Data Management Plan

Name of group/project	The Quantified Heart
Name of researcher(s)/student(s)	Valeria Pannunzio (lead researcher), Carolina Falcão
Description of your research	This study aims at exploring emerging blood pressure self-tracking practices among early self-tracking technology adopters. The overall research question of the study is formulated as; How do early self-tracking technology adopters collect, display, interpret and use self-tracked blood pressure data?
Funding body(ies)	The study is funded by the CardioLab, one of the Delft Design Labs (https://delftdesignlabs.org/cardiolab/); in turn, the CardioLab receives funding from the following bodies; • Delft University of Technology • Philips Electronics Nederland B.V • Nederlandse Hartstichting
Grant number	*n.a.*
Partner organisations	CardioLab, Quantified Self
Project duration	Start: ***2019-06-01*** End: ***2019-09-01***
Date DMP created	***2019-03-26***
Date last update	***2019-05-28***
Version	*1.6*

Name of researcher(s) with roles/responsibilities for data management	Valeria Pannunzio, Carolina Falcão

1. Data Collection

Observational data will be collected during the study. Particularly, two categories of data will be collected;

1. self-tracked data, including blood pressure, steps, heart rate, and sleep quality data.
2. survey data, including data collected through interviews, questionnaires, and webinars with the participants (notes, quotes, and/or audio-video material).

Type of data	**Format**	**Software**	**Estimated data size**
Survey data	*Text, audiovideo*	*Word, Excel, Tableau, Windows Media Player*	*~20 GB*
Self-tracked data	*CSV, .json*	*Excel,* *Tableau,* *Python*	*~30 GB*

The overall volume of collected data is expected to be ~50 GB. The data collection process, visualized in Figure 1., will be organized as follows;

1. Self-tracked blood pressure data will be collected through the Omron RS7 Intelli IT device. Additionally, at participants' will, steps, heart rate, and sleep quality data will be collected through a Fitbit device. Each participant will be shipped an Omron (and, at participants' will, a Fitbit) device by regular mail. By using the Omron (and optionally the Fitbit) devices during their everyday life, participants will collect self-tracked data that will be then transferred via Bluetooth from the local storage of the two devices to the cloud environments of, respectively, the OMRON connect app and the Fitbit app. Omron's and Fitbit's privacy policies respectively, can be found at the following links: https://www.omronconnect.com/eula/nl/nl_def/; https://www.fitbit.com/legal/privacy-policy. Participants will be able to access the data collected through the Omron device by downloading it in a .csv format from the Omron app. Conversely, participants will be able to access the data collected through the Fitbit device through Open Humans, (https://www.openhumans.org/), a platform created to upload, connect, and privately store personal data. Finally, participants will be offered the option to share the aforementioned self-tracked data with the researchers by authorizing the researchers to access the relevant data through a dedicated Open Humans project page. All of the self-tracked data shared with the researchers will be stored in a password-protected SurfDrive folder managed by the researchers.

2. Survey data will be collected through online questionnaires, webinars, and individual videocalls. Respectively, questionnaire data will be collected through Collector (collector.tudelft.nl/); webinar data and individual videocalls data will be collected through Crowdcast (https://www.crowdcast.io). All of the survey data will be stored in a password-protected SurfDrive folder managed by the researchers for the whole duration of the study.

