Hartstichting (Dutch Heart Foundation): Hartstichting - Initial phase

1. General information

1.1. Contact person of the project

Guidance:

Include basic information of the contact person of this project (full name, institute, e-mail address, telephone numbe and ORCID).

1.2. Is there a person responsible for data management in this project? Please include the contact details.

- Yes
- No

Guidance:

It is recommended to consult someone with expertise on data management for writing the DMP. Provide his or her contact information (i.e. full name, institute, e-mail address and telephone) at the additional comment area.

1.3. Is there a back-up data manager? Include contact information.

- Yes
- No

Guidance:

If the answer is positive, provide his or her contact information (i.e. full name, institute, e-mail address and telephone) at the additional comment area.

2. Third-party data reuse

2.1. Did the researchers search for third-party data that could be reused in the current project? - F,A,I,R

- Yes
- No

Guidance:

If answered positively, include a description of the search strategy for finding third-party data sources at the additional comment area.
There are several data catalogues and repositories where datasets are accessible upon specific requirements. See below a list of data repositories:

This item aims to apply the FAIR guiding principles of data management.

Data catalog or repository	Description
Biobanking and Biomoleculares Resources research Infrastructure - BBMRI-NL	The BBMRI-NL biobank catalogue is a searchable database, containing information on all its associated biobanks. To date, there are some 200 biobanks listed. Researchers can search the collection for cohorts on specific disease types, GWAS data, types of biomaterials and so on, and send a request to the biobank coordinators to share data and/or materials. The biobank catalogue is built to facilitate links to other registries and catalogues, like those used by BioSHaRE, LifeLines, PSI, CTMM, TRAIT and others
clinicaltrials.gov	ClinicalTrials.gov is a Web-based resource that provides patients, their family members, health care professionals, researchers, and the public with easy access to information on publicly and privately supported clinical studies on a wide range of diseases and conditions.
Dash	DataONE Dash is a self-service tool for researchers to describe, upload, and share their research data via ONEShare, member repository of the DataONE network.
Dataverse	The Dataverse Network is an open source application to publish, share, reference, extract and analyze research data.
figshare	figshare allows users to upload any file format to be made visualisable in the browser so that figures, datasets, media, papers, posters, presentations and file sets can be disseminated in a way that the current scholarly publishing model does not allow.
Re3data - Registry of Research Data Repositories	Re3data is a global registry of research data repositories from a diverse range of academic disciplines. It provides information on repositories for the permanent storage and access of data sets to researchers, funding bodies, publishers and scholarly institutions. This DataCite service is referenced by multiple publishers in their editorial policies as the best tool to identify the most appropriate data repository and recommended in the European Commission's Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020.
UK Data Archive	The UK Data Archive (UKDA) is a centre of expertise in data acquisition, preservation, dissemination and promotion and is curator of the largest collection of digital data in the social sciences and humanities in the UK.
Yale university Open Data Access - YODA	The Yale University Open Data Access (YODA) Project's mission is to advocate for the responsible sharing of clinical research data, open science, and research transparency. The Project is committed to supporting research focused on improving the health of patients and informing science and public health.
Zenodo	ZENODO builds and operate a simple and innovative service that enables researchers, scientists, EU projects and institutions to share and showcase multidisciplinary research results (data and publications) that are not part of the existing institutional or subject-based repositories of the research communities

2.2 Did the researchers check whether the informed consent form used for the third-party data collection allows the researchers to reuse this data? - F,A,R

- Yes
- No

Guidance:

It is recommended to revise the informed consent forms used by potential third-party data holders in order to avoid issues relating to the privacy of the participants. Researchers can informally ask the third-party data owners whether they do have this information or review the informed consent forms by themselves.
To answer this item, describe how the intention of reuse is approached by the third party data, or simply copyand paste the statemente from the informed consent form.

