DMP template Hartstichting

Record ID	eDETECT
Project title	early Detection of disease in cardiomyopathy mutation carriers - e-DETECT
Grant number	.
Project abstract	eDETECT aims to: 1. Increase awareness among family members of patients with inherited cardiac disease, medical professionals and the general public and 2. Improve early detection of disease signs, enabling risk stratification, prediction of disease and preventing complications including SCD. A campaign to inform the general public and online tools to evaluate indications for cardiogenetic referral will be launched. The yield of novel methods to reach out to family members will be investigated and if successful, implemented into routine cascade testing. Advanced non-invasive imaging and non-invasive electrophysiological measurements will be performed in mutation carriers and asymptomatic family members to detect early signs of cardiac disease, with special attention to sex differences. As exemplar for all inherited cardiomyopathies, early detection at individual and cellular level will focus on patients and their family members with PLN or PKP2 gene founder mutations. Early detection facilitates risk stratification and timely therapeutic interventions, thereby reducing mortality and morbidity. A national infrastructure will be developed to facilitate biomarker discovery and follow-up research. Extensive genetic and phenotypic (including imaging) data will be stored and integrated into a central data warehouse to facilitate data sharing and guarantee long-term sustainability of the registries. As such, e-DETECT will help in the prevention of cardiac dysfunction and SCD among young, apparently healthy individuals. By 2030, the consortium aims to increase the uptake of cascade testing in family members by 50% and thereby detect up to 25% more patients at true risk. This increased uptake will enable regular follow up and timely treatment, thereby reducing morbidity and mortality.

1. PROJECT GROUP

1.1 Who is the contact person of the project? Please include full name, institute, e-mail address, telephone number and ORCID.	J. Peter van Tintelen (UMC Utrecht) j.p.vantintelen-2@umcutrecht.nl +31(0)88 7553810
1.2 Is there a person responsible for data management in this project?	⊗ Yes ◯ No
1.3 Please provide full name, institute, e-mail adress, telephone number and ORCID of the data manager.	Erik van Iperen (Netherlands Heart institute) e.p.vaniperen@amsterdamumc.nl +31(0)205666502
1.4 Is there a back-up data manager?	⊗ Yes ◯ No
1.5 Please provide full name, institute, e-mail adress, telephone number and ORCID of the back-up data manager.	Wanda van Ast (Netherlands Heart institute) wanda.hermans-vanast@durrercenter.nl +310205666499
1.6 Who is the contact person after the project has ended? Please include full name, institute, e-mail adress, telephone number and ORCID.	J. Peter van Tintelen (UMC Utrecht) j.p.vantintelen-2@umcutrecht.nl +31(0)88 7553810

2. DATA REUSE

2.1 Did you search for existing data that could be reused in your project? This could be either third-party data or data of your own institute.	⊗ Yes ◯ No

Catalogues and repositories:

- BBMRI
- clinicaltrials.gov
- Dash
- Dataverse
- Figshare
- re3data
- UK data archive
- YODA
- Zenodo

2.3 Will you reuse existing data in the current project? This could be either third-party data or data of your own institute.	⊗ Yes ◯ No
2.4 Did you check whether the informed consent form used for the existing data collection allows you to reuse this data? Describe how the intention of reuse is approached by the third party data, or simply copy and paste the statement from the informed consent form.	Relevant data has been identified for reuse. In WP2, participants with a PLN and/or PKP2 mutation who are included in the UNRAVEL, iPHORECAST or the NLHI ACM registry, will be included in eDETECT WP2. Participants of UNRAVEL and ACM have been informed about potential reuse of the data.

2.5 What documents or websites exist describing the access, privacy, secondary use and co-authorship rules and policies for reusing the existing data?	In part this has been described under task 4.1. A data set catalogue will be developed using MOLGENIS. Formal documents will be published on the website of the project.
2.6 Who is the contact person responsible for the existing data? Please include full name, institute, e-mail adress, telephone number and ORCID.	Folkert Asselbergs (F.W.Asselbergs@umcutrecht.nl)

3. CREATING AND PROCESSING DATA

3.1 What metadata will be produced on project-level? Will you use a metadata standard?	The data collection is registered in the DCVA catalogue. The metadata scheme is based on the BBMRI metadata scheme.
3.2 Did you search for metadata standards at data level that could be used in the project? Please explain your answer. Include a description of the metadata standard, if applicable.	Snomed and ICD10 codes will be used. Also, items formulated according to the Clinical data Acquisition Standards Harmonization (CDASH) were also used to build the electronic case report forms. Furthermore, the study will be described using MIABIS (for the BBMRI-NL catalogue) which also makes use of the NCI thesaurus standard. Moreover, the data will also be described using the Maelstrom Research Areas of Information terminology. The questionnaires used in WP1 are based on metadata standards (TMSI, OCDF and STAI).