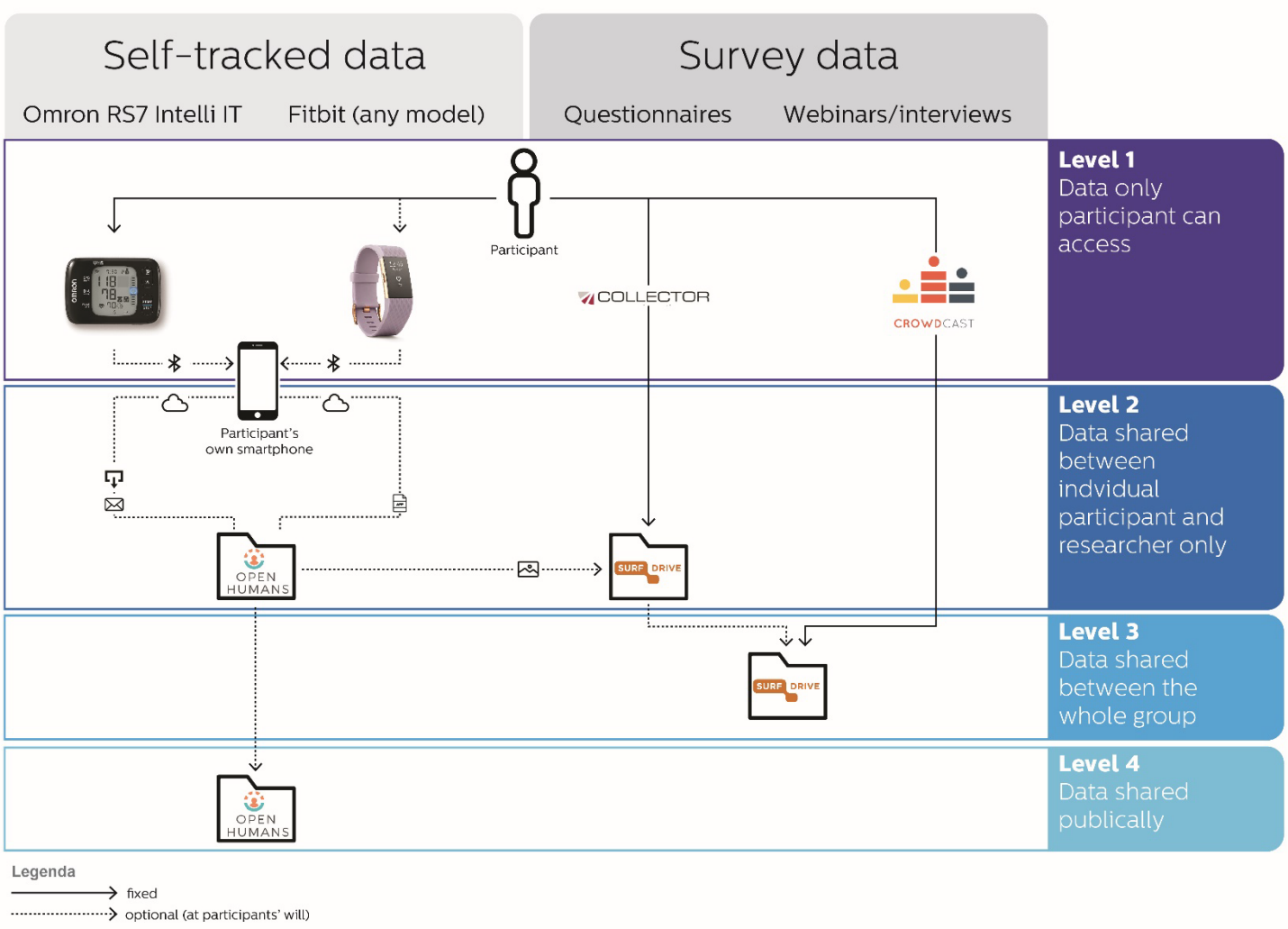

Figure 1. Overview of the data collection infrastructure

2. Data Storage and Back-up

All of the research data will be stored in a dedicated password-protected SurfDrive folder offered by the TU Delft ICT services. The data will be processed through the following softwares; Excel, Tableau, Word, and Windows Media Player. Additionally, graphical representation of the data (e.g. graphs generated through Excel or Tableau) may be imported in Adobe Illustrator to generate visual overviews of the self-tracked data. These data processing operations will be executed on a password protected TU Delft laptop.

3. Data Documentation

Data will be documented following 4TU. Centre for research data guidelines (available at https://researchdata.4tu.nl/fileadmin/editor_upload/pdf/Deposit_Guidelines/Deposit_Guidelines.pdf).

The following table summarizes the expected data labelling elements.

Creator*	Main researchers involved in producing the data.
Title*	Name or title by which the dataset is known.
Contributor	Institution where the data was created or collected. A person or organization responsible for making contributions to the dataset.
Publisher*	A holder of the data (including archives appropriate) or institution which submitted the work. Any others may be listed as contributors.
Publication year*	The year when the data was or will be made publicly available.
Date created*	Date the resource itself was put together; this could be a date range or a single date.
Description*	Concise description of the contents of the dataset. Describe the research objective, type of research, method of data collection and type of data.
Subject	Subject, keyword, or key phrase describing the resource.
Temporal coverage	Indicate the dates to which the data refer. Enter the year, or beginning and ending dates.
Spatial coverage	Describe the geographic area to which the data refer (e.g. municipality, town/city, region, country). The geographic coordinates of the area may be included, if desired.
Identifier	4TU.ResearchData automatically assigns a DOI to a dataset once the entire deposit procedure has been completed. In some cases, a dataset may be known by one or more other (persistent) identifiers.
Language*	The primary language of the resource. When no language is added, 4TU.ResearchData will automatically assign 'English'.
Link to publication	Include the web addresses or DOIs for any publication, important internal reports or other datasets that are related to your dataset.
Licence*	Terms and conditions on how the dataset may be used. Our recommended licence is CC0 as it makes your data maximally reusable. When no licence is selected, 4TU.ResearchData will automatically apply CC0 to your dataset.

