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	**Astrid De Roover 0000-0003-0321-2896**
Contributor name(s) (+ ORCID) & roles	Not applicable
Project number [1] & title	Unravelling the impact of peroxisomal lipid biology on H3K79 methylation and osteoarthritis
Funder(s) GrantID [2]	FWO 12AMS24N
Affiliation(s)	KU Leuven - ROR identifier KU Leuven: 05f950310
Please provide a short project description	In osteoarthritis (OA), the most common joint disease that still lacks a disease-modifying therapy, methylation of lysine 79 at histone 3 (H3K79me) in articular cartilage is reduced. Evidence showed that maintaining H3K79me is crucial to protect against OA. Yet, mechanisms that control H3K79me remain largely elusive. My project aims to address this knowledge gap. During my PhD, I successfully performed an advanced high-content microscopy siRNA screening to identify H3K79me regulators. The results suggested an intriguing link between peroxisomal lipid metabolism and H3K79me, a novel and potentially groundbreaking finding. With my exciting preliminary data, I hypothesize that targeting peroxisomal lipid biology could restore deficient H3K79me in OA and have therapeutic effects. In this project, I aim to unravel for the first time (1) the impact of peroxisomal lipid biology on H3K79me and OA, and the underlying molecular mechanisms, (2) Whether targeting peroxisomal lipid biology can counteract H3K79me deficiency in OA and protect against the disease. To achieve this, I will use patient-derived materials, well-established molecular techniques and OA mouse models, combined with cutting-edge technologies like CRISPR-Cas9 genome editing, lipidomics and single-cell epigenetic CyTOF approaches. A successful project will advance our knowledge of OA pathogenesis and epigenetics, and will contribute to the development of an effective therapy for this widespread disease.

[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
Cellular models of cartilage health and disease	Ex vivo/in vitro assays include quantitative PCR, Western blot analysis, Chromatin Immuno-precipitation, lipidomics and colorimetric matrix assays. The dataset also includes the statistical analysis.	☒ Generate new data ☐ Reuse existing data	☒ Digital ☒ Physical	☐ Audiovisual ☒ Images ☐ Sound ☒ Numerical ☒ Textual ☐ Model ☐ Software ☐ Other:	- Primary raw and processed data: xlsx and csv files, TIFF-files - Statistical analysis: R-code, Graphpad files and associated txt or PDF files	☐ < 1 GB ☒ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	Not applicable
Mouse models of osteoarthritis	Mouse model analysis includes histological scores, radiographic imaging (micro-CT), immunohistochemical analysis, pain assessments, quantitative PCR and Western blot analysis	☒ Generate new data ☐ Reuse existing data	☒ Digital ☒ Physical	☐ Audiovisual ☒ Images ☐ Sound ☒ Numerical ☒ Textual ☐ Model ☐ Software ☐ Other:	- Primary raw and processed data: xlsx and csv files, TIFF-files - Statistical analysis: R-code, Graphpad files and associated txt or PDF files	☐ < 1 GB ☒ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	20 histology slide boxes (storage in the Skeletal Biology and Engineering Research Centre).

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

Bioformatic analysis of hits siRNA screening	Bioinformatic analysis of primary hits using HumanBase and String-db softwares to organize the data into biologically relevant networks and identify the top enriched pathways and key nodes within complex networks	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☒ Images ☐ Sound ☒ Numerical ☒ Textual ☐ Model ☐ Software ☐ Other:	- Primary raw and processed data: xlsx and csv files, TIFF-files - Statistical analysis: R-code, Graphpad files and associated txt or PDF files	☐ < 1 GB ☒ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	Not applicable

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.	We are reusing existing data of a targeted large-scale siRNA screening generated within our team before the start of the current project under supervision of the leading investigator.
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: S56271 ☒ Yes, animal data; provide ECD reference number: P004/2022 ☐ Yes, dual use; provide approval number: ☐ No Additional information: Preliminary data collected under ECD approvals P114-2008 - P198-2012, P159-2016, P018-2017