3.3 What kind of variables will be measured at the study? Specify which tools or instruments will be used for measuring the data. In which document will the variables be described in detail?	e-DETECT is divided across different workpackages. For sake of completeness, we will describe the variables across the different work packages (WPs). WP1: Qualitative Data "Informing relatives at risk of inherited cardaic diseases: a qualitative study with health care professionals, patients and relatives." A. Instrument: Interviews patients and family members Domains: Own experience, informing family members; Actual vs. Active way of informing family members; Informing family members as an obligation or responsibility; Right of not knowing; Mean for transfering the information; Type of information; Terms and conditions for notification of hereditary advice; Providing information. B. Instrument: Focus groups with caregivers Domains: Ways to contact family members, Responsibilities to inform family members; Direct/active way; Information for family members; Conditions / barriers to applying hereditary advice. Used standards/data: Quantitative data: questionnaires are listed, all items included if not otherwise specified. Motivational to inform family members: Informing Relatives inventory (IRI) Family communication style: Adaped version of the questionnaire "Opennes to Discuss Hereditary Cancer in the Family" (OCDF) All items of the "Threatening Medical Situations Inventory" (TMSI) Trait subscale of the State Trait Anxiety Inventory (STAI) WP2: The data dictionary for this work package is available upon request and will be provided by the datamanager. Regarding areas of information that were used: demographics, presentation, family history, exercise history, medication, ECG, saECG, Holter, exercise tolerance test, physiology, MRI, echocardiogram, angiogram, tissue biopsy, arrhythmic events ICD, inappropriate ICD intervention, atrial arrhythmia, pregnancy, heart failure, htx, death, diagnostic tfc. When available and applicable, classification of the AHA or ESC are used. WP3: PICP, PINP, PIIINP and ICTP (peptides involved in collagen synthesis/turnover). miRNAs (miRNA-21, 30c and 133a). Sodium current reduction, characteristics of impulse propagation. Sodium current and action potential (AP) characteristics in iPSC-CM.

3.4 List all sources of data and include the following information (use the answer example as guidance, but feel free to include additional fields if relevant): o Data source; o Software necessary for reading it; o Version of the software; o Format of the data (e.g. .csv, .dat); o Where can the software be accessed. Example answer: Source: Tabular data with physiological and anthropometrical. Software: SPSS 22.0; R Statistics 4.2. Format: .sav; .csv. Software access: www.spss.com; www.r-project.org.	WP1: Questionaires, CASTOR for collecting and SPSS for analyzing. WP1 en 2: Tabular data, SAS software version 9.4 (SAS Institute, Cary, NC, USA) and RStudio version 1.1.414 (Boston, MA, USA) WP2: REDCap database online.
3.5 How are the datasets and raw data going to be named and stored? How will you name your folders? How will you structure your files? What naming conventions will you use? Include: o Name of the file (including date and version) o Description of the file o Where is the data stored Example answer: Description: Final dataset of processed data used for statistical analyses included in the publications Smith et al. 2019. Path and file name: G:/Department/StudyA/Analyses/ 20180629_EW01_FinalDataset.csv G:/Users/StudyA/Documents/Datasets/ StudyA_FinalDataset_01012018.csv.	Within each work package arrangements have been made how to handle file naming. Each file will have a structured name with a date or version number and will refer to an analysis for example.
More information on file naming: See digitalscholarshipleiden.nl and GARP for an example.	
3.6 How will you secure a master file? How will you handle versioning?	Within each work package arrangements have been made how to handle versioning. A version number or date will be included in the file name.
3.7 Will you create a data dictionary? Where is it going to be stored?	The datadictionary is part of the REDCap database, which is hosted by the Netherlands Heart Institute.
3.8 Will established terminologies or ontologies be used in the project? Which terminology/ontology will be used for which variable? Please explain your answer.	In the REDCap database it is documented how each variable needs to be registered.
3.9 What will be the procedure to standardize variables without standard ontologies? Will a codelist be created? How will it be accessible for other researchers? Will it relate to established terminologies/ontologies?	New metadata standards will be published at biosharing.org. We will discuss which standards might be interesting to publish. Work package leaders will be contacted for it at a later stage of the project.

4. DATA COLLECTION AND IT PROFESSIONALISM

4.1 How is existing data going to be combined with new data?	When applicable, the approach described by Rolland et al. (2013) will be used for harmonizing data. In WP2, the eCRF software REDCap will be used. This software allows multi-center data entry.

Data harmonization

Data harmonization can be useful for researchers using third party data, or for studies that will have data collected in multiple center.
Combining different datasets consists of pooling heterogeneous different data sets and transforming them into one merged and complete data set. There are many ways to conduct this procedure, such as making use of common variables (i.e. variables that are common to the different data sets such as age or sex) or by generating new variables from different items. These variables are entitled "common data elements" (Rolland et al. 2015).

