FWO DMP Template

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.

1. General Information	
Name applicant	**Jakub Ksiazkiewicz**
FWO Project Number & Title	**Niche-specific BMP-SMAD responses in lymphatic endothelium (11M4922N)**
Affiliation	☒ KU Leuven ☐ Universiteit Antwerpen ☐ Universiteit Gent ☐ Universiteit Hasselt ☐ Vrije Universiteit Brussel ☐ Other:
2. Data description	
Will you generate/collect new data and/or make use of existing data?	☒ Generate new data ☒ Reuse existing data

Describe the origin, type and format of the data (per dataset) and its (estimated) volume *If you **reuse** existing data, specify the **source** of these data.* *Distinguish data **types** (the kind of content) from data **formats** (the technical format).*	**Generated data:** **Single-cell sequencing data:** **Raw scRNAseq data will be generated by SMART sequencing. Those files contain the information about the transcriptome of individual cells within the sample (expression levels of all genes). The exact size of files is hard to describe and is dependent on number of cells sequenced. The formats of generated raw files id FastQ. Filtered data format can vary between formats like .csv or .mtx depending on the method used. We plan to analyse the sequencing data using R language. The code will be stored as an R script object.** **qPCR data:** **Files in RDML format will be generated for every gene and sample of interest. Those files contain the information about the amount of certain mRNA being produced by the cell, and thus indirectly about gene expression.** **FACS:** **Files containing information about FACS results will be stored in .fcs format. Files are generated by software used by FACS device and carry information about number of cells, their size or presence of certain antigens.** **Microscopy:** **Experiments that involve microscopy will generate images containing results of e.g. immunofluorescence or immunohistochemistry (both giving information about spatial distribution of proteins within tissue). Those files will be stored as .zvi (Zeiss vision image) format as it contains complete metadata and is the most suitable for analysis using Zeiss microscopy software, as well as, Fiji software routinely used by our laboratory. Size of the files can vary from tens to hundreds of megabytes depending on characteristics of the photo.** **Reused data:** **We will reuse already generated and filtered single-cell RNA sequencing data to compare the results of**

	our experiment with existing knowledge. The data will be collected from open access databases like NCBI. The format of files will vary as there is no general standard in which the data is shared. The most common are .csv and .mtx formats. Filtered data files are usually not large – less than 1GB in size.

3. Ethical and legal issues	
Will you use personal data? If so, shortly describe the kind of personal data you will use AND add the reference to your file in your host institution's privacy register. *In case your host institution does not (yet) have a privacy register, a reference is not yet required of course; please add the reference once the privacy register is in place in your host institution.*	☐ Yes ☒ No If yes: - Privacy Registry Reference: - Short description of the kind of personal data that will be used:
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 ☐ No If yes: - Reference to ethical committee approval: **The project is already approved by the Ethical committee for animal experimentation (ECD) of KU Leuven and given the project number of 062/2022.**
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 ☒ No If yes, please comment:
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 ☒ No If yes, please comment:

4. Documentation and metadata	
What documentation will be provided to enable understanding and reuse of the data collected/generated in this project?	**We will provide documentation in form of scientific publications, posters and presentations. We plan to upload the raw data on platforms like Zenodo, as well as storing raw sequencing data on open access databases like GEO. Additionally we will publish the protocols used to obtain the data to ensure future reproducibility.**
Will a metadata standard be used? If so, describe in detail which standard will be used. If not, state in detail which metadata will be created to make the data easy/easier to find and reuse.	☐ Yes ☒ No If yes, please specify: **Every digital dataset will contain metadata describing the characteristic of the process used to generate the data. Sequencing files contain metadata about sequencing conditions. Similarly metadata of microscopy files allows for tracking the conditions in which photo was taken. Additionally, any analogous material will be described and stored with its location saved within database/lab-book for easy localization. Samples are going to be described with great detail to ensure accurate identification.**

