Plan Overview

A Data Management Plan created using DMPonline

Title: Feedback loops: Using data insight to stimulate data sharing across NHS Accident and Emergency departments.

Creator: Euan Vincent

Affiliation: University of Edinburgh

Template: UoE Default DMP template for PGRs

Project abstract:

The four-hour A&E waiting time target is a pledge set out in the Handbook to the NHS Constitution.
(https://eu01.alma.exlibrisgroup.com/leganto/public/44UOE_INST/citation/37819402820002466?auth=SAML). The operational standard is that at least 95% of patients attending A&E should be admitted, transferred or discharged within four hours. In January 2019, only 84.4% of patients were seen within four hours, the worst figure since the target was introduced in 2004. (https://www.bbc.co.uk/news/health-47229719) A study in the Emergency Medicine Journal linked waiting more than 5 hours in emergency care before admission to hospital to a heightened risk of death from any cause within the next 30 days. (https://www.bmj.com/company/newsroom/5-hour-emergency-care-wait-before-admission-linked-to-heightened-death-risk/).

The sharing and analysis of routine data plays a vital role in performance management of NHS A&E departments, allowing staff and management to understand trends and employ strategies that will mitigate risk. As Dame Fiona Caldicott has noted, "The duty to share information can be as important as the duty to protect patient confidentiality." (https://www.nhsconfed.org/articles/why-data-sharing-matters-excellent-care).

I propose to improve sharing of routine data across NHS Accident and Emergency department by developing a simple and easy to use Data Capture Tool using python widgets. Users of this tool will receive access to insight provided through analysis of collected data. This insight will provide understanding to those submitting data of the use of routine NHS data in improving NHS performance through the benchmarking and insight (tend analysis etc.). This feedback loop will improve NHS data literacy and deepen the understanding of the importance of routine data to NHS management which in turn will improve data quality.

The tool will be tested using test data downloaded from the open source A&E attendances dataset, downloaded using the NHSRdatasets package.

ID: 101534

Start date: 05-06-2022

End date: 10-07-2022

Last modified: 03-07-2022

Feedback loops: Using data insight to stimulate data sharing across NHS Accident and Emergency departments.

Administrative Information

1) School or Institute

- CMVM - Centre for Genomic and Experimental Medicine

2) Name and Contact details of supervisor(s)

Mairead Bermingham - mberming@ed.ac.uk

3) Project start date

2022-06-05

4) Project end date

2022-07-10

Data Collection

5) Data Collection

1. Description of Data.
Data will be collected using a webform with embedded widgets (python) and synthesised, processed and analysed using R.
1.2 Types of Data
Quantitative data of the following description and (data types) will be collected:

- Period - period of data submitted (month / year) collected as single date (date)
- Org_code - NHS organisation code for an A&E department (factor /categorical)
- Attendance - Number of NHS A&E attendances within period (integer)
- Breaches - Number of A&E attendances that results in a breach of the four hour waiting time target (integer)
- Consent - Informed consent for data sharing (Boolean / True or False)

1.3 Format and scale of the data
Format
Collected data will be stored in CSV format to maximise interoperability with different software and coding languages.
Volume
Collected data will be small in volume. Over a three year period, less that 500KB of data is expected to be collected.
2. Data collection / generation
2.1 Methodologies for data collection / generation
Testing of the data capture tool (python) will utilise open source NHS datasets, available through the R Package NHSRdatasets (downloadable via CRAN). data will be subset into test and training datasets and the test dataset will be used to test the data capture tool.
The data data collection will be handled using a python script within Jupyter notebook. Downstream processing will be handled using an R script. In each case, scripts will be annotated to improve reusability and scripting techniques will avoid the use of non-standard or little used packages. Where packages are used, clear descriptions will be included as well as links to further resources.
2.2 Data quality and standards
Quality of the R datasets used to test the data capture tool will be checked for missing values and consistency of variable types. The data capture tool will ensure the capture of quality data through widgets (which will permit only correct variable types). Timeliness of data capture is an important metric of data quality, which will be monitored through monthly spot checks by the research lead and direct contact with submitting org units. These measures of quality closely follow those proposed by Laura Sebastian-Coleman (https://www.sciencedirect.com/topics/computer-science/data-quality-metric).

File naming

Documentation & Metadata

6) Documentation & Metadata

Metadata standards and data documentation
The Research Data Alliance standards will be followed for the collection of metadata. And a data dictionary will be created and maintained using the Datameta package (R) to outline each of the collected variables. Documentation for the creation of this dictionary will be included within the R scripts and the dictionary will be widely available through the project resources (shared via Github). A written report will be provided which fully outlines all steps in data creation, documentation and storage. Also available via Github.

Ethics & Legal Compliance

7) Ethics & Legal Compliance

No person identifiable data will be stored within this project.
The project proposal will be subjected to an ethical review board before the commencement of data capture.
Sharing and use of the collected data will be limited to researchers who are able to demonstrate training in data and research ethics and upon review of a proposal for use of the collected data.

Storage and Back-Up

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

Location of data storage
The file structure will be in un-spaced format to aid interoperability and will mirror normal file structure formats for research / coding practices (segregating Raw from processed data and using separate folders for scripts, outputs, tables and figures). This structure will be synced and shared using a Github repository. Access to the repository will be controlled using an access key. Access to the repository will be monitored by the project lead to ensure no access beyond the stated use within a project proposal.
Backup
A backup of the GitHub repository will be taken on a three-monthly basis by the project lead (Euan Vincent) and stored on a USB drive.

Selection and Preservation

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

Data will be stored within a Github repository and backed up regularly.

10) Which data will be retained long-term?

All collected data will be stored long-term for potential re-use for further studies. Such studies will be reviewed before provision of access, however, the collected data may provide a useful resource for future studies, especially in combination with other datasets and so it is anticipated to be shared with research communities, particularly those working within the NHS across the United Kingdom and other health institutions worldwide.

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?

Data will be made available through the university of Edinburgh discussion boards and will be used to prompt discussion among students and further analysis. The final report, including access to the raw and analysed datasets and all scripts and materials will be published via online journals and clear instruction on how to request access to the dataset will be provided. Finally, a user license will be provided within the root folder of the Github repository to further allow the use of the data and tool for anyone with repository access.

Responsibilities & Resources

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

Euan Vincent - principal researcher

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

.