The following articles describe approaches for data-harmonization:
Rolland B, Reid S, Stelling D, et al. Toward Rigorous Data Harmonization in Cancer Epidemiology Research: One Approach. American Journal of Epidemiology. 2015;182(12):1033-1038. doi:10.1093/aje/kwv133.

Fortier I, Burton PR, Robson PJ et al. Quality, quantity and harmony: the DataSHaPER approach to integrating data across bioclinical studies. Int J Epidemiol. 2010;395:1383-1393.

Fortier I, Doiron D, Little J et al. Is rigorous retrospective harmonization possible? Application of the DataSHaPER approach across 53 large studies. Int J Epidemiol. 2011;405:1314-1328.

Doiron D, Burton P, Marcon Y et al. Data harmonization and federated analysis of population-based studies: the BioSHaRE project. Emerg Themes Epidemiol. 2013;101:12.

4.2 How are data edits going to be documented?	Electronic case report forms (eCRFs) built with REDCap will be used across all centers involved for entering data. Logs for data edits will be extracted from REDCap.

4.3 Is the data going to be audited/monitored? Some possibilities are: - check completeness of records - perform in-depth checks for selected records - perform logical and consistency checks - automate checks whenever possible	⊗ Yes ◯ No

Data audits

Data audits can help improving the quality of the data. There are several guidelines for auditing data, such as the NCI Guidelines for Auditing Clinical Trials.

The Nederlandse Federatie van Universitair medisch centra (NFU) has also reported guidelines on data audits.
It should be indicated who is responsible for conducting the audits, and how audit report forms can be accesed.
These forms should be made acessible at a data repository at a further stage.

4.4 Who is responsible for conducting the audits and how can audit report forms be accessed?	REDCap and CASTOR have an audittrail and automated checks when data is collected.

4.6 Are there going to be strategies to prevent data entry mistakes? Please explain your answer.	Validation rules will be used when applicable and computed using REDCap.

4.7 How is the data going to be stored and backed-up during the data collection phase?	Pseudonymized data will be stored and backed-up at the private server of the Netherlands Heart Institute (hosted at transIP). Images will be stored at the CTMM-TraIT server using the BioMedical Imaging Archive (hosted by Vancis). The following storage space is estimated: o Size MRI scans: 135 GB o Size echo studies: 414 GB o Size ECGs: 400 MB o Ajmaline tests: 50 MB Data from WP3 will be stored at the local servers of the Academic Medical Center (AUMC) and of the University Medical Center Utrecht (UMCU). A document displaying the data architecture workflow is under development and will be available upon request.
4.8 What are the access rules of the database across the project members? (i.e. who owns read/write access, etc.)	Researchers have access to the records of their own center in the REDCap database. Researchers have access to records of another center only when this center explicitly provides permission.
4.9 Is it needed to link multiple independently-collected data sets at the participant level?	◯ Yes ⊗ No

More information on anonymization and pseudonymization:

Anonymous data refers to data where re-identification is impossible.

Pseudonymous data is a form of de-identification, in which a part of personal information remains. This concept is not formally defined in the current EU data protection legal framework. (wsgrdataadvisor.com)

There are many techniques for data pseudonymization. The Working Party on the Protection of Individuals with Regard to the Processing of Personal Data has issued an opinion document on different anonymization and pseudonymization techniques. (ec.europa.eu)

Moreover, the UK Data Archive has issued guidelines on qualitative and quantitative data anonymization.

4.13 Which (electronic) data capture software will be used for collecting the data (e.g. eCRF system or wearable device)? Does the software have any of the following specific features: o User logs (i.e. whether it registers user activity on the database, especially desirable for audits); o Data field validation (e.g. only allowing numbers or dates to be typed on a certain variable); o Using reference values (pre-defining minimum and/or maximum values allowed).	REDCap and CASTOR will be used for data collection. These software packages have user logs, data field validation and reference values.
4.14 How are queries going to be generated and managed during the data cleaning process?	R scripts are generated to clean the data and transform the dataset to a usable format.