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:
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: The discoveries made within the project, taken together from the different datasets, could have potential for intellectual property protection and subsequent valorisation by licensing agreements or spin-off activity. Such valorization will not only be dependent on the discoveries itself, but also on the support and priority setting that can be provided by Leuven R&D.
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

[4] See Glossary Flemish Standard Data Management Plan

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*.	Our lab team is using lab notebooks that are kept within the laboratory. Data storage as described above is both digital (quantitative data, protocols, data analysis, images) as well as physical (histology specimens). Lab notebooks are chronological. Datasets are annotated by the investigators performing the experiments and this information is contained within the digital storage environment. Contextual and descriptive features of the data are included within the written and digital data records both at the level of a dataset (e.g. describing how the data were created), but also at the level of individual data elements (e.g. explaining what each variable means or the parameters for generation of datafiles such as images). The following documentation will be provided: (1) a table of content (excel file and csv) with all project-related experiments including experiment number, date of implementation and name of the researcher who stored the experiment, (2) a brief description of the goal of the experiment and related work package (word and txt file), (3) a detailed protocol or link to an existing standard protocol (SOP) which will enable other researcher to repeat the experiment, (4) all data or link to another file with the (raw) data, (5) in case of animal work: a list of the used animals with details such as age, sex, housing and link with LAIS system information, (5) samples that are generated during the experiments and will be stored and listed in a csv file, (6) if appropriate, illustrations of the data with legends and statistical analysis. In case that documentation is written or available in notebooks or stored on other files a link will be provided. (7) Read-me text files providing information about definitions used in the dataset files. With the help of these documentations every authorized researcher will be able (1) to look up all the information of the performed experiments and (2) to repeat the experiment in exactly the same way.
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 siRNA screening: We aim to use ScreenSifter as metadata standard (https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-14-290) 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: Box (KU Leuven)
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) : version back-up on KU Leuven Box and OneDrive
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*	All data will be stored in a protected KU Leuven environment. Research data can only be accessed by a login following KU Leuven's policy for identifier and with password and double authentication.
What are the expected costs for data storage and backup during the research project? How will these costs be covered?	Current pricing for the KU Leuven Shared Network Drive is € 503,66 / TB / year. Datasets for this project are considered to require less than 50 GB. The costs both in short and long-term are covered by the project.

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):
What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	Current pricing for the KU Leuven Shared Network Drive is € 503,66 / TB / year. Datasets for this project are considered to require less than 50 GB. The costs both in short and long-term are covered by the project and the lab's historical financial resources. Upon publication the datasets will be included in the KU Leuven Research Data Repository (RDR).

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 : Cellular models, mouse models and bioinformatics analysis data used in publication. ☒ Yes, as embargoed data (temporary restriction): Cellular models, mouse models and bioinformatics analysis data not yet published. ☒ Yes, as restricted data (upon approval, or institutional access only): Cellular models, mouse models and bioinformatics analysis data not yet published. ☒ No (closed access): Cellular models, mouse models and bioinformatics analysis data not yet published. ☐ Other, please specify:
If access is restricted, please specify who will be able to access the data and under what conditions.	Project leaders (Prof. S. Monteagudo & Prof. R. Lories). Backup options will be departmental chair and departmental manager.

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 may be used for valorisation and will – in that case – not be made fully public
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 datasets (NIH – see above) – Lipidmaps (NIH – see above) ☐ 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)

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?	50 GB per year per author of the dataset in Leuven RDR are free. Hence we do not foresee any financial burden to share our data via this repository

7. Responsibilities	
Who will manage data documentation and metadata during the research project?	Astrid De Roover - Silvia Monteagudo (PI) – Rik Lories (co-lab director) – Frederique Cornelis (lab-manager)
Who will manage data storage and backup during the research project?	Astrid De Roover - Silvia Monteagudo (PI) – Rik Lories (co-lab director) – Frederique Cornelis (lab-manager)
Who will manage data preservation and sharing?	Silvia Monteagudo (PI) – Rik Lories (co-lab director) – Frederique Cornelis (lab-manager)
Who will update and implement this DMP?	Astrid De Roover - Silvia Monteagudo (PI)