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	**Prof. Joris Vermeesch (0000-0002-3071-1191)**
Contributor name(s) (+ ORCID) & roles	Co-supervisor(s) appointed at a co-applicant: Prof. Toomas Kivisild (0000-0002-6297-7808) Prof. Isabelle Cleynen (0000-0003-0857-7683) Prof. Yves Moreau (0000-0002-4647-6560) Prof. Bernard Thienpont (0000-0002-8772-6845) Prof. Diether Lambrechts (0000-0002-3429-302X) Prof. Dr. Maarten Naesens (0000-0002-5625-0792)
Project number [1] & title	S003422N**:** MICADO: Multiomic Integration of cell-free DNA profiles toAdvance Disease Outcome
Funder(s) GrantID [2]	FWO-SBO
Affiliation(s)	X **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	Genomic medicine moves healthcare from being reactive to disease to being predictive to disease onset or from cure to prevention. Next generation sequencing technologies have leveraged novel concepts to diagnose, treat and monitor diseases. Analysis of free-floating cell-free DNA (cfDNA) from blood or other body fluids - called “liquid biopsy” - represents an emerging tool that enables non-invasive monitoring of tissue dynamics in multiple human physiological and pathological conditions, including pregnancy, cancer and other disorders. The partners in this consortium are developing innovative approaches for (epi)genome-wide cfDNA analysis that can lead to novel applications to monitor health and disease. In this project we will increase the cfDNA knowledge base in health and disease, develop novel analysis tools for mining and integration of cfDNA genomic and epigenome data that would enable development of new liquid biopsy-derived biomarkers. Apart from disease-specific (diagnostic) biomarkers, we will also leverage the full potential of (epi)genomic data for personalized polygenic risk scores (PRS) assessments to enable patient stratification and guided patient/disease management. Further, the integration of multi-omics data will provide novel and unique approaches for patient stratification, population screening as well as accurate disease detection. We will also demonstrate the value of the separate and combined cfDNA omics approaches in different use cases: 1) we will map the sensitivity and specificity of the developed tools in cancer management; 2) we will demonstrate the stratification potential in different conditions relevant for society and/or different pharmaceutical companies and 3) we will demonstrate the value for specific biomedical applications. Our technology will allow for earlier intervention, improve patient outcomes, increase overall population health and reduce health care costs.

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
Clinical or demographical data	Demographic, clinical, histological data of cancer, transplanted patients, healthy controls, pregnant	☒ Generate new data ☒ Reuse existing data	☒ Digital ☒ Physical	☒ Numerical ☒ Textual	.csv .xls/.xlsx	☒ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	
Biological samples (blood, plasma samples, tissue biopsies, cfDNA, genomic DNA, gDNA)	Plasma samples and tissue biopsies and extracted cfDNA and gDNA (where applicable) that will be included in the respective	Reuse of existing data (retrospective samples) and new data (prospectively collected samples)	☒ Physical			☒ NA	Biological samples routinely stored in the Biobank within dedicated studies according to regulations of the UZ Leuven Biobank

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

	studies						
Sequencing data	Genomic and methylation genome-wide and targeted data on (cf)DNA, gDNA from tissue biopsies or blood (where applicable)	☒ Generate new data ☒ Reuse existing data	☒ Digital	☒ Experimental	Sequencing data: .fastq.gz Reference genomes: .fasta Aligned reads: .bam, .bai Methylation calling files: .bedgraph, .bed .txt, .csv,.xls/.xlsx	☒ > 50 TB	
Single cell data (expression ang fragmentation)	10X single cell (ATAC + gene Expression) or similar on tissue biopsies	☒ Generate new data	☒ Digital	☒ Experimental	Sequencing data: .fastq.gz Reference genomes: .fasta Aligned reads: .bam, .bai, 10x Cell Ranger output files: .bam, .mtx, .tsv .csv,.htlm, Analysis with Seurat package: .R, .Rdata, .rds .csv, .xls/.xlsx .jpeg	☐ < 1 TB	

Array data	Illumina array data on gDNA	☒ Generate new data	☒ Digital	☒ Experimental	.idat, .csv., .xls/.xlsx, flat text files	☐ < 1 TB	

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

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.	Ethical approval by the Ethical Review Committee of the University Hospitals UZ/KU Leuven already obtained before the start of the project (S61883; S62548; S63253; S62795; S62285; S63720; S64325; S65028; S65304; S53364; S65158).
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: ☐ Yes, dual use; provide approval number: ☐ No Additional information: The use of clinical data and samples already included in this study is approved by the Ethical Review Committee of the University Hospitals UZ/KU Leuven (S61883; S62548; S63253; S62795; S62285; S63720; S64325; S65028; S65304; S53364; S65158).

