Plan Overview

A Data Management Plan created using DMPonline

Title: B213753

Creator: [redacted]

Affiliation: University of Edinburgh

Template: UoE Default DMP template for PGRs

Project abstract:

The data capture tool will look at the (length of stay) 'LOS_model' dataset and attempt to model hospital length of stay (LOS) in the ten different trusts within the dataset. Being able to analyse the differences, if any, in length of stay in different trusts will allow for better planning and allocation of resources, or perhaps highlight areas for improvement with future more in-depth analyses.

ID: 101550

Last modified: 30-06-2022

B213753

Administrative Information

2022-05-30

Data Collection

Data is generated via the LOS_model dataset contained within NHSRdatasets package in R. This is an artificially generated open source data set presented as a "tibble" in R and will be saved as a csv file locally.

Data is also collected using an interactive data collection tool using python and a Jupyter notebook. Data entry is manual and row-by-row, with an a csv file as an output.

The collected data will further have a data dictionary appended to it, saved as a R dataset (RDS) file as its final form.

The collected data consists of numerical quantitative data (Age, Length of stay, ID) and character qualitative data (Organisation, Death).

Future data will be collected using the Jupyter notebook form and saved as .csv output, using anonymised data from participating organisations. This data will change over time as more organisations add data with the aim to improve the regression model and drive data driven insights. The data will be collected as a retrospectively, sampling random admissions over the last year for each organisation.

Documentation & Metadata

Documentation for the NHSRdatasets package can be found at https://github.com/nhs-r-community/NHSRdatasets.

The project is documented throughout using markdown and code annotation. This will facilitate providing reusable, adaptative and reusable code.

Metadata and the data dictionary will be stored as attribute data onto the collected dataset in R, saved as a R dataset (RDS) file.

Ethics & Legal Compliance

The dataset used contained artificially generated data, with no data protection or ethical concerns.

However, consideration is given to use of the data collection tool in real-world data. Hence only non-identifiable data is collected (e.g. an arbitrary patient ID). In addition, the data collection tool will check for data user consent for their data to be processed and shared, and only those who consent will have their data saved to a dataset file.

Access to the dataset, tool and outputs will need to be restricted for real world data in compliance with Data protection regulation.

Storage and Back-Up

Data is stored within the Noteable cloud environment in an organised file structure. Access to the project environment is only via the author's personal Edinburgh university login.

The project folder will also be available as a public github repository, for the purposes of assessment and development. This will not be the case if the project is deployed due to data security concerns.

Selection and Preservation

If the project is to use real-world data, the dataset should be stored in a recognised research data repository such as those available in the University of Edinburgh (DataShare/ DataVault).

This will allow re-use and sharing of the tool and/or dataset with approved individuals or institutions.

The source code for dataset processing and data collection should be preserved for re-use and sharing, as this will contain no sensitive data.

Collected real-world data may be retained long-term depending on the requirements from the stakeholders, any re-use/research opportunity and the consent of the data-users.

Data Sharing

- No: go to 13

If the tool uses real-world data, then data can only be shared depending on the requirements from the stakeholders, any re-use/research opportunity and the consent of the data-users. Future access will be via secure cloud-based repositories so that sensitive data is not downloaded to local machines.

Responsibilities & Resources

The author will be responsible for data management.