DATA MANAGEMENT PLAN (FWO DMP)

Xiaozheng Liu

DMP_1231025N

ADMIN DETAILS

Project Name: Targeting the metabolism of lung resident cells to inhibit metastasis formation - DMP_1231025N
Grant title: 1231025N
Principal Investigator / Researcher: Prof. Sarah-Maria Fendt / Xiaozheng Liu
Project data contact: Prof. Sarah-Maria Fendt
Institution: KU Leuven

1. GENERAL INFORMATION

Name applicant
Sarah-Maria Fendt
Xiaozheng Liu

FWO Project Number & Title
1231025N Targeting the metabolism of lung resident cells to inhibit metastasis formation

Affiliation

- KU Leuven

KU Leuven-VIB Center for Cancer Biology

2. DATA DESCRIPTION

Will you generate/collect new data and/or make use of existing data?

- Generating new data
- Using existing data

Describe in detail the origin, type and format of the data (per dataset) and its (estimated) volume. This may be easiest in a table (see example) or as a data flow and per WP or objective of the project.

Type of data	Format	Volume	How created
Microscopy images	.tif	1-10MB	All digital images from 2D/3D tissue cell cultures, taken with microscopes

Mass-spectrometry (LC/GC-MS) data	.ms	14MB/sample	Data generated by GC-MS and/or LC-MS
Excel files	.xls	various volume	Collecting numeric data of multiple experiments in excel formats
Luminescence and absorbance data collected in excel files	.xls	20KB - 200KB	Data collected from assays will be kept within excel files
H&E and immunohistochemistry (IHC) /immunofluorescent (IF) staining of mice lungs/liver	.qptiff	100GB	High quality scans of H&E / IHC/ IF staining of mice lungs and liver
Written notes	paper based	/	Written notes of experiments and experimental set-up
Other *in vivo* mouse data	.xls, .jpg, .png, .ms	10GB	Body weight, blood metabolite measurements, metabolomic data, tumor burden, etc. are recorded manually and recorded in Excel files. Images of mouse tissues will be taken with a digital camera and saved as .jpg image files.
Gene expression	.xls	<1GB	Gene expression from RNA extracted from cells/tissue are measured using PCR with data exported as an Excel file.
Statistical analysis	.pdf, .png, .wmf, .svg, .pzfx	5 GB	Prism8 (GraphPad) is used for all statistical analyses, producing program-specific files (.pzfx). Final figure forms are exported from this program in the following file types: .pdf, .png, .wmf, .svg.
(single cell)RNA sequencing data and analyses	.fastq, .mtx, .bam, .R, .RData, .txt, .rnk, .html, .xlsx, .png	100GB	Sequencing data file (.fastq), count matrix (.mtx) and genome-aligned (.bam) files will be generated. Data analysis will be conducted in R, generating R script files (.R) with analysis output (.RData, and .txt). Gene set enrichment analysis (GSEA) will be conducted, requiring libraries of pathways (.gmt) and normalized transcriptome reads in a ranked format (.rnk). The data will be will visualized in .bedGraph and .tdf files. Output files are produced in various formats (.html, .xlsx, .png images).

3. LEGAL AND ETHICAL ISSUES

Will you use personal data? If so, shortly describe the kind of personal data you will use. Add the reference to your file in KU Leuven's Register of Data Processing for Research and Public Service

Purposes (PRET application). Be aware that registering the fact that you process personal data is a legal obligation.

- Yes

Clinical data from patients will be accessed through the UZ Leuven biobanks and will only contain information on the age of the patient, initials and a personal identification number as well as health data (e.g. description of characteristics of physical features of the body, medical history and medical test information (such as blood samples results from scans and biopsies)). GDPR approval was obtained before applying for ethical approval at the KU/UZ Leuven EC by Prof. Christine Desmedt. The approval is available and outlined in EC S64410.

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, add the reference to the formal approval by the relevant ethical review committee(s)

- Yes

In the project, we perform in vivo experiments with the use of mice. All animal experiments have been approved by the Ethical Committee for Animal Experimentation (ECD) at KU Leuven. Below the three ethical approvals used in the project can be found:

Ethical approval projects	
	P007/2020
	P048/2020
	P025/2020

Does your work possibly result in research data with potential for tech transfer and valorisation? Will IP restrictions be claimed for the data you created? If so, for what data and which restrictions will be asserted?

- Yes

IP arising from this work is managed as per the framework agreement between the VIB (VIB Tech Transfer) and the KUL, the two participating institutes in this study.

Do existing 3rd party agreements restrict dissemination or exploitation of the data you (re)use? If so, to what data do they relate and what restrictions are in place?

- Yes

Yes, as above, dissemination or exploitation of the data is managed according to the framework agreement between the VIB and the KUL.

4. DOCUMENTATION AND METADATA

What documentation will be provided to enable reuse of the data collected/generated in this project?

All methods and details of data collection and experiments will be recorded. Notes in lab books and transcription to Word or excel files will be conducted by lead postdoctoral researcher of the project (Xiaozheng Liu) and - possibly - also by a lab technician (Janine Theile). All experiments will be labelled according a specific numbering/code and will be given the adequate file folders containing raw data of each experiment. These raw data files will contain the needed information about data collection and used methods, the analysis and normalization methods. All files will be stored on the KU Leuven L-Drive and OneDrive, with sharing possibilities (managed by KU Leuven).

