

Working with data types and structures in Python and R_B208593

Administrative Information

1) School or Institute

- CMVM - Usher Institute

2) Name and Contact details of supervisor(s)

Dr Mairead Bermingham, mairead.bermingham@ed.ac.uk

3) Project start date

2022-06-06

4) Project end date

2022-07-04

Data Collection

5) Data Collection

The data will be collected from NHS England accident and emergency attendances and admissions records from large type 1 emergency departments. The raw quantitative data will be stored in CSV format in NHS databases.

The initial raw data, the processed data, scripts (RScripts, Python scripts) and graphs will be stored in clearly labelled folders to allow replicability of the data process.

Documentation & Metadata

6) Documentation & Metadata

Dataset will be analysed using RStudio and Python notebook. Raw files will be processed in RStudio and Jupyter notebook. Python will be used to create a data capture tool. The RStudio will used to process the raw data and to create a data dictionary for the processed data.

Ethics & Legal Compliance

7) Ethics & Legal Compliance

We must be adhered to GDPR and obtain consent from the end-user to process and share the data collected with the data capture tool.

Storage and Back-Up

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

All information (dataset raw files, R scripts, Jupyter notebook scripts, data dictionary) and data will be stored in secure storage location such as DataStore.

Selection and Preservation

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

To ensure the long-term preservation and storage of your data it is best to deposit it in a recognised research data repository. This could be either DataShare or DataVault (<https://www.ed.ac.uk/information-services/research-support/research-data-service/after>) at the UoE or a specialist disciplinary repository such as NERC datacentres, NCBI or EBI.

10) Which data will be retained long-term?

Processed data, R scripts, Jupyter notebook scripts, data dictionaries

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 anonymized data can be accessed via open data repository such as DataShare.

Responsibilities & Resources

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

I will be responsible for the research data management of this project.

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

I will require further training as I am a first-year MSc student in Data Science.