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	**Mohammad Sarfi & 0000-0003-2724-1528**
Contributor name(s) (+ ORCID) & roles	
Project number [1] & title	**11A5G25N & Fighting liver cancer heterogeneity and aggressiveness by inhibiting Hippo and Oxytocin Receptor signaling**
Funder(s) GrantID [2]	**Fonds voor Wetenschappelijk Onderzoek - Research Foundation Flanders (FWO)**
Affiliation(s)	■ 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	Liver cancer represents a highly heterogeneous group of malignancies, including hepatocellular carcinoma (HCC), cholangiocarcinoma (CCA), and tumors with mixed epithelial and mesenchymal phenotypes. This heterogeneity contributes significantly to tumor aggressiveness and resistance to current therapies, yet the mechanisms driving cellular plasticity in liver cancer remain poorly understood. To address this, my colleagues and I developed a novel genetic mouse model of liver cancer in which cholangiocarcinoma cells spontaneously dedifferentiate and acquire mesenchymal characteristics without the need for external stimuli. This in situ transformation closely resembles human liver cancer progression and contrasts with earlier in vitro or xenograft models that rely on artificial induction. Our model recapitulates key histological features of human CCA, including desmoplastic stroma and tumor cell populations coexpressing epithelial and mesenchymal markers. Importantly, we observed that the mesenchymal phenotype arises through transdifferentiation of epithelial cancer cells rather than recruitment of mesenchymal cells from the surrounding tissue. To investigate the molecular regulators of this process, we conducted a targeted genetic screen and identified the Hippo pathway effectors YAP and TAZ, as well as the oxytocin receptor (OXTR), as critical drivers of tumor progression and transdifferentiation. Notably, deletion of these factors impaired the epithelial-to-mesenchymal transition and the development of aggressive tumor features. These findings suggest that YAP/TAZ and OXTR are key mediators of cancer cell plasticity rather than mere promoters of proliferation. My ongoing project aims to elucidate the mechanistic roles of these factors in driving transdifferentiation, providing essential insights into the biology of aggressive liver cancers and potential therapeutic targets.

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
Microscopy images	Confocal microscopy images of mouse livers	Generate new data	Digital	Images	Tiff, OIB, jpeg, czi	<100GB	-
Flow cytometry	Flow cytometry of purified mouse liver cells	Generate new data	Digital	Other	.fcs	1GB	-
Sequencing files	Single cells or bulk sequencing will be performed at KU Leuven core facilities	Generate new data	digital	other	.fastq, .cloupe	100-500GB	-
DNA vectors	DNA vectors will be designed by SnapeGene	Generate new data	Digital	Other	.dna	200MB	-

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

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

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.	NA
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: ☒ Yes, animal data; provide ECD reference number: 251/2024 ☐ Yes, dual use; provide approval number: ☐ No Additional information: All experiments on mice are conducted according to institutional, national and European animal regulations. We already have ethical clearance files approved by the animal ethics committee at the KU Leuven. (KUL ECD project number 251/2024). For some MPs ethical approval will be requested once needed.
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: NA

[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 do not exclude that the proposed work could result in research data with potential for tech transfer and valorization. As the promoter of this grant is a member of VIB, VIB has a policy to actively monitor research data for such potential. If there is substantial potential, the invention will be thoroughly assessed, and in a number of cases the invention will be IP protected (mostly patent protection or copyright protection). As such, the IP protection does not withhold the research data from being made public. In the case, a decision is taken to file a patent application it will be planned so that publications need not be delayed.
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*.	**All experiments will be performed from an experimental design document that details the different conditions, and the metadata linked to the experiment. For all experiment both a hard copy and electronic copy will be maintained. The shared resources (animal, cell line, DNA, siRNA, sgRNA, shRNA) will be saved on electronic lab Notebook ELN. The experimental approach will follow a Standard Operating Procedure (SOP) which is also shared on L-drive.**
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: *No specific metadata standard will be used. Following metadata (non-limiting list) will be recorded, when applicable, for every experiment:* *-date* *-type of experiment (in vivo, in vitro, computational modelling, bio-informatics)* *-source of data (mouse, database)* *-type of material generated (DNA, protein, digital)* *-location of physical material (digital database)* *-device and details of program used* *-compounds, formulation and application* *-applicable ethical form*

