FWO DMP Template - Flemish Standard Data Management Plan

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	Angelica Pagliazzi ORCID: 0000-0002-9128-6264
Contributor name(s) (+ ORCID) & roles	Promotor: Prof. Dr. Maarten Naesens Co-promotors: Prof. Joris Vermeesch Prof. Bernard Thienpont
Project number[1] & title	1S93023N - Remote allograft rejection diagnosis by epigenetic changes in circulating cell-free DNA.
Funder(s) GrantID[2]	1S93023N (grant application number)
Affiliation(s)	X KU Leuven ☐ Universiteit Antwerpen ☐ Universiteit Gent ☐ Universiteit Hasselt ☐ Vrije Universiteit Brussel ☐ Other: Provide ROR[3] identifier when possible: https://ror.org/05f950310

[1] "Project number" refers to the institutional project number. This question is optional since not every institution has an internal project number different from the GrantID. 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.

[3] Research Organization Registry Community. https://ror.org/

Please provide a short project description	The monitoring of patients after kidney transplantation is mainly based on markers of kidney function, not sensitive and specific enough to detect allograft rejection. The diagnosis of allograft rejection relies on the histological analysis of kidney biopsies, which at the University Hospital Leuven are performed at pre-established timepoints (3, 12 and 24 months after transplantation) and in case of any change in the clinical status of the patient. The main objective of the project is to improve the post-transplant monitoring of kidney transplanted patients by developing a non-invasive biomarker accurate enough to allow for stratification of patients according to their individual risk of rejection, to help in the request and clinical interpretation of kidney biopsy and to guide the adaptation of the immunosuppressive strategy. This PhD project builds on a multidisciplinary collaboration between the Nephrology and Renal Transplantation Research Group (Prof. Maarten Naesens, supervisor) and the Laboratories for Cytogenetics and Genome Research (Prof. Joris Vermeesch, co-supervisor) and for Functional Epigenetics (Prof. Bernard Thienpont, co-supervisor). The project is organized in three work packages (WP). The first WP is dedicated to the development and optimization of new model to classify kidney transplanted patients with and without allograft rejection, via the analysis of the epigenomic profile of plasma cfDNA. For this discovery phase, we opted for a case-control study design, including in total 150 plasma samples retrospectively selected from our BIOBANK Renal Transplantation (S53364, clinicaltrials.gov NCT01331668). Each plasma cfDNA sample will be analysed with two technical approaches in parallel: after enzymatic conversion of cfDNA (to allow the study of cfDNA methylation), we will proceed with genome-wide low-coverage sequencing and with targeted sequencing for specific regions of interest. For the targeted methylation sequencing approach, we have created (in collaboration of the team of Prof. Thienpont) a unique set of genomic regions, including differentially methylated regions between the different cell-types involved in allograft rejection (tubular epithelial cells, kidney endothelial cells, immune cells), regions enriched for single nucleotide polymorphisms (SNPs) with high probability to carry heterologous alleles in donor and recipient, promoter regions and gene bodies of genes involved in inflammation and allograft rejection, and viral genome sequences. The selection of this regions of interest was based a publicly available methylation sequencing dataset (GSE186458) and on the data of an ongoing study in our laboratory, focused on single-cell RNA sequencing (scRNAseq) of a unique set of 18 kidney transplant biopsies. To increase the accuracy of our analysis of the genome-wide epigenomic profile of plasma cfDNA, and adapt it to the kidney transplant cohort, in the first WP we will profile 10 biopsies with clear clinical phenotypes

	(N=5 control biopsies; N=5 rejection biopsies collected in the BIOBANK Renal Transplantation S53364 of University Hospitals Leuven) using 10X Single Cell Multiome ATAC + Gene Expression, to investigate chromatin accessibility and to infer nucleosome positioning patterns of specific cell types in the kidney allograft. In the second WP, we will test the diagnostic performance of the newly developed biomarker for diagnosis of allograft rejection in an independent validation cohort of 300 prospectively collected plasma samples from kidney transplanted patients coming at University Hospital Leuven for their post-transplant follow-up (plasma samples stored in the BIOBANK Renal Transplantation). Finally, the third work package will be dedicated to the analysis of longitudinal plasma samples from (n=30) kidney transplanted patients with diagnosis of rejection, in order to assess the value of our biomarker in predicting the onset of graft rejection.

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[4].