Will a metadata standard be used? If so, describe in detail which standard will be used. If no, state in detail which metadata will be created to make the data easy/easier to find and reuse.

- No

For metabolomic analysis, the lab has internal metadata standards describing the collection and processing of the metabolites. Guides are included in the lab L-drive to describe the data collection and specific analysis methods of GC/LC-MS data. All files will be labelled and organized to be easily interpreted in the future.

5. DATA STORAGE AND BACKUP DURING THE FWO PROJECT

Where will the data be stored?

How is backup of the data provided?

Data saved on the L-drive and OneDrive, both provided by KU Leuven, is managed and backed up by the KU Leuven IT services.

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

Yes, the L-drive and OneDrive have sufficient storage capacity for the outlined project.

What are the expected costs for data storage and back up during the project? How will these costs be covered?

Annual costs of the L-drive consist of around 475.7 euros each year to have a total of 5 TB storage. This cost and capacity include the performance of mirror copies of the stored data, for safety backup purposes. We expect that 5TB will be more than sufficient to store all data generated as part of the project. These costs will be covered by the budget of the project lead (Prof. Sarah-Maria Fendt).

Data security: how will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?
Personal computers and lab-computers are protected via password access, set up by KU Leuven IT and changed annually. Off-site access to the L-drive is possible through KU Leuven data access points and is also password protected. Files can only be modified by lab-members that are allowed to access the L-drive folders from the lab of Prof. Sarah-Maria Fendt.

6. DATA PRESERVATION AFTER THE FWO PROJECT

Which data will be retained for the expected 5 year period after the end of the project? In case only a selection of the data can/will be preserved, clearly state the reasons for this (legal or contractual restrictions, physical preservation issues, ...).
All data will be retained for the expected 5-year period after the end of the project. The data will be stored on the university's central servers (L-drive with automatic back-up procedures and access right management), conforming with the KU Leuven research data management policy.

Where will the data be archived (= stored for the longer term)?
All raw data will be stored on the KU Leuven L-Drive until publication, upon which the data will be transferred to the KU Leuven K-drive Archive storage.

What are the expected costs for data preservation during the retention period of 5 years? How will the costs be covered?
The annual cost of L-Drive storage is 475.7 € per 5TB of storage space per year. We expect that 5 TB will be more than sufficient for long-term storage of all data generated as part of the project. Upon publication, the data will be reorganized by figure and stored on the KU Leuven Archive K-drive. The annual cost of the K-drive is around 47.57€ per 1TB. We believe 1-2TB should be sufficient to for the archiving of all data related to the project. These costs will be covered by the budget of the project lead (Prof. Sarah-Maria Fendt).

7. DATA SHARING AND REUSE

Are there any factors restricting or preventing the sharing of (some of) the data (e.g. as defined in an agreement with a 3rd party, legal restrictions)?

- Yes, Intellectual Property Rights
- Yes, Ethical aspects

Intellectual Property arising from this work is managed as per the framework agreement between the VIB (VIB Tech Transfer) and the KU Leuven, the two participating institutes in this study. Patient data will be made available in restricted access repositories. Requests for access to the patient data will be assessed on a case-by-case basis by the Data Access Committee, KU/UZ Leuven and the project lead (Prof. Sarah-Maria Fendt).

Which data will be made available after the end of the project?
At the moment, the generated raw data will not be made publicly available. The journal articles and its accompanying key data will be made available upon completing the project.

Where/how will the data be made available for reuse?

- In an Open Access repository
- In a restricted access repository
- Upon request by mail

Upon publication of the research results, the full datasets will be made available upon request. All other data that is not added to the publication, is kept open in the hard drive of the lab. This way, all lab member can access the data at any time.
Any other requests for sharing datasets will be evaluated on a case-by-case basis by Prof. Sarah-Maria Fendt herself.

When will the data be made available?

- Upon publication of the research results

Upon publication of the results, the summarized datasets and key findings with their interpretation will be made available in open access through the publication of a journal article.
At the end of the project, remaining data will be made available to the rest of the lab members.

Who will be able to access the data and under what conditions?
The summarized dataset of published results will be uploaded in a .cvs format as an open access dataset, upon request of the non-predatory academic journal. Therefore, it will be available to anyone for any purpose, provided that they give appropriate credit to the creators.
All other datasets will be made available to members of the lab of Prof. Sarah-Maria Fendt and the PI herself.

What are the expected costs for data sharing? How will the costs be covered?
Costs for data sharing will be discussed with collaborators on a case-by-case basis. To minimize data management costs, free-to-use data repositories will be used when possible. Data management will be covered by own funding.

8. RESPONSIBILITIES

Who will be responsible for data documentation & metadata?
For data documentation and metadata, Prof. Sarah-Maria Fendt accepts responsibility. Xiaozheng Liu will be responsible for experimental data.

Who will be responsible for data storage & back up during the project?

Xiaozheng Liu will be primarily responsible for collecting, generating and storing all data and documentation. KU Leuven IT services are responsible for maintenance and back-up of all drives used for data storage.

Who will be responsible for ensuring data preservation and reuse ?
For the data preservation and reuse, our principal investigator Prof. Sarah-Maria Fendt accepts responsibility. Also, our local IT-manager (Urbain Schepereel) and the ICTS KU Leuven department are taking care of accurate preservation of all data.

Who bears the end responsibility for updating & implementing this DMP?
The PI - Prof. Sarah-Maria Fendt - bears the end responsibility of updating & implementing this DMP.