|  |  |
| --- | --- |
|  | |
| Data Management–Assessment 2  TU060 : Case Study: MLHealth/TriCARE – *Predict* Project.  Data Management Plan  Data Protection Impact Assessment | |
|  | |
| Ciaran Finnegan – Part Time – First Year 2021/2022  MSc in Computer Science (Data Science)  Student No : D21124026  15/4/2022 |  |
|  |  |

Table of Contents

1 Introduction 3

1.1 Purpose of Report 3

1.2 The Data Management Plan 3

1.3 The Data Protection Impact Assessment 3

2 Data Management Plan 4

2.1 Project Predict: Data Flow Lifecycle 4

2.2 Project Predict: Data Quality Issues 5

2.3 Project Predict: Data Bias and Remediation 5

2.4 Project Predict: Data Privacy and Security 5

3 Data Protection Impact Assessment 6

3.1 Project Predict: Basis for Lawful Data Processing 6

3.2 Project Predict: data Controllers and Processors 6

3.3 Project Predict: Safeguards 6

3.4 Project Predict: Data Collection and Consent 6

3.5 Project Predict: Ethical and Privacy Risk Matrix 7

4 Conclusions 8

5 References 9

# Introduction

## Purpose of Report

The..

## The Data Management Plan

The..

## The Data Protection Impact Assessment

The..

# Data Management Plan

## Project Predict: Data Flow Lifecycle

System data flow diagrams can often follow a Context Diagram/ Diagram 0 format**[1]**, but for our purposes we need to understand the lifecycle of data as it moves through the ***Predict* Project**.

The diagram below borrows from the circular format used by Folio3 Dynamics**[2]** to represent the key data lifecycle phases for the project. Stakeholders and individuals are marked in ***Italic Blue*** and the associated data types that are input/output by each phase are included.

Timeline

Description automatically generated

Figure 1 – Predict Project Data Lifecycle

For the purposes of this report the ML process is considered as an independent entity. It generates a triage recommendation and will ultimately act as a proxy for assessments carried out by nurses.

## Project Predict: Data Quality Issues (Wk10)

Below are a brief list of potential data quality issues and remediations, with which Project Predict should have a strategy to tackle;

*Publish a formal PDCA Model*

Be upfront and declare that a continuous commitment to data quality is a central tenant for Project Predict throughout the entire four-year lifecycle. A documented Plan-Do-Act-Check (PDCA) model will inspire confident that the sensitive medical data held on patients by Project Predict is being treated with the upmost respect. (Taylor et al., 2013) provide excellent guidelines in how to apply this in a healthcare setting, such as Project Predict[3].

*Ensure Sensor Accuracy*

DigiHealth must incorporate appropriate validation routines into the system implemented for Project Predict so that faulty Medic sensors cannot introduce erroneous data, outside of possible medical ranges. The potential challenge with such Medical Internet of Things (IoT) is one of the challenges discussed in tr 2019 paper by Krishnan and Shasidhar[4].

Similarly, sensor data must always be complete. Missing health metrics should also generate a system alert for correction.

*Realtime Update of Patient Dashboards for Nurses/ML Process*

Timeliness is a critical issue in the triage alerting process for Project Predict. Sensor data should be transmitted in real time, and not through a scheduled batch update. Patient dashboards that may need urgent attention should render as quickly as possible, with supplementary alerting to nurses/ML systems if immediate diagnosis is recommended. Senor data must also be timestamped to confirm that it is the most recent copy of data, and also allow a future trend analysis.

*Data Store Integrity*

The data model within the TriCARE datastore will store daily sensor data, Triage recommendations and outcomes, and then, at a later stage the Year Three Trial survey information. All of this information, which is from different sources, must seamlessly tie into a single patient data entity in the TriCARE systems, as implemented by DigiHealth.

*Data Audits*

The project needs to avoid an ad-hoc approach to checking data quality, and reliance on TriCARE employees conducting occasional reviews to capture data inconsistencies. A formal ‘data audit’ should take place each month by nominated TriCARE team members, which rewards the discovery of data issues and/or suggestions around data quality management.

*Data Storage*

Lastly, to avoid data synchronisation issues it is strongly encouraged that there is ‘one version of the truth. TriCARE should manage a single datastore, through the DigiHealth cloud infrastructure, and allow appropriate access to MLHealth researchers. There is research on current options to allow data replication across healthcare sites[5] but in this is likely beyond the scope of Project Predict and introduces a possibility of unnecessary data duplication and data consistency errors.

## Project Predict: Data Bias and Remediation (Wk9)

Ipso Lorem..

Ipso Lorem..

## Project Predict: Data Privacy and Security (Wk11)

Ipso Lorem..

Ipso Lorem..

# Data Protection Impact Assessment

## Project Predict: Basis for Lawful Data Processing

Ipso Lorem…

Ipso Lorem…

## Project Predict: Data Controllers and Processors

Ipso Lorem…

Ipso Lorem…

## Project Predict: Safeguards

Ipso Lorem…

Ipso Lorem…

## Project Predict: Data Collection and Consent

Ipso Lorem…

Ipso Lorem…

## Project Predict: Ethical and Privacy Risk Matrix

Ipso Lorem…

Ipso Lorem…

<Diagram>

# Conclusions

One could ipso lorem…

.

# References

[1] Polkovnikov, I. (2016). Unified Control and Data Flow Diagrams Applied to Software Engineering and other Systems. Retrieved 15 April 2022, from https://doi.org/10.48550/arXiv.1610.02374

[2] Saud, D. (2022). What Are The Three Main Goals Of Data Lifecycle Management (DLM)? - Folio3 Dynamics Blog. Retrieved 5 May 2022, from https://dynamics.folio3.com/blog/what-are-the-three-main-goals-of-data-lifecycle-management-dlm/