				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**
Clinical data	Demographic, clinical, histological data of kidney transplanted patients included in the study	Reuse existing data	Digital	Numerical and textual	SAS format .csv .xls/.xlsx	< 100 MB	
Biological samples (plasma samples, tissue biopsies)	Plasma samples and tissue biopsies that will be included in the study (see WP1 and WP2)	Reuse existing data (retrospective samples) and generate new data (prospectively collected samples)	Physical				Biological samples routinely stored in the BIOBANK Renal Transplantation S53364 (according to all guidelines and regulations of the UZ Leuven Biobank)
10X Single Cell gene expression data	10X single cell RNA sequencing data on post-transplant kidney biopsies	Reuse existing data	Digital	Experimental	Sequencing data: .fastq.gz Reference genome (downloaded from 10X website): .fasta	< 1 TB	

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

					10x Cell Ranger output files: .bam .mtx .tsv .csv .htlm Analysis with Seurat package: .R .Rdata .rds .csv .xls/.xlsx .jpeg		
10X Single Cell Multiome (ATAC + gene expression) data	10X single cell ATAC + gene expression on post-transplant biopsies	Generate new data	Digital	Experimental	Sequencing data: .fastq.gz Reference genome (downloaded from 10X website): .fasta 10x Cell Ranger output files: .bam .mtx .tsv .csv .htlm Analysis with Seurat package:	< 1 TB	

					.R .Rdata .rds .csv .xls/.xlsx .jpeg		
cfDNA extracted from plasma samples	cfDNA extracted from plasma samples stored in the BIOBANK Renal Transplantation	Generate new data	Physical				Around 500 (in 1.5 ml Eppendorf tubes and subsequently in 0.2 ml Thermo Scientific tubes) labelled tubes stored in BIOBANK Renal Transplantation S53364 (according to guidelines of the UZ Leuven Biobank)
cfDNA sequencing data	cfDNA methylation sequencing data (genome-wide and targeted) and data analysis files	Generate new data	Digital	Experimental	Sequencing data: .fastq.gz Reference genome: .fasta Aligned reads: .bam Methylation calling files: .bedgraph .txt .csv .xls/.xlsx	< 5 TB	

GUIDANCE: *DATA CAN BE DIGITAL OR PHYSICAL (FOR EXAMPLE BIOBANK, BIOLOGICAL SAMPLES, …). DATA TYPE: DATA ARE OFTEN GROUPED BY TYPE (OBSERVATIONAL, EXPERIMENTAL ETC.), FORMAT AND/OR COLLECTION/GENERATION METHOD.* *EXAMPLES OF DATA TYPES: OBSERVATIONAL (E.G. SURVEY RESULTS, SENSOR READINGS, SENSORY OBSERVATIONS); EXPERIMENTAL (E.G. MICROSCOPY, SPECTROSCOPY, CHROMATOGRAMS, GENE SEQUENCES); COMPILED/AGGREGATED DATA[5] (E.G. TEXT & DATA MINING, DERIVED VARIABLES, 3D MODELLING); SIMULATION DATA (E.G. CLIMATE MODELS); SOFTWARE, ETC.* *EXAMPLES OF DATA FORMATS: TABULAR DATA (.POR,. SPSS, STRUCTURED TEXT OR MARK-UP FILE XML, .TAB, .CSV), TEXTUAL DATA (.RTF, .XML, .TXT), GEOSPATIAL DATA (.DWG,. GML, ..), IMAGE DATA, AUDIO DATA, VIDEO DATA, DOCUMENTATION & COMPUTATIONAL SCRIPT.* *DIGITAL DATA VOLUME: PLEASE ESTIMATE THE UPPER LIMIT OF THE VOLUME OF THE DATA PER DATASET OR DATA TYPE.* *PHYSICAL VOLUME: PLEASE ESTIMATE THE PHYSICAL VOLUME OF THE RESEARCH MATERIALS (FOR EXAMPLE THE NUMBER OF RELEVANT BIOLOGICAL SAMPLES THAT NEED TO BE STORED AND PRESERVED DURING THE PROJECT AND/OR AFTER).*	
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.	Clinical data: personal data include demographic data (e.g. age, sex), histopathological data (e.g. histopathological diagnosis, disease activity and chronicity) and clinical data (e.g. kidney function markers, therapeutic regimen, donor-specific antibody testing results, etc.) will be retrieved from the included patients' files (patients transplanted at University Hospital Leuven, S53364). All patient data are pseudonymized. Biological samples: the biological samples used in this project will be retrieved from the BIOBANK Renal Transplantation (S53364) of University Hospitals Leuven. All biological samples (e.g. plasma, tissue) are stored in labelled tubes in -20°C or -80°C freezers. Electronic laboratory databases in .xls format are used to keep track of the of these samples and their link to the original study sample ID. All biological samples are stored according to the guidelines of the UZ Leuven Biobank. 10X Single Cell gene expression data: sequencing fastq files and Cell Ranger output files from single cell RNA sequencing on post-transplant biopsies are stored on KU Leuven Large Volume Storage (L: drive).

