Data Management Assignment 2

Lecturer: Dr Emma Murphy

Assignment Context and Aim

"Technological design decisions should not dictate our societal interactions and the structure of our communities, but rather should support our values and fundamental rights." (European Data Protection Supervisor, 2015)

The aim of this assessment is to apply the data management concepts, methods and ethical theories that we have studied to a practical case study in the area of digital health.

Assignment Brief

Based on the case study detailed below you need to create a data management plan and data protection impact assessment by answering the questions using the relevant templates provided below. Some questions in the data management plan and impact assessment can be answered using the details in the case study. However, some aspects of the data management plan will require you to create individual data management solutions. Therefore, your work should not be identical to that of other students. For every decision that you make ensure that you have a clear rationale and reference relevant literature as necessary.

Case Study

The *MLHealth* research group in the Department of Computer Science at the University of Dublin and the *HomeHeart Clinic* at the Eastern City Hospital have obtained funding in Ireland for the *HeartRisk* project. The aim of this project is to develop and validate machine learning models for the prediction of potential rehospitalisation and/or exacerbation (worsening) of symptoms in previously hospitalised older people over the age of 65 with a heart condition. The models are intended to aid identification of high-risk patients in order to deliver early clinical and lifestyle interventions.

The aims of the project is to use data from patient monitoring over a 2 year period combined with previous health information for 1000 patients in order to:

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.

For the two year duration of the project, the *HomeHeart Clinic* team will deliver remote triage by monitoring patients using commercial personal sensing devices such as activity trackers, blood pressure monitors and digital weighing scales. All commercial sensing devices that will be used are manufactured by the French company *Medic*.

The *HomeHeart Clinic* team manage remote patient monitoring using a central system that aggregates sensor data for each patient. This system is developed and maintained by a multinational technology company that specialises in digital health applications (*DigiCare*). 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.

The *HomeHeart Clinic* nurses will also conduct a phone check-up every week to discuss and record any exacerbations of symptoms, healthcare visits or hospitalisations. Existing triage protocols for remote care will be in place so that additional calls may be made to patients to check on readings that breach thresholds so that nurses can advise patients of any action that they need to take i.e. contact their GP, consultant or ambulance in an emergency. Any additional triage interventions will also be recorded for the study.

In order for the researchers in the *MLHealth group* at the University of Dublin to build and validate the proposed machine learning models, the *HomeHeart* team will share necessary data from the 1000 patients in their care with the research group. The *HomeHeart* team will give researchers access to the *Digicare* system that aggregates live sensor data for each patient (examples of this data in table 3). They will also share 12 months of relevant past medical history for each patient and demographic information for each patient.

Digicare is a collaborating partner on the project and has agreed to integrate the machine learning models that will present a risk score for rehospitalisation and risk of exacerbation of heart conditions into their existing interface. These risk scores will be used to help the *HomeHeart* team manage the care of their patients and determine appropriate interventions. It is intended that the machine learning models will be developed during the first year of the project and will be released for refinement and validation into the *Digicare* interface for the last year of the project.

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.

Health Condition	No. of Patients
Hypertension (High Blood Pressure):	423
Heart Failure	191
Angina	92
Coronary Heart Disease	192
Other Heart Condition	102

Table 1: Type of Heart Condition

Reported Disability	No. of Patients
Partially Sighted	20
Blind	2

Mild Hearing Loss	30
Severe Hearing Loss	7
Intellectual Disability	3
Mild cognitive impairment (MCI)	10
Physical disability	33

Table 2 Number of patients that reported having a disability

Parameter	Metric	No. of Readings per day
Sleep	Hours and minutes	1
Activity	Step Count	1
Heart Rate	Beats per minute	5
Blood Pressure	Systolic number Diastolic number	2
Weight (Heart failure patients only)	KG	1

Table 3 Overview of sensor data per patient

Ethical approval for the project has been approved by the relevant healthcare bodies and participants have already been recruited. However before the study can begin, a data management plan and data protection impact assessment must be completed. You are an employee at the *HomeHeart Clinic* and you have been tasked with writing a Data Management Plan for the project. Based on this data management plan you need to complete a Data Protection Impact Assessment using the templates below.

About MLHealth
ML Health is a leading research group in the field of machine learning and healthcare with leading experts in the field of machine learning, data science, data ethics and digital health. It is based at the Department of Computer Science at the University of Dublin.

About the HomeHeart Clinic
The *HomeHeart Clinic* is located at the Eastern City Hospital and specialises in providing remote care and triage to patients with heart conditions. It has over 50 staff dedicated to the clinic made up of nurses, consultants, researchers and administrative staff.

About Digicare
Digicare is large multinational corporation (established in 2000) with offices in Ireland, Germany and the US with over 5,000 employees in the organisation. The team assigned to the *HeartRisk* project are based in Ireland. *Digicare* uses a cloud-based data storage platform with secure servers physically located in data centres in Ireland, Germany and the US.

About Medic
Medic is an SME based in France that manufactures medical sensors and devices to monitor health parameters such as blood pressure, blood glucose, activity, sleep ect. Data from devices is stored to a cloud-based data storage platform with secure servers physically located in data centres in France.

Data Management Plan Template

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.

1. With the aid of a diagram, outline the flow of data that will need to be managed and shared for this project including data types and the stakeholders or individuals that are responsible

   [15 marks]

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

   [10 marks]

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

   [15 marks]

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

   [10 marks]

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.

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

   [5 marks]

2. If the project involves multiple organisations, identify the data controller(s) and processor(s)? (Justify your answer)

   [10 marks]

3. How will you apply safeguards to ensure the processing remains lawful e.g. Pseudonymisation, anonymisation?

   [10 marks]

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

   [10 marks]

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]

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

   [10 marks]

Submission Details:

Submit your assignment to Brightspace as a word document, 1.5 spaced, size 12 font with clear heading structures and using APA referencing style. The total word count should not exceed 2,500 words (excluding references).

Deadline: Sunday 2 May 2021 23:30

Assessment Rubric
Marks for each section of the Data Management Plan and Data Protection Impact Assessment Plan are labelled for each question.

Learning Outcome Alignment

1	Demonstrate cognisance of the roles and responsibilities of data management stakeholders and be able to critically evaluate their significance in relation to relevant EU and International legislation.	X
2	Analyse the critical issues for data protection and data governance in public and private sector organisations across different domains such as health, education, transport and energy.	
3	Identify the role data quality plays, discuss deficits and limitations and explain appropriate remedies	X
4	Contribute to key ethical debates in data science and machine learning (i.e. data quality, bias in data, informed consent and privacy issues)	X
5	Design and implement data balancing and fairness strategies to overcome bias in data	X
6	Critically analyse different GDPR functions and responsibilities for organisations of various sizes	X
7	Critically analyse the impact of data privacy needs on organisations and individuals	
8	Design and implement anonymisation strategies and examine related issues	X
9	Design a data management plan for a complex multistakeholder prediction tool in the field of digital health.	X
10	Conduct a Data Protection Impact Assessment (DPIA)	X