.

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 yes, please specify (where appropriate per dataset or data type) which metadata standard will be used: Metadata standards will be used for genomics (http://www.dcc.ac.uk/resources/metadata-standards/genome-metadata). For all other data, metadata will be created using the Dublin core (http://www.dcc.ac.uk/resources/metadatastandards/dublin-core). 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: K-drive (large raw data files)

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 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*	Researchers involved in the project can control who they give access to the files on their personal OneDrive. To access the KU Leuven servers, access is provided and controlled by the research manager, Hannelie Korf. The KU Leuven ICTS data center hosts the network storage, with a mirror available in the second ICTS center. This ensures additional back-up capacity, recovery of lost data and long term data availability. The access is controlled by KU Leuven security groups and it is password protected. Data related to Patients and controls are pseudonymized, which is managed by the research manager, Hannelie Korf. The list linking the anonymized code with patient records (name, EMDNR number, clinical data) is stored on the hepatology common J-drive, and is locked under password protection. Password access if provided for read-only access to the students and technicians who are working directly with the samples. Hannelie Korf retains the sole ability to modify this file under alternative password protection. All samples are processed and stored as per their anonymized code only.
What are the expected costs for data storage and backup during the research project? How will these costs be covered?	Personal and Shared network drive (I, J) 503,66 euro/TB per year. All large primary data will be moved to LVS. LVS storage costs per 5 Tb (KU Leuven ICTS): 104,42euro/year. Expected amount of data (50TB). The costs have been budgeted on the grant.

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): Digital data will be stored at the Archive (K:) server from KU Leuven ICTS. HCR probes and physical samples will be stored in the freezers from the laboratory of Hepatology. Code scripts will be stored on Github.
What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	We expect the costs to gradually increase up to 3000 euro/year. After the project, data preservation costs will be covered by other grants.

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.	Since the data consists of genetic and health data, the data will only be shared under restrictions. The Informed consent form allows the sharing of pseudonymised data, however, the pseudonymisation key will not be shared. Sharing of data will only be possible after agreement of the DAC.
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: Since the data consists of genetic and health data, the data will only be shared under restrictions.

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): Experimental data will be made available through a data repository such as EGA, ncbi, github, Genbank, FigShare (https://figshare.com/), Dryad (https://datadryad.org/) or https://zenodo.org/depending on the type of data. We will explore the possibilities via online repositories and will use the website www.re3data.org
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?	The transfer costs depend on the data repository selected. Costs will be covered by the project funding.

7. Responsibilities	
Who will manage data documentation and metadata during the research project?	PhD student (to be hired), supervisor Prof. Verbeek
Who will manage data storage and backup during the research project?	PhD student (to be hired), supervisor Prof. Verbeek
Who will manage data preservation and sharing?	PhD student (to be hired), supervisor Prof. Verbeek
Who will update and implement this DMP?	PhD student (to be hired), supervisor Prof. Verbeek