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) ☐ Teams ☐ Sharepoint online ☐ Sharepoint on-premis ☐ Large Volume Storage ☐ ManGO ☐ Digital vault ☒ Other: *The data will be stored on the university's central servers with automatic daily back-up procedures.*
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 stored on the university's central servers with automatic daily back-up procedures.*
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 *Storage space on the J-drive and L-drive are unlimited and can be expanded upon simple request. Upon closure of the project (or upon request), all data will be transferred to the archive on the K-drive for long term storage. Data placed on the K-drive is more strictly secured with only very specific members of the lab. Only the head of the lab has the authority to have data removed by the IT department.* 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*	*The following measurements are taking to ensure secure data storage and to prevent modification by unauthorized persons: controlled physical access to the building, firewalling (on both departmental and individual server levels), encrypted communications, network compartmentalization & MAC authentication, least-known ports for well-known services, bruteforce intrusion detection & isolation, individual account expiry in accordance with contract of employment, ACL's (Access Control Lists).*
What are the expected costs for data storage and backup during the research project? How will these costs be covered?	L-drive, 128,38 euro/TB/year in blocks of 5TB K-drive, 128,38 euro/TB/year in blocks of 5TB J-drive, 519 euro/TB/year

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): Hard copy notebooks will be archived in the host institute's building. Digital data will be archived in team folders on the storage (L-Drive). Upon closure of the project, all digital data will be transferred to the archive on the K-drive for long term storage. Data placed on the K-drive is more strictly secured with only very specific members of the lab having the authority to place data on the K-drive and only the head of lab has the authority to have data removed by the IT department. The physical data in the form DNA vectors will be cryopreserved in the lab or at the KULeuven biobank.
What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	Digital storage on the K-drive is expected to cost 1200 euro per year (at current rates) for 10TB. Cost of physical storage at the KULeuven Biobank is currently still unclear. Costs of storage will be carried by the labs future grants and VIB dotation.

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: All data that are not under IP. Data, with the exception of personal/otherwise confidential data, are available upon publication of the results. All omics data and code & scripts will be made publicly available. Other data can be made available upon request.
If access is restricted, please specify who will be able to access the data and under what conditions.	Published data are accessible to all. The full dataset will be uploaded in their respective databases as an open access dataset under a CC-BY license. Therefore, it will be available to anyone for any purpose, provided that they give appropriate credit to the creators. Unpublished data that are considered confidential will only be accessible to the partners. For data shared directly by the PI, a material transfer agreement (and a non-disclosure agreement if applicable) will be concluded with the beneficiaries in order to clearly describe the types of reuse that are permitted. Data under IP will not be shared with peers.
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: No restricting factors foreseen but sharing of data might be temporarily postponed due to IP potential.

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) In an Open Access repository or upon request by mail
When will the data be made available?	☒ Upon publication of research results ☐ Specific date (specify) ☐ Other (specify) Confidential data will be kept in house.
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) Published data are accessible to all. The full dataset will be uploaded in their respective databases as an open access dataset under a CC-BY license.
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?	Most of the data are put without costs on public repositories. Peers can use the data at no cost under the condition of co-authorship. Commercial organizations will have to pay a fee that will be determined by the Legal departments of the universities involved in this project.

7. Responsibilities	
Who will manage data documentation and metadata during the research project?	**Mohammad Sarfi**
Who will manage data storage and backup during the research project?	**Mohammad Sarfi and IT support KU Leuven**
Who will manage data preservation and sharing?	**The host institute's IT team is responsible for digital preservation**
Who will update and implement this DMP?	**The PI bears the end responsibility of updating and implementing this DMP**