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	Lore Van Santvliet https://orcid.org/0000-0002-2344-3504
Contributor name(s) (+ ORCID) & roles	Promotors Prof. Maarten De Vos and Prof. Bert Vandenberk
Project number [1] & title	Synthetic data in healthcare
Funder(s) GrantID [2]	1107725N
Affiliation(s)	x KU Leuven ☐ Universiteit Antwerpen ☐ Universiteit Gent ☐ Universiteit Hasselt ☐ Vrije Universiteit Brussel ☐ Other: ROR identifier KU Leuven: 05f950310
Please provide a short project description	This project tackles challenges associated to synthetical, medical data and its applications in healthcare. It aims to combine data-driven and physics-based methods in the context of cardiac electrophysiology and cardiac digital twins.

[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.

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 data Master@Heart	12-lead ECG, cardiac CT scans and demographics of the Master@Heart study.	☐ Generate new data ☒ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☒ Images ☐ Sound ☒ Numerical ☒ Textual ☐ Model ☐ Software ☐ Other:	ECGs: excel files (xls) CT scans: DICOM format Demographics: excel files (xls)	☐ < 1 GB ☐ < 100 GB ☒ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	
CARP studio	Cardiac modeling software	☐ Generate new data ☒ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☐ Images ☐ Sound ☐ Numerical ☐ Textual ☐ Model ☒ Software ☐ Other:	.exec file	☒ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	
Clinical data female healthy subjects	12-lead ECGs, cardiac CT scans and demographics of healthy subjects.	☐ Generate new data ☒ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☒ Images ☐ Sound ☒ Numerical ☒ Textual ☐ Model	ECGs: excel files (xls) CT scans: DICOM format Demographics: excel files (xls)	☐ < 1 GB ☐ < 100 GB ☒ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	

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

				☐ Software ☐ Other:			
Clinical data tetralogy of Fallot patients	12-lead ECGs, cardiac CT and MRI scans and demographics of ToF patients.	☐ Generate new data ☒ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☒ Images ☐ Sound ☒ Numerical ☒ Textual ☐ Model ☐ Software ☐ Other:	ECGs: excel files (xls) CT/MRI scans: DICOM format Demographics: excel files (xls)	☐ < 1 GB ☐ < 100 GB ☒ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	
Clinical data hypertrophic cardiomyopathy patients	12-lead ECGs, cardiac CT and MRI scans and demographics of HCM patients.	☐ Generate new data ☒ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☒ Images ☐ Sound ☒ Numerical ☒ Textual ☐ Model ☐ Software ☐ Other:	ECGs: excel files (xls) CT/MRI scans: DICOM format Demographics: excel files (xls)	☐ < 1 GB ☐ < 100 GB ☒ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	
Clinical data post-myocardial infarction patients	12-lead ECGs, cardiac CT and MRI scans and demographics of MI patients.	☐ Generate new data ☒ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☒ Images ☐ Sound ☒ Numerical ☒ Textual ☐ Model ☐ Software ☐ Other:	ECGs: excel files (xls) CT/MRI scans: DICOM format Demographics: excel files (xls)	☐ < 1 GB ☐ < 100 GB ☒ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	
Statistical shape models, synthetic heart shapes	SSMs of various patient groups and anatomical structures of the	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☐ Images ☐ Sound	SSM: hierarchical data format (.h5)	☐ < 1 GB ☒ < 100 GB ☐ < 1 TB	

and modeling code	heart, including the code to generate these models, and synthetic heart shapes derived from these models.			☒ Numerical ☐ Textual ☒ Model ☒ Software ☐ Other:	Code: Python and bash scripts (.py and .sh) Synthetic heart shapes: The Visualization Toolkit format (.vtk)	☐ < 5 TB ☐ > 5 TB ☐ NA	
Generative AI models and synthetic ECG database	GANs and diffusion models for synthetic ECG generation, including model weights and training code, and synthetic 12-lead ECGs	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☐ Images ☐ Sound ☒ Numerical ☐ Textual ☒ Model ☒ Software ☐ Other:	Model weights: .model files Training code: python scripts (.py) ECGs: comma separated value format (.csv)	☐ < 1 GB ☐ < 100 GB ☒ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	
Simulated ECG database	Simulated 12-lead ECGs, including code for its construction	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☐ Images ☐ Sound ☒ Numerical ☐ Textual ☐ Model ☒ Software ☐ Other:	ECGs: comma separated value format (.csv) Code: python scripts (.py)	☐ < 1 GB ☐ < 100 GB ☒ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	
Surrogate model	Surrogate model for cardiac digital twin calibration, including model weights and	☒ Generate new data ☐ Reuse existing data	☒ Digital ☐ Physical	☐ Audiovisual ☐ Images ☐ Sound ☐ Numerical ☐ Textual ☒ Model	Model weights: .model files Training code: python scripts (.py)	☒ < 1 GB ☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	

	code for its construction			☒ Software ☐ Other:			