5. PRIVACY AND INTEGRITY

5.1 Does the project need approval by a medical ethical committee, animal ethical committee, biobank committee or another ethical committee? To which committee was the approval requested? What is the current status? Please explain your answer.	Yes. Ethical approval is needed for the active approach in WP1, and clinical investigations in WP2. Ethical approval to collect buccal mucosa samples has been given. Approval by the medical ethical committee of the UMCU has been given at an earlier stage.
5.2 Is it necessary for your project to obtain informed consent of the participants?	⊗ Yes ◯ No
5.3 What is the procedure to obtain informed consent of the participants? Will you use paper or digital forms? How are the forms going to be made accessible at a later stage?	Informed consent forms will be stored at the local UMCs collecting data (in archives, for paper version), and on a secure database for digital version.
5.4 Does the informed consent state that data created in this project can be reused for new projects? Please explain your answer.	Yes, the informed consent states that the data from this project are collected for general research, meaning it can be used for other projects.
5.5 Is there a committee assigned to review privacy and integrity issues of the project? Please explain your answer.	Yes. The Research Data Management platforms of the three participating UMCs offer solutions for reviewing privacy and integrity issues and will be contacted for providing this assistance when needed.
5.6 Will you use wearable devices?	◯ Yes ⊗ No

Wearable devices

Data captured through wearable devices poses potential privacy and integrity risks related to its underlying processes. Besides, privacy concerns, the quality of the data might be threatened by attributability issues (e.g. the device being used by someone other than the participant) or technology related issues (such as data limit, lack of wireless connection, inadequate calibration). Many of these issues have been discussed in the paper eSource in Clinical Research: A Data Management Perspective on the Use of Mobile Health Technology, issued by the Society for Clinical Data Management.

5.8 How will sensitive data be handled to ensure it is stored and transferred securely? How will the identity of the participants be protected? Example answers: 1) Identifying data will remain within the institution and will never be published. 2) All data will be stored pseudonymized. 3) It is not possible to trace back patients. 4) All data will be sent through a secure connection.	All data will be stored pseudonymized. Mail will be send using ZorgMail Veilig Verzenden (version 1.8.0.0).

6. BUDGET

6.1 How will data management be costed in the project?	The budget used for WP4 Data management is applicable for all related data management issues.

Costing data management

To cost research data management in advance can substantially reduce the costs of the project. The UK Data Archive provides a data management costing tool that can be helpful for answering this item.

6.2 Estimate the costs for sustainability (long-term storage). Make an estimation of the disk space needed for long-term storage of the data (after the data cleaning process). Indicate which (meta)data will be stored in long-term and where.	There will be no additional costs for long-term storage. Metadata will be registered in the DCVA catalogue.
6.3 Did you realize the cost estimation for data management and sustainability?	_____

7. DATA SHARING

7.1 Will you make data available for reuse at the end of your project?	⊗ Yes ◯ No
7.2 Which data is going to be made available for sharing and how is it going to be accessible (totally open, under co-authorship agreements, embargo period etc)? Access rules may vary between different types of data.	The REDCap database is already available for reuse. Also other data collected in this project will be made available.
7.3 How are the datasets going to be made findable? Which catalogues and/or repositories will be used to place the (meta)data of the project?	The clinical dataset has been made findable through the DCVA catalogue.
7.4 How is your data going to be identified? List all the persistent identifiers used for data and documentation (e.g. DOI, ISBN).	This hasn't been decided yet. The DCVA catalogue doesn't provide a PID.
7.5 Are your metadata available and sufficient for other researchers to understand your data?	The metadata from REDCap is available for other researchers. The REDCap database is provided with guidance. The questionaires of WP1 are standardized and validated, and can be interpreted by anyone familiair with these standards.

7.7 Which kind of formal documents (e.g. agreements, contracts) were produced? Where is it located and how can it be accessed? Which of these formal documents need to be retained /preserved for contractual, legal or regulatory purposes? Consider: o Research proposal (subsidieaanvraag) o Research protocol o Data/material transfer agreements o Intellectual property agreements o Informed consent forms o Data dictionaries Example answer: Research protocol (K:/Department/StudyA/Protocol/StudyA_ResearchProtocol_Final_20180101.pdf). Informed consent forms: Digital version: K:/Department/StudyA/Forms/InformedConsentform_ID1001.pdf; Paper version: Department of experimental immunology, AMC.	The documents created are contract with the Heart Foundation, the research proposal, cooperation agreement, MTA, ethical approval and biobank approval. The project leader saved these documents on the AMC and UMCU research drives. The data dictionairies are available in REDCap and CASTOR.

8. INTELLECTUAL PROPERTY

8.1 Did you generate intellectual property?	◯ Yes ⊗ No
8.3 Can you publish the existing dataset (as a part of the new, combined dataset) or is it protected by intellectual property?	The dataset isn't protected by intellectual property.

INITIAL PHASE QUESTIONS COMPLETED

Initital phase questions filled in? For the initial phase DMP all mandatory questions should be filled in. Choosing "Yes" will send an email to Durrer Center. Durrer Center will then assess the DMP and provide feedback within three weeks. Please note: this message will only be send the first time you select "Yes".	⊗ Yes

MID-TERM QUESTIONS COMPLETED

Mid-term questions filled in? For the mid-term DMP all questions should be filled in. Choosing "Yes" will send an email to Durrer Center. Durrer Center will then assess the DMP and provide feedback within three weeks. Please note: this message will only be send the first time you select "Yes".	◯ Yes