1. General Project Information	
Name Grant Holder & ORCID	**Jan D'hooge (0000-0002-2346-142X) (Principal investigator))**
Contributor name(s) (+ ORCID) & roles	**Somayeh Akbarisaghezchi (Junior researcher)** **Andrea Pulido(0000-0003-4464-4433) (Junior researcher)**
Project number [1] & title	**Machine learning to predict cardiovascular events and response to therapy based on echocardiographic-derived functional and morphological characteristics of the heart.**
Funder(s) GrantID [2]	G0C7120N
Affiliation(s)	⛌ KU Leuven ☐ Universiteit Antwerpen ☐ Universiteit Gent ☐ Universiteit Hasselt ☐ Vrije Universiteit Brussel ☐ Other: ROR identifier KU Leuven: 05f950310
Please provide a short project description	Cardiovascular disease remains a major health problem worldwide, as it is responsible for about 30% of all deaths. When diagnosing the heart, ultrasonic imaging remains the modality of choice not only due to the fact that it is non-invasive, mobile and relatively cheap but also because it can generate images in real-time and at a high rate (e.g. conventionally about 30 images/second can be generated). Although worldwide a lot of research efforts focus on estimating cardiac morphological and functional parameters in an accurate and robust manner, little attention has been given to aid the clinician in further interpreting the obtained measurements. Nevertheless, it is well recognized that these data sets are complex and hard to interpret even by experts. Within this project, we will take advantage of state-of-the-art machine learning methodologies in order to develop a tool that can support the physician in interpreting echocardiographic data and therefore guide the decision-making process. More specifically, we will extract information on local cardiac function and shape – after correcting them for confounding factors such as age or gender and determine their (individual and joint) added prognostic power. As first application domains, we will predict the risk of developing future cardiac disease on the one hand and the response to biventricular pacemaker therapy in heart failure patients on the other.

[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
Intervendor in vivo Dataset	**<u>In vivo patient dataset</u>** Sixty-three subjects (5 healthy volunteers and 58 patients) were examined with 7 different ultrasound machines	☐ Generate new data ☒ Reuse existing data	☒ Digital ☐ Physical	☒ Images ☒ Numerical ☒ Textual	*.srd *.mat *.xlsx	☐ < 100 GB ☒ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	
DOPPLER-CIP	**<u>In vivo patient dataset</u>** Non-randomized study enrolling about 1200 patients with *suspicion* of ongoing (chronic) myocardial ischemia. The dataset contains Demographic Data medical history, Imaging modalities.	☐ Generate new data ☒ Reuse existing data	☒ Digital ☐ Physical	☒ Images ☒ Numerical ☒ Textual	*.DICOM *.txt *.xlsx *.pdf	☐ < 100 GB ☒ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	

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

				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
Echonet dataset	An open-access Contains over 10k echocardiogram, or cardiac ultrasound, videos from unique patients at Stanford University Medical Center.	☐ Generate new data ☒ Reuse existing data	☒ Digital ☐ Physical	☒ Images ☒ Numerical ☒ Textual	*.DICOM *.xlsx *.pdf	☐ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☒ NA	
CAMUS Dataset	An open-access The CAMUS dataset, containing 2D apical four-chamber and two-chamber view sequences acquired from 500 patients	☐ Generate new data ☒ Reuse existing data	☒ Digital ☐ Physical	☒ Images ☒ Numerical ☒ Textual	*.nifti *.ipynb	☒ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	
Synthetic Cardiac Ultrasound Recordings Dataset	An open-access Contains 1492 synthetic cardiac ultrasound recordings with ground truth left ventricular motion.	☐ Generate new data ☒ Reuse existing data	☒ Digital ☐ Physical	☒ Images ☒ Numerical ☒ Textual	*.mat *.m	☒ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	
Realistic Synthetic dataset	An open-access Contains 105 synthetic sequences with: Healthy and ischemic motion patterns , along with the most common apical probe orientations; and vendor-specific image quality from seven different systems. Ground truth deformation is also provided.	☐ Generate new data ☒ Reuse existing data		☒ Images ☒ Numerical ☒ Textual	*.mat *.m	☒ < 100 GB ☐ < 1 TB ☐ < 5 TB ☐ > 5 TB ☐ NA	

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.	Intervendor in vivo Dataset: https://doi.org/10.1016/j.jcmg.2017.02.014 Synthetic Cardiac Ultrasound Recordings Dataset https://doi.org/10.48804/AFQH7E Realistic Synthetic dataset https://doi.org/10.1109/tuffc.2017.2786300 CAMUS Dataset https://www.creatis.insa-lyon.fr/Challenge/camus/databasesTesting.html Echonet Dataset https://aimi.stanford.edu/echonet-dynamic-cardiac-ultrasound Doppler CIP: https://cordis.europa.eu/project/id/223615/reporting
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: S66096 S64686
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: S66096 S64686

[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: DOPPLER CIP: All ultrasound data is co-owned by the DOPPLER-CIP consortium, Permission by individual consortium members is required for secondary use of the data. Intervedor Dataset: Not applicable, its only permitted to use by the members of the lab.
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: See previous question

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*.	**For the open access datasets** Can be access through their respective URL **For the In-vivo datasets** Doppler-CIP is already stored in the UZ secure server. The Data is only available upon requesting it. The data explanation is provided in xlsx files and pdf. Intervendor in vivo is also stored at UZ Leuven servers and only can be used as get the permissions
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: If no, please specify (where appropriate per dataset or data type) which metadata will be created:

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: UZ Leuven Servers
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)
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*	The patients dataset are stored in secured server at UZ Leuven where IT has the appropriate security measures to restrict access. The open access dataset are for everyone to use, so they don't need any security.

What are the expected costs for data storage and backup during the research project? How will these costs be covered?	Disk storage costs 1200 Euro/Tb/year in our hospital (backed-up). For 2000 echo studies @400Mb/study this adds up to about 0.8Tb to be stored for 3y. ***The FWO scholarship covered these costs.***

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):
What are the expected costs for data preservation during the expected retention period? How will these costs be covered?	Disk storage costs 1200 Euro/Tb/year in our hospital (backed-up). For 2000 echo studies @400Mb/study this adds up to about 0.8Tb to be stored for 3y. ***The FWO scholarship covered these costs.***

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:
If access is restricted, please specify who will be able to access the data and under what conditions.	To access the In vivo patients datasets (Doppler CIP and Intervendor in vivo) the request should be made directly to the PI of the project
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:** Patient dataset such as Doppler CIP and Intervendor In vivo Dataset are not to be share as they contain personal information
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) ☐ 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)
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?	**NA**

7. Responsibilities	
Who will manage data documentation and metadata during the research project?	**Andrea Pulido**

Who will manage data storage and backup during the research project?	**IT department UZ Gasthuisberg**
Who will manage data preservation and sharing?	**IT department UZ Gasthuisberg**
Who will update and implement this DMP?	**Andrea Pulido**