4. Data Access

The self-tracked data will be formally owned by the participants who collected it. All personal and self-tracked data collected by the researchers will be destroyed at the end of the study. Conversely, survey data, in a transcribed and anonymized form, will be formally owned by the researchers and stored in 4TU. Centre for research data repository after the end of the study. The lead researcher will maintain data control credentials over the repository data; in case of staff changes or illness of the CardioLab researchers, data control credentials will be made accessible to Rudolf van Heur, CardioLab program coordinator.

5. Data Sharing and Reuse

Once anonymized, transcribed survey data is expected to be linked to one or more scientific publications and to be archived in the 4TU.ResearchData repository.
It is expected that the archived data might be reused by design and health researchers interested in attitudes of early technology adopters towards self-monitoring technologies.

6. Data Preservation and Archiving

Data will be either be discarded or retained for preservation and long-term access following criteria of data quality and expected research relevance. The discretion on retaining or discarding data will be assigned to CardioLab researchers. The retained data will be archived on 4TU.ResearchData. Archiving the data on 4TU.ResearchData is expected to happen at no cost.

Guidance

Data Management Plan

4TU.Centre for Research Data is supporting Open Science, and with that respect advocates the FAIR data principles to make research data Findable, Accessible, Interoperable and Reusable (FAIR). Creating a good data management plan contributes to these principles and stimulates researchers to publish their data in a 'FAIR' manner.

This template has been designed specifically for researchers of the three technical universities: TU Delft, University of Twente and Eindhoven University of Technology. The template may be appropriate for Principal Investigators collaborating in a research project or for research students working on a PhD or Masters project.
Your research funder may have its own Data Management Plan template. If your funder does not provide a template, you can use this template.
The Data Management Plan consists of 6 sections. Each section is accompanied with a Checklist, providing the most important questions to be answered in your Data Management Plan. Feel free to add content elements when your particular research project requires it.
More detailed information on each section will show up when you click at the 'further guidance' button.

1. Data Collection

Give a short description of the data, including amount (if known) and content. If possible give a rough estimate of the number of files.
What type of data you're generating impacts how you will manage the data and how long it needs to be preserved. There are four main types of research data:

- **Observational data**: captured in real time, typically cannot be reproduced exactly.
  Examples: sensor readings, sensory (human) observations, survey results, images
- **Experimental data**: from labs and equipment, can often be reproduced but may be expensive to do so.
  Examples: gene sequences, chromatograms, spectroscopy, microscopy
- **Simulation data**: from models, can typically be reproduced if the input data is known.
  Examples: climate models, economic models, biogeochemical models
- **Derived or compiled data**: after data mining or statistical analysis has been done, can be reproduced if analysis is documented.
  Examples: text and data mining, derived variables, compiled database, 3D models

Data types could include text, numbers, images, 3D models, software, audio files, video files, reports, surveys, etc.
Provide information on the existence (or not) of similar data and the possibilities for integration and reuse.

File formats
In planning a research project, it's important that you consider which file formats you will use to store your data. In some cases, this will be dictated by the software you're using or the conventions of your discipline, but in other cases you may have to make a choice between several options. These are likely to be some of the key factors in your decision-making:

- What software and formats you or colleagues have used in past projects
- Any discipline-specific norms (and any peer support that comes with them)

- What software is compatible with hardware you already have
- Whether you have funding for new software for the job
- How you plan to analyse, sort, or store your data

But you should also consider:

- What formats will be easiest to share with colleagues for future projects
- What formats are at risk of obsolescence, because of new versions or their dependence on particular software and/or hardware
- What formats it will be possible to open and read in the future
- What formats will be easiest to annotate with metadata so that you and others can interpret them days, months, or years in the future

Detailed guidance can be found in our documentation on recommended file formats: http://researchdata.4tu.nl/fileadmin/editor_upload/File_formats/Preferred_formats.pdf

In some cases, you may be best off using one format for data collection and analysis and converting your data to a standard format for archiving once your project is complete. After conversions, data should be checked for errors or changes that may be caused by the export process.