3. Creating and processing data

3.1. Did the authors search for metadata standards that could be used in the project? Include a description, if applicable. - I

- No
- Yes

Guidance:

Indicate here if there are metadata standards already available for the data that is going to be collected. These data standards can be found at biosharing.org or in previous studies using similar data.

3.2. List all the variables that will be measured at the study. Specify which tools/instruments will be used for measuring the data. - I,R

Guidance:

List all variables that are supposed to be measured at the current project. Also, include the instruments used to record this data. A table may be used for displaying the answer. Alternatively, make the data dictionary of the study available upon request.

3.3. What will be the procedure to standardize metadata for variables without standard ontologies? - F,I,R

Guidance:

Often, there are no ontologies available for all the study variables. This poses as a challenge for making data as interoperable as possible., On the other hand, it gives the researchers the opportunity to produce their own ontology and eventually help future researchers that will need to work with the same variable. Publishing new ontologies gives the possibility for collaborations between different researchers that could end in high-quality publications.

4. Data collection and IT professionalism

4.1 How existing third-party data is going to be combined with new data? - I,R

Guidance:

Describe which strategy willl be used for combining data (i.e. data harmonization) will be used. This procedure 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, shuch 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).

Some authors have described their approaches for data-harmonization .See below a list of articles on biomedical 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. PMC free article

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. PMC free article 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. PMC free article 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. PMC free article

4.2 How are data edits going to be documented?

Guidance:

For ensuring the integrity of the data, it is recommended to register everytime the generated data gets edited. That relates not only to the actual dataset, but also to the documents derived during the project (including the DMP).

Generally a dataset can be edited in two ways:real-time (i.e. into the electronic data capture system) or programmed on the software that reads the data stored on the dabase (e.g. SAS).

To certify the transparancy of the data it is recommended to document the data edits and report at the data management plan.

A template for data-edits documentation (both real-time and programmed was suggested by Brand and colleagues (2011).

4.3. Is the data going to be audited/monitored?

- No
- Yes

Guidance:

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 some 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. Are there going to be strategies to prevent data entry mistakes? If so, describe.

- No
- Yes

Guidance:

Several strategies can be taken to improve the accuracy of data entry. Such as using validation rules (by including reference range values that cannot be crossed during data entry) or using two persons to enter the data.

It is highly recommended to use an electronic case report form (eCRF) software for entering data. The most popular software are RedCAP, OpenClinica and CASTOR.

4.5. How is the data going to be stored and backed-up during the data collection phase? - F,A

Guidance:

Describe briefly what will be the approach used.There are several ways to back-up data. Besides using private devices, researchers can make use of cloud systems such as the BeeHub offered by SURFsara, B2SAFE by EUDAT and figshare. Often, the local IT departments offer solutions for storing data in a safe way.

5. Privacy and integrity

5.1. Does the project needs approval by a medical ethical committee?

- No
- Yes

Guidance:

If applicable, include to which committee the approval was requested and the current status of it.

5.2. Describe the procedure that will be used to obtain informed consent of the participants.

Guidance:

Inform how informed consent is going to be requested (e.g. by paper or digital forms), and also think in advance how this forms are going to be made acessible at a later stage. Make sure the participant gets enough information about the potential reuse of the data by third-parties.

5.3. Is there a committee assigned to review privacy and integrity issues of the project?

- No
- Yes

Guidance:

Hiring a Data and Safety Monitoring Board (DSMB) can provide additional guarantee of safety for human research. A DSMB is a group of experts studying the growing stream of data from an ongoing clinical trial at regular intervals and then advise the principal investigator and the data manager on whether or not to set or modify the study. Setting a DSMB is primarily motivated by integrity and security of the research project.

The focus is primarily on ensuring the safety of patients and to monitor the inclusion and quality of study's implementations. This involves estimating additional risks that subjects run during the project. Besides risk, the DSMB also assesses other issues such as the overall size, the speed of recruitment and duration of the research.

6. Budget

6.1. Explain how data management will be costed in the project.

Guidance:

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.

