MY PLAN (FWO DMP)

DMP TITLE

ADMIN DETAILS

Project Name: My plan (FWO DMP) - DMP title
Principal Investigator / Researcher: Peter Witters
Institution: KU Leuven

1. GENERAL INFORMATION

Name applicant
Peter Witters

FWO Project Number & Title
18B4322N
Towards treatments of Congenital Disorders of Glycosylation

Affiliation
KULeuven

2. DATA DESCRIPTION

Will you generate/collect new data and/or make use of existing data?
We will collect new 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. If you reuse existing data, specify the source of these data. Distinguish data types (the kind of content) from data formats (the technical format).
1. Patient Level
Tissue biopsies will be taken from patient diagnosed at the UZ Gasthuisberg. From these tissues the clinical partners will grow fibroblast cell lines that will be used for the metabolomics assays.

2. Digital data (metabolomics)
For the metabolic assays described in the project we will generate the following files: ".raw" files which are the un-processed data files from the Mass Spectrometers, ".pmd" files containing the processing

information and finally the ".quan" files containing the info on the actual processed files which were used to generate the excel files (.xlsx format) containing information on the identified metabolites, the peak abundance (as area under the curve) and –if applicable- the fractional contribution. Each project receives a unique code ("MCF####") containing all of the aforementioned files. Each step (running, processing and report) of a specific project can be traced back to the responsible scientist conducting the actual steps. Estimated volume of data is 2 terabytes. Data will be processed using in-house metabolite library, the incorporation profiles of >250 metabolites -located in different metabolic pathways- are measured and processed using an in-house developed software tool Emfasys (Expert Metabolomics and Fluxomics Analysis System). These tools are being developed independently of this project, and therefore are not described further in this document.

Digital lab books are collated by the lead post-doctoral scientist (Matthew Bird), and the PhD students on the project using the program Findings app 2.

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.

1. Patient Level

Tissue biopsies will be taken from patient diagnosed at the UZ Gasthuisberg. From these tissues the clinical partners will grow fibroblast cell lines that will be used for the metabolomics assays.

2. Digital data (metabolomics)

For the metabolic assays described in the project we will generate the following files: ".raw" files which are the un-processed data files from the Mass Spectrometers, ".pmd" files containing the processing information and finally the ".quan" files containing the info on the actual processed files which were used to generate the excel files (.xlsx format) containing information on the identified metabolites, the peak abundance (as area under the curve) and –if applicable- the fractional contribution. Each project receives a unique code ("MCF####") containing all of the aforementioned files. Each step (running, processing and report) of a specific project can be traced back to the responsible scientist conducting the actual steps. Estimated volume of data is 2 terabytes. Data will be processed using in-house metabolite library, the incorporation profiles of >250 metabolites -located in different metabolic pathways- are measured and processed using an in-house developed software tool Emfasys (Expert Metabolomics and Fluxomics Analysis System). These tools are being developed independently of this project, and therefore are not described further in this document.

Digital lab books are collated by the lead post-doctoral scientist (Matthew Bird), and the PhD students on the project using the program Findings app 2.

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)

The UZL ethics review board has reviewed this project, and granted us permission to conduct the study as per dossier S60206.

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?

IP arising from this work is managed as per the framework agreement between the VIB 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, 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?

Digital lab books are collated by the lead post-doctoral scientist (Matthew Bird), and the PhD students on the project using the program Findings app 2. This application is a central repository for experiments, data and methods. All data is organised into subfolders describing either the experiment or method type, and is further organised by date. All files are fully searchable. The application stores the data on dropbox (this will not contain any identifying data). It is still explored whether this app could also function with the J-drive. Files can either be shared between users of the findings app, or exported as PDF documents. The findings app data is additionally backed up to the J-drive once per month. Similarly, all experiments, data processing steps, and methods are additionally exported as PDF documents and readme files (.txt) to the J-drive where they are available for all members of the consortium to accurately repeat all data processing if required.

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.

We will use an adjusted (for metabolomics) Dublin core Metadata initiative.

5. DATA STORAGE AND BACKUP DURING THE FWO PROJECT

Where will the data be stored?

All data will be stored on the Large (L) storage drive at KUL.

How is backup of the data provided?

The Large storage drive at KUL is maintained and backed up by KUL 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 – L drive KUL

What are the expected costs for data storage and back up during the project? How will these costs be covered?
Cost of L-drive is 900 euro/year and can hold 5Tb of data.

Data security: how will you ensure that the data are securely stored and not accessed or modified by unauthorized persons?
The data are stored on the KUL L-drive, which is protected under the KUL network. Patients and controls are recorded as pseudonyms, which is managed by the lead post-doctoral scientist, Matthew Bird. The list linking the pseudonym with patient records (name, EMDNR number, genetic mutation) 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 who are working directly with the samples on a need to know basis. Matthew Bird retains the sole ability to modify this file under alternative password protection. All samples are accordingly processed and recorded as per their pseudonym only.

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 raw data will be retained for at least 5 years from the data of publication of the work. As per above, patient information is identified in data files by pseudonyms only. For publication purposes, this information can be cross referenced to the patient record to obtain genetic diagnosis information. Informed consent is collected for new patients enrolled into this study to this effect as per study protocol and informed consent dossiers under number S60206, and for archived material, that was collected in the routine course of diagnosis, the use of these samples has previously been approved for this study under ethics applications number S58977.

Where will the data be archived (= stored for the longer term)?
L – drive at KUL

What are the expected costs for data preservation during the retention period of 5 years? How will the costs be covered?
Cost of L-drive is 900 euro/yr. The cost for data storage is provided by the Metabolomics Expertise Center (group of B Ghesquière)

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)?
Data is collected as a monocenter study. Data can accordingly be shared upon request where there is no longer any link remaining identifying the patients (pseudonymized). Eg, pseudonym and genetic diagnosis are the only identifiers to be recorded in the data, and would be communicated. This sharing will only take place in the framework of an agreement following the GDPR recommendations.

Which data will be made available after the end of the project?
A summary of our data reporting key findings will be published in open access non-predatory journals. Subsequently, the raw data from these publications will be stored and made available by uploading the raw files into database repositories such as MetaboLights. Full genetic (DNA sequencing) information can only be made available in a controlled access repository.

Where/how will the data be made available for reuse?
Analysed data summaries can be shared electronically by email. Transfer of raw data, if requested, would be by transfer to external hard drive

When will the data be made available?
A summary of our data reporting key findings will be published in open access non-predatory journals, upon which data will be available upon request.

Who will be able to access the data and under what conditions?
Any user can request access to the work for non-commercial uses. Commercial use of the data must be negotiated with the VIB as above.

What are the expected costs for data sharing? How will the costs be covered?
Costs of data sharing would be negotiated upon request. Expected costs of €200 for hard drive purchases and postage per request

8. RESPONSIBILITIES

Who will be responsible for data documentation & metadata?
Bart Ghesquière accepts responsibility for the data documentation & metadata recording.

Who will be responsible for data storage & back up during the project?
Data management, storage and back up will be performed by the students and postdocs associated with this project. Bart Ghesquière will monitor.

Who will be responsible for ensuring data preservation and reuse ?

As study leader, and technical specialist, Bart Ghesquière will be responsible for data storage. He will also monitor data sharing requests, and share requests with the other consortium members for review. Data will be automatically be made available after the publication of results to any requestor using the data for non-commercial purposes. Commercial use of the data will be negotiated through the VIB tech transfer office.

Who bears the end responsibility for updating & implementing this DMP?

Bart Ghesquière and Peter Witters accept responsibility for updating & implementing this DMP.