Version control

Because digital research data can so easily be copied, over-written or changed, researchers need to take steps to protect its authenticity. Research time is wasted and valuable data put at risk if researchers work with outdated versions of files.
Version control can prevent this. Control is particularly important if data is being used by multiple members of a research team, or if research files are shared across different locations.

A regime to synchronize different copies or versions of files will improve research efficiency and help guarantee the authenticity of the data. Good practice generally involves the keeping of a single master file, to which all changes are recorded. Version control mechanisms should be established and documented before any data is collected or generated.

Examples of version control systems for code are Git and GitHub.

Consult your local IT Service for more advice.

2. Data Storage and Back-up

It is the responsibility of the researcher to ensure that their research data is regularly backed-up and stored securely for the life of the project.
When your research project involves collaboration with several researchers, and a large amount of research data, 4TU.Centre for Research Data offers the possibility of establishing a 'data lab'.
One of the data labs provided is DataverseNL. You can store data in the DataverseNL environment and grant multiple individuals controlled access to them. Contact your local front office for more information or for setting up a dataverse for you.

There are four other options for data storage:
Networked drives: University fileserver – As these are secure and backed-up regularly, they are ideal for master copies of your research data.
Local drives: PCs and Laptops – Data can be lost because local drives can fail, or the computer may be lost or stolen. These are convenient for short-term storage and data processing but should not be relied upon for storing master copies, unless backed-up regularly.
Remote or Cloud storage – commonly used services, such as Dropbox and Google Drive, will not be appropriate for sensitive data, and their service level agreements should be studied before using them to store your research data.
As an alternative for these commercial cloud storage services, SURF has developed SURFdrive. Each university staff member may use SURFdrive (100 GB storage, access via institutional account).
External portable storage devices – External hard drives, USB drives, DVDs and CDs. These are very convenient, being cheap and portable, but not recommended for long-term storage as their longevity is uncertain and they can be easily damaged.

Consult your local IT Service for advice about secure storage options.

You may choose to only back up certain data, or to back up files you use every day more regularly than others. The basic rule of thumb is: The more important the data and the more often they change, the more regularly they need to be backed up.
If your files take up a large amount of space and backing up all of them (or backing them up sufficiently frequently) would be difficult or expensive, you may want to focus on backing up specific key information, programs, algorithms, or documentations that you would need in order to re-create the data in case of data loss.

3. Data Documentation

Describe the types of documentation that will accompany the data to help secondary users to understand and reuse it. It is recommended to createmetadata records along with your data files to describe or contextualise thedata. This includes: Title, Date of creation, Creator(s), Keywords, Description, and under what conditions it can be accessed.

Documentation may also include details on the methodology used, analytical and procedural information, and data specific information (parameters and/or variables used, column headings, codes/symbols used, etc.)
It's good research practice to add data documentation as a Readme.txt file (a plain text file) to the folder where the corresponding datasets are saved.

Wherever possible you should identify and use existing community standards.
An example of commonly used generic metadata can be found here.

For more information on common metadata standards, organized by discipline, see: http://www.dcc.ac.uk/resources/metadata-standards

File naming
Organising your files and folders effectively and efficiently can save you time and make collaboration easier by ensuring you are working on the correct version of the data. A good file name makes it easy to identify, locate and retrieve your data. There is no one recommended way to name your files and folders, but you should name your files consistently. If you work as part of a research group, you should decide on a file and folding naming system with your colleagues.
See for practical information: http://guides.lib.purdue.edu/c.php?g=353013&p=2378293

Identifiers
An identifier is a reference number or name for a data object and forms a key part of your documentation and metadata. To be useful over the long-term, identifiers need to be unique (globally unique if possible) and persistent (the identifier should not change over time).
The emerging identifier standard for publicly available datasets is the Digital Object Identifier (DOI). Although DOIs have been traditionally used for journal articles, they can now be assigned to datasets.
4TU.ResearchData will automatically assign a DOI to a dataset that you make available to allow easy citation and discoverability.

