

School of Computer Science

Data Management

in Fulfilment of

SPEC9997

Maksymilian Drzezdzon

C15311966

Degree: TU060/1

Module Coordinator: Emma Murphy

Declaration of Ownership: I declare that the attached work is entirely my own and that all sources have been acknowledged: 🗹  
**Date: 2021/03/24**

Data Management Plan &

Data Protection Impact Assessment



***Maksymilian Drzezdzon***

*C15311966*

This Project submitted in partial fulfilment of requirements of Technological University Dublin for the degree of

M.Sc. in Computer Science

May 2022

Table of Contents

[Introduction 1](#_Toc69574550)

[Data Management Plan 3](#_Toc69574551)

[Data Flow 3](#_Toc69574552)

[Handling Data Quality 4](#_Toc69574553)

[Accuracy 4](#_Toc69574554)

[Completeness 4](#_Toc69574555)

[Consistency 4](#_Toc69574556)

[Data Integrity 4](#_Toc69574557)

[Reasonability 4](#_Toc69574558)

[Timeliness 4](#_Toc69574559)

[Uniqueness 4](#_Toc69574560)

[Validity 4](#_Toc69574561)

[Data Bias 5](#_Toc69574562)

[Data Privacy and Security 5](#_Toc69574563)

[Data Protection Impact Assessment 5](#_Toc69574564)

[Processing Personal Data 6](#_Toc69574565)

[Data Stakeholders 6](#_Toc69574566)

[Data Safeguards 6](#_Toc69574567)

[Data Security 6](#_Toc69574568)

[Data Consent 6](#_Toc69574569)

[Ethical Risks 6](#_Toc69574570)

[References 7](#_Toc69574571)

# Introduction

This section outlines and summarizes the scenario for both the reader and the author of this project.

This project is based on a case study of a fictional high-tech company that aims to apply machine learning techniques to are outlined below:

1. Develop machine learning models to predict rehospitalisation and exacerbations of

health symptoms related to their heart condition.

2. Compare actual patient outcomes with their predicted scores to validate the machine

learning models that they are based on.

3. Integrate machine learning models into a remote monitoring system and present risk

scores for each patient to nurses to enable them to make decisions on clinical care

and lifestyle interventions.

remote triage by monitoring patients using sensing devices and activity trackers,

gathering blood pressure monitors and digital weighing scales.

For each patient, HomeHeart nurses in the hospital can access a dashboard where they can

view each piece of sensor data captured from Medic commercial sensing devices by day,

week or month.

Any additional triage interventions will also be recorded for the study.

The HomeHeart team will give researchers access to the Digicare system that aggregates

live sensor data for each patient. table 3

They will also share 12 months of relevant past

medical history for each patient and demographic information for each patient.

Patients who are taking part in the study are between the ages of 65 and 97 (Average age:

73 years, Standard Deviation: 5 years); 623 Males; 377 Females. For every patient

demographic information such as age, gender, heart condition (table 1), comorbidities (other

health conditions) and any reported disabilities (see table 2) has also been collected.

significant ratio diff in male to female

Blind 2 -- this will be fun

Intellectual Disability 3 -- big no no

Mild cognitive impairment (MCI) -- also bad

You are an employee at the HomeHeart Clinic

Data Protection Impact Assessment

Identify actors in scenario

Patients - data stakeholders

HomeHeart - remote care for patients with heart conditions, nurses, consultants, researchers and admin staff

Digicare - cloud based data storage, servers in ie, de and us

MLHealth - research group

Medic - sensor manufacturer - they handle data from device, stored in cloud based in france

# Data Management Plan

“A data management plan (DMP) is a formal document that describes the data you expect to acquire or generate during the course of a project, and how you will manage, maintain and protect your data. The following template is a modified and truncated version of a DMP.” (add citation)