Hartstichting (Dutch Heart Foundation): Hartstichting - Mid-term phase

1. Project group

2. Third-party data reuse

2.1 Were the items 2.1 and 2.2 of the initial phase of this DMP already answered?

- Yes
- No

2.2. Inform whether there are documents relevant regarding data access, privacy and co-authorship rules and reuse of third-party data. Moreover, include details on which documents are available/need to be requested.

- No
- Yes

Guidance:

If this project makes use of third-party data, consider whether the following documents have been searched:

- Data access policies
- Informed consent form (including a statement on data reuse)
- Data transfer agreements
- Co-authorship agreements
- Data dictionaries/codebooks
- Study protocol
- Intellectual Property Rights agreements

2.3. Indicate who is the contact person and/or data manager of the third-party data source. Include full name, institution, telephone number and e-mail address.

3. Creating and processing data

3.1 List all the sources of data, including the following information:

- Data source
- Software (including version)
- Format of the data (including extension)
- Whether licenses are needed for using the software
- Access link for the software

Example Answer:

Data source	**Format (Extension)**	**Software (Version)**	**License type**	**Access link**
Demographics Cross-sectional data: 2000 participants	Tabular (.csv/.sav), ~10MB	SPSS 22	License fee applied; for member of an UMC discount is available at SurfSpot	www.surfspot.nl
Syntaxes (code for statistical analysis)	Text (.sps) ~1MB	SPSS		
Imaging (cardiac MRI) ~1000 images	.dicom ~20GB	DMAT xx	License fee applied; for member of an UMC discount is available at SurfSpot	www.trait.nl
DNA/Protein sequencing	.abi ~100GB		*FILL THIS IN*	

3.2 Inform how are the datasets/raw data going to be named and stored. Use the table to organize your answer. For biomaterials, inform how to contact the biobank where the samples are stored.

Example Answer:

Data source	Description	Path and file name
Clinical	Final dataset of processed data used for statistical analyses included in the publications Smith et al. 2019	G:/Users/StudyA/Documents/Datasets/StudyA_FinalDataset_01012018.csv
Clinical/QOL	Tabular dataset including all items of the questionnaire SF-36, not scored	G:/Users/StudyA/Documents/Datasets/StudyA_SF36complete_05062017.csv
Syntaxes	Codes for analyses reported in the manuscript Smith et al. 2019	G:/Users/StudyA/Documents/Datasets/StudyA_Syntaxes.sps
Serum samples	Serum samples of participants stored in a biobank	Biosamples stored at the Durrer Center www.durrercenter.com

3.3. List all the metadata standards used in the project

4. Data collection and IT professionalism

4.1 Inform which kind of formal documents (e.g. agreements, contracts) will be/were produced. Indicate where it is located/how can it be accessed and whether they need to be retained.

Consider the following documents and include additional documents if necessary:

- Research proposal
- Research protocol
- Data/material transfer agreements
- IPR agreements
- Informed consent forms
- Data dictionaries
- Contracts

Example Answer:

Description	Path and file name	Needs to be retained?
Research protocol	G:/Users/StudyA/Documents/Documents/StudyA_ResearchProtocolFinal_01012018.pdf	YES
Data dictionary (StudyA_FinalDataset_01012018)	G:/Users/StudyA/Documents/Documents/ DataDictionary_FinalDataset_01012018.csv	
Informed consent forms	Digital version: H:/private/StudyA/forms/InformedConsentform_ID1001.pdf Paper version: Department of experimental immunology, AMC	

4.2 Describe the access rules of the database across the project members. Please read the guidance for defitions of standard roles.

