Data Management Plan for the NHS England Accident and Emergency attendances and admissions data from the NHSRdataset

Administrative Information

1) School or Institute

- CMVM - Usher Institute

2) Name and Contact details of supervisor(s)

Mairead Birmingham

3) Project start date

2022-06-01

4) Project end date

2022-07-04

Data Collection

5) Data Collection

Overview:
The data will be collected via an online form for administrators of the hospital to allow for a more complete and data driven view of possible gaps to close and improvements to be mode. The collected data will be added to the NHSR dataset hosted in the Edingburgh DataStore.
1. Data collection
-Data creation:
Using one of the free data sets from the NHSRdatasets package that exists to support skills development in the NHS-R community, I selected the NHS England accident and emergency attendances and admissions (ae_attendances) which is sourced from NHS England Statistical Work Areas. The ae_attendances data includes monthly reported attendances, four-hour breaches and admissions for all A&E departments in England for 2016/17 through 2018/19 (Apr-Mar). No

identifiable information is visible within the data set, and only a single row per A&E department per month data is available, hence it is at low-risk of accidental release of confidential data. This is a credible source and the selected data set ae_attendances is complete with no missing value.

-**Data storage**: stored on Edinburgh DataStore for a period of 3 years. As data scientist need to ensure a copy of the source (raw) data is protected and can be retrieved if the data becomes corrupted. Hence put password for "RawData" folder

2. Data format

-Collected data are saved as cvs file. R Markdown document is saved as .Rmd. and output rendered as HTML. The data capture tool is saved as .ipynb and output rendered as PFD file.

-**Synthesis**: I used R for data exploration, data analysis and data visualisation. For the data capture tool, I used Python-programming and interactive widgets in Jupyter Notebook.

Documentation & Metadata

6) Documentation & Metadata

Documentation includes:

-Data dictionary appended to the collected data

-Metadata included attributes (e.g. author, date, number of rows, data types, brief description of the original data) and saved in the RawData folder.

-Using best practice for filling structures, original data set is saved in the RawData folder, collected data is saved as cvs file in the working Data folder, outputs are saved in the Outputs folder, R scripts and Python scripts are kept separately in their own respective folder.

Version control using Github repository.

Ethics & Legal Compliance

7) Ethics & Legal Compliance

-The author has completed the Data Protection Courses and the mandatory Data Protection Training prior to conducting to the research.

-No sensitive or personal identifiable data are visible in the data set.

-Written consent are obtained by the data capture tool before collecting and saving data.

-Data creator is authorised and identified for authenticity, following ISO15489:2016 data management standard

-To retain trust and to ensure no unintended bias from the data-driven applications, service users are informed on the reason for the data collection, how the data will be used, signpost to the open data repository where they can view the data now and in future to benchmark their performance.

Storage and Back-Up

8) Where will your data be stored and backed-up during the project?

Data will be stored in the UoE's DataStore, and data is backed up daily.

DataStore provides enterprise-class storage with guaranteed backup and resilience. Data is retained on DataStore until deletion by the data owner. The backups provide resilience in the case of accidental deletion and against incidents affecting the main DataStore storage. The data are automatically replicated to an off-site disaster recovery facility, with 10 days of file history visible online. Off-site tape backups keep 60 days of history of the filesystem. The 60 day rolling snapshots allow important data to be recovered to a prior state, by request if beyond the visible period.

Sensitive data stored on DataStore will be further protected by the use of 256 bit encryption as required by University policy

Selection and Preservation

9) Where will the data be stored long-term?

Public repositories on gitHub can be used to share the data and source code.
Data sharing at the 2 stages:
1. During: n/a as there's no other collaborator
2. After: For the data to be truly accessible, I will add an open source license to my repository and place it as a text file named 'LICENSE.txt' at the root of my repository.

10) Which data will be retained long-term?

We will retain the original and collected data, the meta data and the R markdown file for documentation and future-use.

The data will be preserve through an open-access repository (Github) with a license obtained to enable the data to be truly accessible. The cost for the license will be born by the author.

Data Sharing

11) Will the data produced from your project be made open?

- Yes: go to 12

12) How will you maximize data discoverability & access?

The data created from this project will be shared with the cohort for training and search purposes.
The data will be preserve through an open-access repository (Github) with a license obtained to enable the data to be truly accessible.

Responsibilities & Resources

14) Who will be responsible for the research data management of this project?

This is an individual project, the responsible person in contact is the author.
The project is supervised by a Senior Lecturer - Mairead Birmingham.

15) Will you require any training or resources to properly manage your research data throughout this project?

The author will undertake relevant mandatory Data Management Training in prior to the start of the project.
The original dataset was free, and the cost of obtaining a license for the collected data to be accessible and discoverable is born by the author.