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 supported by literature and theory.

Case Study

The *MLHealth* research group in the Department of Computer Science at the University of Dublin and the Irish-based commercial triage company *TriCARE* have obtained funding in Ireland for the *Predict* project. The *Predict* project aims to develop machine learning models to predict health outcomes for people with who have chronic health conditions such as heart disease and diabetes. This project will be led by both the *MLHealth* research group and *TriCARE* and they have recruited a multinational technology company (*DigiHealth*) that specialises in the design of systems for digital health for the duration of the *Predict* project.

TriCARE delivers remote healthcare by monitoring their patients using commercial personal sensing devices such as activity trackers, blood pressure and blood glucose monitors produced by a company called *Medic*. *TriCARE* currently manage the data using a central system that aggregates sensor data for each patient. For each patient, triage nurses can access a dashboard where they can view each piece of sensor data by day, week or month. They know that when certain readings are high the patient is more likely to suffer an exacerbation of their condition but it takes time to view each reading separately. The availability of a risk score that could predict exacerbations of health conditions sent daily to the Triage team would greatly help with the individual care of each patient and potentially prevent emergency hospitalisations.

MLHealth researchers will design machine learning models that can process the current and past data of the *TriCARE* patients to predict potential exacerbations (worsening) of health conditions. Under the direction of the *MLHealth* research group at the University of Dublin, *DigiHealth* will build and integrate the machine learning models into the existing interface used by TriCARE. In order for the *MLHealth* group to design the machine learning models and for *DigiHealth* to build and integrate them, *TriCARE* will share data from 500 patients in their care in Ireland. These patients are between the ages of 25 and 82 (Average age: 64, Standard Deviation: 20); 343 Males; 157 Females. For every patient demographic information such as age range, gender, health condition (table 1) and any reported disabilities (see table 2) will be shared. *TriCARE* will give access to their system that aggregates live sensor data for each patient as well as 12 months of past sensor data for each patient (see table 3).

The project has a total duration of 4 years and during year 3, the *MLHealth* group and *TriCARE* will run a trial involving a 200 of the 500 patients described above that will last for 12 months. Each trial participant will continue to use sensor devices supported by the *TriCARE* team to manage their health.

Additionally, they will take part in trial interview sessions where *MLHealth* researchers will gather further information using measures outlined in table 4 to track the health and wellbeing of participants during the trial. They hope to use this data to validate the prediction scores from the machine learning models developed during the project.

A member of the *MLHealth* research team will conduct interviews with participants at three timepoints; pre-trial, mid-trial, post-trial. These interviews will be conducted in the participant's home, or may be conducted by phone if this is suitable and more convenient for the participant. During the interview, the researcher will administer scales and questionnaires to measure key variables (see Table 4). A semi-structured qualitative interview will also be conducted to ascertain the health status of each participant at each timepoint.

Health Condition	**No. of Patients**
Hypertension (High Blood Pressure):	423
Heart Failure	120
Angina	133
Diabetes:	150

Table 1: Number of patients with health conditions (some patients have more than one condition)

Reported Disability	**No. of Patients**
Partially Sighted	20
Blind	5
Mild Hearing Loss	30
Severe Hearing Loss	7
Intellectual Disability or cognitive impairment	0
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 (if Heart Disease)	Systolic number Diastolic number	2
Blood Glucose (if Diabetes)	millimoles per litre	3

Table 3 Overview of sensor data per patient

Parameter	Metric	Timepoints
Sleep	Pittsburgh Sleep Quality Index (Buysse et al, 1989)	Pre-trial, mid-trial, post-trial
Activity	Rapid Assessment of Physical Activity (RAPA) (Topolski et al, 2006)	Pre-trial, mid-trial, post-trial
Self-Efficacy for Managing Health	Self-Efficacy for Managing Chronic Disease 6-Item Scale (Lorig, 2001)	Pre-trial, mid-trial, post-trial
Demographics	Age; Gender; Marital status; Education; Living alone/with others; Employment status; Primary occupation	Pre-trial

Table 4 Overview of trial data per patient

Ethical approval for the project has been approved in principle by the relevant healthcare bodies for the design of this technology but they require a data management plan and data protection impact assessment before the project can begin. You are a researcher in the *MLHealth* research group 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 (DPIA) using the templates below. You will present both DMP and DPIA internally to the project before formally submitting it to the relevant committee, so you can be honest in assessing risk and providing strategies for risk mitigation and recommending any changes to the data management process.

About MLHealth
ML Health is a 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 TriCARE
TriCARE is an SME (established in 2009) with approximately 100 employees including nurses, doctors and administrative staff. The company is located across three branches in Ireland.

About DigiHealth
DigiHealth 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 *Predict* project are based in Ireland. DigiHealth uses a cloud-based data storage platform with secure servers physically located in data centres in both Ireland and Germany.

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 etc. Data from devices is stored to a cloud-based data storage platform with secure servers physically located in data centres in France.

References

Buysse, D.J., Reynolds, C.F., Monk, T.H., Berman, S.R. & Kupfer, D.J. (1989) The Pittsburgh Sleep Quality Index: A new instrument for psychiatric practice and research. *Psychiatry Research, 28,* 193- 213. http://www.goodmedicine.org.uk/files/assessment,%20pittsburgh%20psqi.pdf

Lorig, K.R., Sobel, D.S., Ritter, P.L., Laurent, D. & Hobbs, M. (2001) Effect of a self-management program on patients with chronic disease. *Effective Clinical Practice, 4,* 256-262. https://www.unmc.edu/centric/_documents/SEMCD6ISinfo.pdf

Topolski, T.D., LoGerfo, J., Patrick, D.L., Williams, B., Walwick, J. & Patrick, M.B. (2006) The Rapid Assessment of Physical Activity (RAPA) among older adults. *Preventing Chronic Disease, 3(4),* 1-7. https://depts.washington.edu/hprc/wp-content/uploads/2021/02/RAPA-English.pdf

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, describe 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 throughout the data lifecycle during the project.

   [15 marks]

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

   [15 marks]

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

   [10 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? (Justify your answer)

   [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 data processing remains lawful at all stages of the project e.g. Pseudonymisation, anonymisation? Provide a clear rationale for your approach.

   [10 marks]

4. If relying on consent to process personal data, how will this be collected and what is the impact if consent is withheld or withdrawn? Provide a clear rationale for your approach.

[5 marks]

5. What are the critical ethical and privacy risks for this project and how can you mitigate for each risk identified?
   You should use a matrix like the one printed below to assign a rating to each risk identified based on severity and likelihood.

[20 marks]

	Negligible (1)	Minor (2)	Moderate (3)	Major (4)	Extreme (5)
Almost Certain (5)	5	10	15	20	25
Likely (4)	4	8	12	16	20
Possible (3)	3	6	9	12	15
Unlikely (2)	2	4	6	8	10
Rare/Remote (1)	1	2	3	4	5

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: Monday 9 May 2022 17:00

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