4. Data Access

During the research project you will want to keep your research data safe and secure. You will want to determine who has access to your data and what they are authorised to do with it. Data security is needed to prevent unauthorised access or disclosure and changes to or destruction of data. The principle investigators are responsible for ensuring data security. The level of security required depends upon the nature of the data – personal or sensitive data need higher levels of security.

It is possible that you will need remote access to your data, if you are working from more than one location, or not at the university. A number of individuals may require access to the data, possibly with different privileges to read, write, update or delete. This may be accomplished by keeping a copy of the data on the university shared network filestore, where it is password protected. The use of cloud storage to share data depends upon the level of security needed.

It is possible that your project may need to arrange for access to third party data that may have specific limitations in how they can be distributed (based on IP or the agreement by which your project obtained the data). When your research project has received data under confidentiality or other restrictions you will have to identify and explain these restrictions in your data management plan.

Ownership of research data must be clarified prior to, or at the beginning of a project. Future storage and reuse are directly affected by the intellectual property rights of research data.
Ownership of the data and copyrighted datasets will depend on whether the project was created as part of sponsored research; the employment status of the creator; whether third-party data has been utilised during the conduct of research, and, in case of an 'encoded work' whether substantial university resources were used in the creation of the encoded work.

The conditions under which the data may be made available by the data repository to other researchers are determined by you as the Principal Investigator depositing the data.

When depositing to 4TU.ResearchData you can choose from two general access conditions:
Open access; there are no additional restrictions on access to the data or publication of results.
Embargo period; you can request that an embargo period be imposed on your data, whereby no access to the data would be permitted until after the date you specify.

5. Data Sharing and Reuse

Your research is valuable and important, and so is the data that it is based on. By publishing your data, you make it available to the scholarly community, who can study and build upon your work. Your work will become more visible and typically be cited more frequently.
At the end of your research project, your funder may require you to share your research data, by publishing it with no access restrictions (open access). Some journal publishers also require the data supporting the research article to be published.
When disseminating your data, you need to think about who would be interested in your research findings, and the way how to reach this audience (by newsletter, community website, press release, attending seminars or conferences, etc.)

When working with personal data you have a duty to ensure that any data you gather and subsequently use is handled correctly. Ethical guidelines are issued by funding organisations and also produced by the University. In addition, laws such as the Wet bescherming persoonsgegevens, which governs the processing of personal data, must be adhered to.

You also may need to think about how you want others to reuse your data. If you want your data to be as widely used as possible, the Creative Commons Attribution Only licence (CC-BY), would be most useful. This license lets others distribute, remix, tweak, and build upon your work, even commercially, as long as they credit you for the original creation. For additional information about Creative Commons license options, visit their website.

Some data repositories have licences that depositors must grant as a condition of deposit. When depositing to 4TU.ResearchData you must agree to the terms and conditions specified in the Deposit Licence Agreement. This is a legal document which sets out your rights and responsibilities as depositor and ours as the data distributor.
Depositors can elect to apply an embargo to the research data so that public access is deferred for a specific period (typically no more than two years). Embargo may be appropriate in cases where the researcher needs to maintain the data in a managed repository environment like 4TU.ResearchData while deferring any access to the data pending further data collection, analysis, publication of results, etc.

If data is generated using specifically-developed software, it may be necessary to provide a copy of the software, noting operating requirements, with the data.

6. Data Preservation and Archiving

You need to decide what data to preserve and what data to dispose of after the end of the project.
You, as researcher, may have your own view on how long you want to retain the data. This will be influenced by the discipline, the type of data created and whether further work

or publications will be based on it. Where you are unsure what data might need to be held you should seek advice from your supervisor.

Data selected for long-term preservation will normally be submitted to a funder established data centre, disciplinary data repository or an institutional data repository. 4TU.ResearchData is a repository for technical-scientific research data that stores the data in a permanent and sustainable manner, according to the guidelines of the international Data Seal of Approval. Being a Trusted Digital Repository, 4TU.ResearchData is demonstrating to researchers that it is taking appropriate measures to ensure the long-term availability and quality of data it holds.

When depositing in 4TU.ResearchData it might be helpful to read our deposit guidelines. If you need any further help or information please contact 4TU.ResearchData (researchdata@4tu.nl)

Other data repositories can be found at re3data.

Most funders regard costs for archiving the data or preparing it for archive as allowable as long as they are justified and incurred within the life of the project. For funding in Horizon 2020, we strongly recommend to include these costs in Annex II of the application form. 4TU.ResearchData can give you support and advice.