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: Ethical approval was obtained by the EC of UZ Leuven S61883; S62548; S63253; S62795; S62285; S63720; S64325; S65028; S65304; S53364; S65158, G-2021-3755
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: we foresee to patent our pipelines or identified biomarkers and to eventually license the patents to interested companies or create a spin-off.
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: If the generated within the project data will have the valorization potential and lead to IP creation, it will be protected and regulated accordingly to the intellectual property rights and ownership.

3. Documentation and Metadata

[4] See Glossary Flemish Standard Data Management Plan

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*.	Wet lab protocols are described in detail and recorded in Word files and PDF files, stored in appropriately labelled folders on project-specific KU Leuven OneDrive or UZ Leuven M drive. For some wet lab procedures SOPs from the UZ diagnostic unit will be followed. Where applicable, final bioinformatic scripts will be tracked in Jypiter notebooks and for reproducibility and data analysis will be upload on GitHub platform or e.g. Figshare, which will be accompanied by a README.txt file. Sequencing data will be collected and stored either on KU Leuven Large Volume Storage (L: Drive) and mainly at VSC Flemish Super Computer. A metadata file will be provided with the clear description of the raw data and how they were generated; the metadata file will be kept together with the sequencing data. Clinical data will be stored in RedCap system, or in an Excel file, provided of a README sheet at UZ Leuven M 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.*	Sequencing data will be stored on VSC, accompanied by a metadata file, containing the necessary information to find and re-use specific files (sample key, technical parameters). Sequencing data require specific metadata when submitted to access-controlled repositories (e.g., EGA). 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; bioinformatics scripts will be provided in repositories like Figshare or GitHub.

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) ☒ OneDrive (KU Leuven) ☐ Sharepoint online ☐ Sharepoint on-premis ☒ Large Volume Storage ☐ Digital Vault ☒ Other: Vlaamse Super Computer (VSC) and UZ Leuven server
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) Data is stored on KU/UZ Leuven and VSC servers with back-up capacities.
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:
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*	Data are stored on RedCap, UZ or KU Leuven IT infrastructure (KU Leuven Large Volume Storage, KU Leuven One Drive, UZ Leuven Server and VSC Flemish Super Computer), requiring for the access a Multifactor Authentication. Also, initial access is defined by the corresponding PI research group, so it will be only available to authorized personnel.

What are the expected costs for data storage and backup during the research project? How will these costs be covered?	VSC Staging storage: € 30 / TB / year. The costs for data storage for this project are foreseen and allocated within the project budget.

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): VSC archive for raw digital files and after publication sequencing data will be deposited to European Genome-phenome Archive/GEO data repositories with controlled access meaning that a third party can obtain access to the data only following approval by the KU Leuven/UZ Leuven Data Access Committee (DAC).
What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	VSC archive storage: € 30 / TB / year. The costs for data storage for this project are allocated within the project budget and within future projects.

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: 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, but deposited to deposited to European Genome-phenome Archive/GEO data repositories with controlled access meaning that a third party can obtain access to the data only following approval by the KU Leuven/UZ Leuven Data Access Committee (DAC). Anonymized aggregated datasets could be made available after the publication. Scripts, algorithms and software tools will be described in manuscripts as supplementary files and/or on GitHub (https://github.com), or Figshare repositories. Research results will be published as preprints and as Open Access in peer reviewed journals.
If access is restricted, please specify who will be able to access the data and under what conditions.	Access to human data will be granted by the data access committee to bonafide researchers affiliated with recognized research institutions upon a proper Data Transfer Agreement (DTA) is in place between UZ/KU Leuven DAC and other research institution.

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: Due to nature of the data and also potential intellectual property, data access to human data will restricted according to the specified clauses in the informed consent forms for the different studies or due to associated intellectual property rights.
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) EGA/GEO Algorithms, scripts and software: The relevant algorithms, scripts and software tools driving the project will be described in manuscripts and/or on GitHub (https://github.com) or figshare, if no novel intellectual property rights are associated. (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 or/and as Open Access in peer reviewed journal
When will the data be made available?	☒ Upon publication of research results ☐ Specific date (specify) ☒ Other (specify) Generated data associated with intellectual property rights might not be immediately available upon publication.

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)
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?	We don't expect additional costs for data sharing

7. Responsibilities	
Who will manage data documentation and metadata during the research project?	Each of the PIs involved in the program will be responsible for their specific projects for data and metadata documentation
Who will manage data storage and backup during the research project?	Each of the PIs involved in the program will be responsible for data storage and backup during the research project
Who will manage data preservation and sharing?	Each of the PIs involved in the program will be responsible for data preservation and sharing

Who will update and implement this DMP?	The coordinator of the project, or designated research personnel will be updating & implementing this DMP