5. Data storage & backup during the FWO project	
Where will the data be stored?	**Internally digital data will be stored in a network drives supported and secured by KU Leuven. Additionally, like mentioned above, the raw data will be shared within open access online databases after publication. Analog data (in the form of tissues, staining slides) will be stored in conditions required by a specific dataset (room temperature, 4°C, -20°C or -80°C). Data in the form of lab-books and protocols will be stored accordingly.**
How will the data be backed up?	**Network drives managed by KU Leuven are constantly being backed up allowing the user to restore any files that may be accidentally lost. Additionally, uploading the files in above mentioned online databases provides additional backup in case of emergency.**

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: **We have currently sufficient storage available to store the data generated by throughout the project on network drives (>500GB) managed by KU Leuven. However, if the drive capacity will not be enough, we certainly will expand our drive volume.**
What are the expected costs for data storage and backup during the project? How will these costs be covered? *Although FWO has no earmarked budget at its disposal to support correct research data management, FWO allows for part of **the allocated project budget** to be used to cover the cost incurred.*	**Cost of storage on internal network drives of KU Leuven is around 500€ per 1Tb yearly. We expect to generate about 100GB of data per year on average. This will cost about 50€ per year and will be paid by the laboratory. However, if the size of the data will drastically exceed our expectations we will consider using a part of bench-fee to cover this expense.**
Data security: how will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?	**Drives on which the data will be stored are accessible only after gaining permission by a specific KU Leuven account with a two-step authentication via mobile device. This ensures the security and limits the access by a third-party people to an absolute minimum.**

6. Data preservation after the end of the FWO project FWO expects that data generated during the project are retained for a period of minimally 5 years after the end of the project, in as far as legal and contractual agreements allow.	
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, ...).	**We aim to retain all the digital data for more than 5 years after the end of the project. However, the analogue data – e.g. staining slides - are not suitable for long-term storage as they can degrade over time.**
Where will these data be archived (= stored for the long term)?	**The data will be stored long-term on a dedicated archive drive within KU Leuven network. Additionally, certain data will be placed within online databases mentioned above.**

What are the expected costs for data preservation during these 5 years? How will the costs be covered? *Although FWO has no earmarked budget at its disposal to support correct research data management, FWO allows for part of **the allocated project budget** to be used to cover the cost incurred.*	**Costs of archives containing large datasets like sequencing data or imaging is estimated to be around 900€ per year and will be paid from laboratory founds.**

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 ☐ No If yes, please specify: **Only analogue data cannot be shared directly as it needs a specific storage conditions and even then can not be eligible to sharing due to degradation.**
Which data will be made available after the end of the project?	**We plan to make all the raw published data available for the public. Including sequencing results, programming code used to analyse the data, images, results of qPCR, statistical analyses etc.**
Where/how will the data be made available for reuse?	☒ In an Open Access repository ☐ In a restricted access repository ☐ Upon request by mail ☐ Other (specify):
When will the data be made available?	**After publication.**
Who will be able to access the data and under what conditions?	**We aim to place the data within online databases that do not have any access restrictions, like e.g. GEO for sequencing data. Thus, everyone will be able to access the data.**

What are the expected costs for data sharing? How will these costs be covered? *Although FWO has no earmarked budget at its disposal to support correct research data management, FWO allows for part of **the allocated project budget** to be used to cover the cost incurred.*	**We aim to submit our data to free repositories like Zenodo created by CERN. However, this is still under discussion and can change and part of the allocated project budget can be used for this purpose.**

8. Responsibilities	
Who will be responsible for the data documentation & metadata?	**PhD student**
Who will be responsible for data storage & back up during the project?	**PhD student and PI**
Who will be responsible for ensuring data preservation and sharing?	**PhD student and PI**
Who bears the end responsibility for updating & implementing this DMP? *Default response: The PI bears the overall responsibility for updating & implementing this DMP*	**The PI bears the overall responsibility for updating & implementing this DMP**