[5] These data are generated by combining multiple existing datasets.

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, please describe these issues further and refer to specific datasets or data types when appropriate.	☒ Yes, human subject data ☐ Yes, animal data ☐ Yes, dual use ☐ No If yes, please describe: The use of clinical data and samples included in this study is approved by the Ethical Review Committee of the University Hospitals Leuven: - BIOBANK Renal Transplantation S53364 - Multi-omics profiling of kidney transplant injury phenotypes S64904
Will you process personal data[6]? If so, briefly describe the kind of personal data you will use. Please refer to specific datasets or data types when appropriate. If available, add the reference to your file in your host institution's privacy register.	☒ Yes ☐ No If yes: - Short description of the kind of personal data that will be used: personal data include demographic data (e.g. age, sex), histopathological data (e.g. histopathological diagnosis, disease activity and chronicity) and clinical data (e.g. kidney function markers, therapeutic regimen, donor-specific antibody testing results, etc.) will be retrieved from the patient file of patients transplanted at University Hospital Leuven (S53364). In addition, plasma samples from the BIOBANK Renal Transplantation will be used for cfDNA extraction and sequencing (S53364 and S64904). All data will be pseudonymized. All patients included have signed the ICF of S53364. - Privacy Registry Reference: S53364

[6] 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 envision possible valorization by tech transfer to companies working in the field of liquid biopsy and/or biomarker development in kidney transplantation: we foresee to patent our pipeline and to eventually license the patent to interested companies.
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).	Wet lab protocols are explained in detail and recorded in Word files and PDF files, stored in appropriately labelled folders on KU Leuven OneDrive. Finalized bioIT scripts for data analysis will be upload on GitHub platform, accompanied by a README.txt file. Sequencing data will be collected per sample on KU Leuven Large Volume Storage (L: Drive) and/or VSC Flemish Super Computer. A metadata file will be provided with the clear description of what the raw data files represent and how they were generated; the metadata file will be kept together with the sequencing data. Clinical data are stored in an Excel file, provided of a README sheet.
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: Sequencing data will be store on KU Leuven IT infrastructure, accompanied by a .csv, Excel and/or Word file, containing the necessary information to find and re-use specific files (sample key, date of processing, technical parameters). Sequencing data require specific metadata when submitted to public repositories (e.g. Gene Expression Omnibus, etc.). Data documentation will be tailored to their ultimate deposition in public repositories. When depositing data in a repository, the final dataset will be accompanied by detailed information regarding technical and analytical methods used to generate and analyze the data, to allow for independent reproduction.

4. Data Storage & Back-up during the Research Project

Where will the data be stored?	The data will be stored on password protected KU Leuven IT infrastructure (KU Leuven Large Volume Storage, KU Leuven One Drive, VSC Flemish Super Computer), in accordance with the KU Leuven SOPs, the principles of GDPR 2016/679, and the Belgian privacy law. All biological samples are stored in the BIOBANK Renal Transplantation (S53364) of University Hospitals Leuven, according to the guidelines of the UZ Leuven Biobank.
How will the data be backed up? *WHAT STORAGE AND BACKUP PROCEDURES WILL BE IN PLACE TO PREVENT DATA LOSS? DESCRIBE THE LOCATIONS, STORAGE MEDIA AND PROCEDURES THAT WILL BE USED FOR STORING AND BACKING UP DIGITAL AND NON-DIGITAL DATA DURING RESEARCH.*[7] *REFER TO INSTITUTION-SPECIFIC POLICIES REGARDING BACKUP PROCEDURES WHEN APPROPRIATE.*	Automatic daily back-up procedures and version tracking is guaranteed by KU Leuven IT infrastructure.
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 yes, please specify concisely: There is enough storage space for this project combining: KU Leuven OneDrive, KU Leuven Large Storage Volume and VSC staging. If no, please specify:

[7] Source: Ghent University Generic DMP Evaluation Rubric: https://osf.io/2z5g3/

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.* [7]	Data are stored on KU Leuven IT infrastructure (KU Leuven Large Volume Storage, KU Leuven One Drive, VSC Flemish Super Computer), requiring for the access a Multifactor Authentication.
What are the expected costs for data storage and backup during the research project? How will these costs be covered?	KU Leuven Large Volume Storage Price: € 104,42 / TB / year VSC Staging storage: € 20 / TB / year. The costs for data storage for this project are covered by our own funding and the SBO project Multiomic Integration of cell-free DNA profiles to Advance Disease Outcome (MICADO) project (S003422N).