Guidance:

Please use standard role definition when applicable. For this DMP we adhere to the following definitions:

Example Answer:

Role	Institute	Add	Edit	Delete	Read	Start date	End date
PI	AMC	Y	Y	Y	Y	01-01-2019	
Data monitor 1	AMC	Y	Y	N	N	01-02-2019	01-10-2019
Project leader	AMC	N	N	N	Y	01-01-2019	
Data monitor 2	AMC	Y	Y	N	N	01-10-2019	

4.3 Is it needed to link multiple independently-collected data sets at the participant level?

If yes:

- How is this data linkage performed in a privacy-sensitive manner?
- Describe the procedure to harmonize different datasets (e.g. third-party, multicenter).
- Inform which common data elements between datasets will be used for aiding data harmonization?

- No
- Yes

Guidance:

Data harmonizartion consists of pooling heterogeneous different data sets and transforming them into 1 merged data set. There are many

ways to conduct this procedure, shuch 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).

Some authors have wrote on data-harmonization procedures. A list including some of these articles is displayed below,

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. PMC free article

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. PMC free article 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. PMC free article 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. PMC free article

4.4 Inform which (electronic) data capture software will be used for collecting the data (e.g. eCRF system or wearable device). Inform whether the software has these specific features:

- User logs (i.e. whether it registers user activity on the database, especially desirable for audits);
- Data field validation (e.g. only allowing numbers or dates to be typed on a certain variable).
- Using reference values (pre-defining minimum and/or maximum values allowed)

4.5. Describe how the data will be cleaned after being captured. Will there be any logs of users accessing the database for control purposes?

5. Privacy and integrity

5.1 How will sensitive data be handled to ensure it is stored and transferred securely?

5.2. Is it needed to link multiple independently-collected data sets at the participant level? If yes, how is this data linkage performed in a privacy-sensitive manner?

Guidance:

form if the data (qualitative and quantitative) is going to be anonymized or pseudonymized. Provide a brief description of how the procedure is going to be conducted.

The General Data Protection Regulation (GDPR) wil take full effect in 2018. Among the implications of this regulation, the preservation the identity of a participant is compulsory.

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

Pseudyonymous data is a form o de-identification, in which a part of personal information remains. This concept is not formally defined in the current EU data protection legal framework.4 Pseudonymization is a form of de-identification, in which part of the personal. In legal terms, the difference between anonymous and pseudonymous data is the way it categorizes the personal data.

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 anonymisation and pseudonymisation techniques.

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

5.3 How are privacy/accuracy concerns for data captured through wearable devices going to be addressed?

Guidance:

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 Pespective on the Use of Mobile Health Technology" , issued by the Society for Clinical Data Management.

6. Budget

6.1 Estimate the costs for sustainability (long-term storage). Make an estimation of the disks pace 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).

7. Data sharing

7.1. Identify potential reusers of the data. Moreover, include a desciption of the strategy used to make this data findable by others.

Guidance:

Points to consider:

Will the data only be interesting for researchers?

From which field?

Will the data also be available for commercial reuse?

7.2 Inform 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

Guidance:

A persistent identifier helps making data not only human-readable but also machine readable, which is crucial for making it findable. There are several ways to generate a persistent identifier. The Digital Object Identifier (DOI) is one of the most used persistent identifier systems. They published a list of (DOI) registration agencies that can help researchers generating a DOI code for their data. All the documents (including the data management plan) generated during data collection (including the dataset itself) must be assigned with a persistent identifier.

7.3 How is your data going to be identified? List all the persistent identifiers.

7.4. Describe the data access policies of the project. Explain your considerations.

Example Answer:

Access to data can be "open" or "closed" with a whole spectrum between these categories. Often researchers impose specific requirements for sharing their data. Consider these requirements (i.e. if

Are the data going to be made available through open access (direct access or after an embargo period) or restricted access? Explain your considerations.

7.5 Are there Data/Material Transfer Agreements needed ? If yes, please inform how to access this document.

- Yes
- No

7.6. Provide a systematic description of how the data will be recorded in this project (i.e.metadata schemes). List all sources of data (e.g. clinical, wearable, imaging). Indicate which metadata standards are being used (both for data formats and data content).