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.	Clinical data Master@Heart: https://doi.org/10.1136/bmjsem-2021-001048 Other retrospective, clinical datasets: exist in the hospital, but no publications for these specific datasets exist. CARP studio software: https://numericor.at/rlb/wordpress/resources/
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: Ethical approval for the use of the clinical data in this study was granted by the Ethics Committee Research UZ/KU, under reference number S6133 and S68896. ☐ Yes, animal data; provide ECD reference number: ☐ Yes, dual use; provide approval number: ☐ No Additional information: This pertains to the 5 clinical datasets.
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: Personal data are the ECGs, imaging records and demographics of all 5 clinical datasets. Ethical approval for the use of the clinical data in this study was granted by the Ethics Committee Research UZ/KU, under reference number S6133 and S68896.

[4] 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:
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). *RDM guidance on documentation and metadata*.	All code that is shared, will be shared with code documentation and README files, containing a link to the paper describing the project where the code is used. Example code for usage of the SSMs, generative AI models and surrogate models will be provided to facilitate (re)use, in addition to a link to the paper describing their creation. For the synthetic meshes and synthetic and simulated ECG databases, a clear description of the data format is added to the Zenodo repository, with a link to the paper describing their creation. Clinical data: - CT and MRI scans: DICOM format, with imbedded metadata - ECG data: xls files with descriptive headers - demographics: xls files with descriptive headers

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: Clinical data: DICOM If no, please specify (where appropriate per dataset or data type) which metadata will be created: Synthetic and simulated ECG databases and synthetic meshes: Since the data are synthetically generated and do not originate from real patients, no direct patient metadata is included. When synthetic data are generated conditionally (e.g., age or diagnosis-based generation), metadata will reflect the conditioning parameters or sampled distributions used, and will be explicitly labeled as synthetic or statistically derived. This approach promotes transparency and avoids misinterpretation while enabling use in demographic-specific analysis or benchmarking.

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) ☐ Teams ☐ Sharepoint online ☐ Sharepoint on-premis ☐ Large Volume Storage ☐ ManGO ☐ Digital vault ☒ Other: Secure research server with controlled access managed by ESAT.

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 on ESAT servers is backed up daily and replicated to an off-site storage system housed in the ICTS data center.
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, the available storage on the ESAT servers and in the ICTS data center is larger than the maximum estimated volume of the datasets. ☐ 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*	The clinical datasets are stored on the ESAT servers, where access is regulated by an access control list (ACL) that grants: - read/write access to the project owner - read-only access to specific users The ACL is managed by the project owner. Client computers can access the data using SMB2 from specific IP ranges of NFSv4 from specific (IT-managed) systems.
What are the expected costs for data storage and backup during the research project? How will these costs be covered?	All expected costs will be automatically covered by ESAT, the host department.

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): Secure research server with controlled access managed by ESAT.
What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	All expected costs will be automatically covered by ESAT, the host department.

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: Open data: results and methods that do not contain personal patient data. For example: SSMs, synthetic heart shapes, simulated ECGs, synthetic ECGs created using generative AI, model weights of generative AI models. Restricted data: all clinical datasets, which contain personal patient data.
If access is restricted, please specify who will be able to access the data and under what conditions.	The data will become available for projects which have the required ethical approvals.
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: I cannot share the clinical datasets (all 5), since they contain sensitive patient information, and ethical permission needs to be obtained before other researchers can use these data.
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): Zenodo (synthetic heart shapes), GitHub and GitLab (code) ☐ Other (specify)

When will the data be made available?	☒ Upon publication of research results ☐ Specific date (specify) ☐ Other (specify)
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) Data Transfer Agreements (DTAs) will be used for sharing the clinical datasets with institutes that participate in certain research projects. Such DTA already exist for sharing the Master@Heart and female healthy subject clinical datasets with the Medical University of Graz, Austria. Possibly, in the future, additional DTAs will be made. For the results and models, an MIT license will be used. For the synthetic datasets, a CC-BY 4.0 license will be used.
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: 10.5281/zenodo.14261122 (This is a DOI for the synthetic meshes.) ☐ No
What are the expected costs for data sharing? How will these costs be covered?	There are no expected costs for data sharing.

7. Responsibilities	
Who will manage data documentation and metadata during the research project?	Lore Van Santvliet

Who will manage data storage and backup during the research project?	ESAT, KU Leuven, the host department
Who will manage data preservation and sharing?	The promotors of the project: Prof. Maarten De Vos and Prof. Bert Vandenberk
Who will update and implement this DMP?	Lore Van Santvliet