## Data Flow

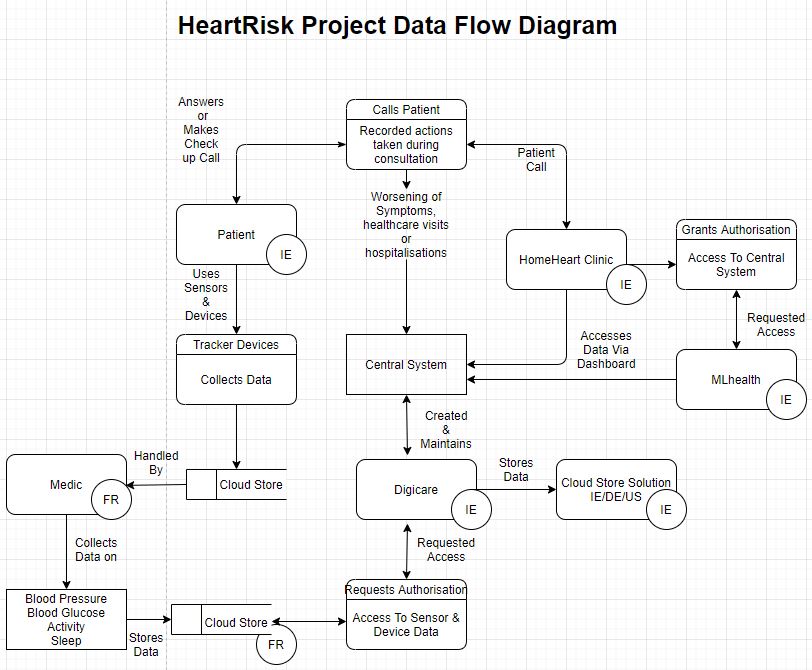


Figure : Data Flow Diagram

Digicare has data centers in Ireland, Germany and the United States. Their data center based in Ireland should be utilized for this project, keeping data within the EU and the country in which the project is being carried out.

The initial data collection is handled by Medic via their medical devices. This data will need to be access by Digicare for further processing. With such a broad range of data stakeholders, the proposed solution would be to use a private network, which HomeHeart clinic and MLhealth would need a VPN to access, it would also need to be configured by Digicare. Further details on this will be described in further sections in data security and privacy.

The data types for the above diagram would consist of text for names, addresses among anything else that is supplied, I’d need to request a list column to provide a more extensive list. Anything on medical records would also be foreign to me, a short brief on that would also be required for the purpose of the evaluation.

The sensor data gathered by Medic devices would be numeric in nature. All of this would be rendered on the Digicare systems dashboard accessed by HomeHeart Clinic and MLhealth.

## Handling Data Quality

2. What data quality issues have you identified and how will you remedy them?

[10 marks] – 250 words

When assessing data quality its important to define what high and low-quality data are first. High quality data is data that meets the expectations and needs of a data consumer, if it is not suitable for its intended purpose, its low quality. (International, DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition, 2017)

This section will asses data gathered and potential areas that could be lacking when evaluating data quality.

Data quality can be assessed using data quality dimensions outlined by DAMA International. (International, DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition, 2017)

### Accuracy

### Completeness

### Consistency

### Data Integrity

### Reasonability

### Timeliness

### Uniqueness

### Validity

## Data Bias

3. Is there any potential for data bias in this project and if so, what strategies will you

use to address this?

[15 marks] - 375 words

## Data Privacy and Security

4. What measures will you take to ensure and maintain data privacy and security for

individuals?

[10 marks] – 250 words

# Data Protection Impact Assessment

Data Protection Impact Assessment Template

Data Protection Impact Assessments can be used to identify and mitigate against any data

protection related risks arising from a new project, which may affect your organisation or the

individuals it engages with.

DPIAs are important tools for negating risk, and for demonstrating compliance with the

GDPR. The following template is a modified and truncated version of a DPIA. These

questions should be answered based on the GDPR and related 2018 Irish Data Protection

Act.

## Processing Personal Data

1. If you are processing personal data, what is the lawful basis for processing this

data?

[5 marks] – 125 words

## Data Stakeholders

2. If the project involves multiple organisations, identify the data controller(s) and

processor(s)? (Justify your answer)

[10 marks] – 250 words

## Data Safeguards

3. How will you apply safeguards to ensure the processing remains lawful e.g.

Pseudonymisation, anonymisation?

[10 marks] – 250 words

## Data Security

4. How will personal data be secured throughout its entire lifecycle?

[10 marks] – 250 words

## Data Consent

5. If relying on consent to process personal data, how will this be collected and what is

the impact if consent is withheld or withdrawn?

[5 marks] – 125 words

## Ethical Risks

6. What are the critical ethical risks for this project and how can you mitigate for them.

[10 marks] – 250 words

# References

International, D. (2017). DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. In D. International, *DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition* (pp. 449-494). Technics Publications.

International, D. (2017). DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition. In D. International, *DAMA-DMBOK: Data Management Body of Knowledge: 2nd Edition* (pp. 123-166). Technics Publications.