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...).	Leftover biological samples will be stored in the BIOBANK Renal Transplantation (S53364 of UZ Leuven; clinicaltrials.gov NCT01331668), in accordance with the guidelines of Biobank UZ Leuven. Relevant generated pseudonymized data, and corresponding metadata, will be stored for a minimum of 10 years after the end of the project for reproducibility, verification and potential reuse. In particular, raw sequencing data and finale results will be kept for long-term storage. Intermediate files which are easily and cost-effectively reproducible may not be kept for long-term storage.

Where will these data be archived (stored and curated for the long-term)?	Generated data will be stored on KU Leuven servers with back-up capacities (KU Leuven Large Volume Storage, KU Leuven One Drive, the VSC Flemish Supercomputer Centre). Developed bioIT scripts will be stored on KU Leuven ITCS serves, as well as on public repositories such as Github.
What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	KU Leuven Large Volume Storage Price: € 104,42 / TB / year VSC Archive storage: € 70 / TB / year. The costs for data storage for this project are covered by our own funding and the SBO project Multiomic Integration of cell-free DNA profiles to Advance Disease Outcome (MICADO) project (S003422N).

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, in an Open Access repository ☐ Yes, in a restricted access repository (after approval, institutional access only, …) ☐ No (closed access) ☐ Other, please specify: Pseudonymized (coded) data will not be shared, unless a proper Data Transfer Agreement (DTA) or Material Transfer Agreement (MTA) is in place. This implies that pseudonymized data will not be made public, also not after the end of the project. Publicly relevant anonymized or aggregated datasets could be made available during or after the end of the project. Scripts, algorithms and software tools will be described in manuscripts and/or on GitHub (https://github.com). Research results will be published as BioRxiv preprints and as Open Access in peer reviewed journal.
If access is restricted, please specify who will be able to access the data and under what conditions.	Pseudonymized (coded) data will not be shared, unless a proper Data Transfer Agreement (DTA) or Material Transfer Agreement (MTA) is in place.
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.	Pseudonymized (coded) data will not be shared, unless a proper Data Transfer Agreement (DTA) or Material Transfer Agreement (MTA) is in place. Fully anonymized and aggregated data can be shared in Open Access repositories, if necessary upon publication, or upon request by third parties.

Where will the data be made available? If already known, please provide a repository per dataset or data type.	Anonymized data can be made available in Open Access repositories, such as the KU Leuven RDR (research data repository), or Gene Expression Omnibus (GEO). Algorithms, scripts and software: All the relevant algorithms, scripts and software tools driving the project will be described in manuscripts and/or on GitHub (https://github.com). (Pre-print) publications will also be automatically added to our institutional repository, Lirias 2.0, based on the authors name and ORCID ID. Research results will be published as BioRxiv preprints and as Open Access in peer reviewed journal.
When will the data be made available? *This could be a specific date (dd/mm/yyyy) or an indication such as 'upon publication of research results'.*	The data will be made available upon publication (via the required link in the publication) of research results or upon request by third parties.
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.* *Example Answer: E.g. "Data from the project that can be shared will be made available under a Creative Commons Attribution license (CC-BY 4.0), so that users have to give credit to the original data creators."* [8]	Relevant fully anonymized and aggregated data will be made available under a creative commons attribution license (cc-by 4.0, cc-by-nc-4.0) and interested parties will thereby be allowed to access data directly, and they will give credit to the authors for the data used by citing the corresponding DOI. Pseudonymized (coded) data will be shared with third parties only according to proper Data Transfer Agreement (DTA) or Material Transfer Agreement (MTA).

[8] Source: Ghent University Generic DMP Evaluation Rubric: https://osf.io/2z5g3/

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 ☐ No If yes: Shared datasets will be provided of a specific PID/DOI/GEO accession number to ensure dataset retrieval for third party use and citation.
What are the expected costs for data sharing? How will these costs be covered?	Usually, data repository in Open Access repositories is free of cost. If there are any costs associated with data sharing to third parties, the costs of this data transfer will be negotiated in the DTA/MTA.

7. Responsibilities	
Who will manage data documentation and metadata during the research project?	The promotor of the study and the head of the Nephrology and Renal Transplantation Research Group will be responsible for data documentation and metadata.
Who will manage data storage and backup during the research project?	The research and technical staff will ensure data storage and back up, with support from KU Leuven ICTS. Final responsibility for data storage and back-up lies with the promotor of this project.
Who will manage data preservation and sharing?	The promotor of the project will ensure data preservation and reuse. All requests for data sharing and reuse should be directed to the promotor of the study.
Who will update and implement this DMP?	The promotor bears the end responsibility of updating & implementing this DMP.