Working with data types and structures in Python and R

Administrative Information

1) School or Institute

- Other (please state)

Deanery of Molecular, Genetic and Population Health Sciences

2) Name and Contact details of supervisor(s)

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

3) Project start date

2022-06-03

4) Project end date

2022-07-04

Data Collection

5) Data Collection

To obtain a hospital length of stay (LOS) data (LOS_model) model, we artificially generated hospital data with fictional patients at ten hospitals, including LOS, age, and date status data. Data were generated to learn generalized linear models (GLM) concepts, modeling death or LOS. LOS, death numbers, and age are all integer,

Documentation & Metadata

6) Documentation & Metadata

The data I will manage are from the NHSRdatasets package. This package has been created to support skills development in the NHS-R community (https://nhsrcommunity.com). It contains several free datasets including Hospital length of stay (LOS) data (LOS_model) which I have selected.

Ethics & Legal Compliance

7) Ethics & Legal Compliance

This project fundamentally follows general data protection regulation (GDPR) since it is commonly implemented with the most excellent principles in many nations. Secondly, many controllers have roles in mitigating risks by following human rights and GDPR. In addition, DPIA is used to detect the risks.

Storage and Back-Up

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

The secure data collection process should not use flash drives and paper to record all the data. Moreover, the information is stored and backed up to digitally safer NHS servers with highly protected networks using a more guarded app made in Scotland for uploading all data. The backup is saved in the UoE's DataStore facility as primary back-up and on One Drive for Business as secondary back-up.

Selection and Preservation

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

For long-term preservation and storage of your data, it is best to deposit it in a recognised research data repository of DataShare and DataVault (https://www.ed.ac.uk/information-services/research-support/research-data-service/after) at the UoE.

10) Which data will be retained long-term?

We do not need personal information like name, or contact details, including telephone numbers and addresses. Necessary information is variable to measure the achievements of the research.

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?

Publication and secondary uses of datasets for future research and project for public benefits.
During the project we will grant internal collaborators access to the data via DataStore and DataSync, the university's secure storage and sharing platforms. When using DataSync the data will be encrypted at all times and the encryption keys shared with collaborators in a secure manner. Once the project is complete, we will preserve our personal, sensitive research data in a closed repository: DataVault. We will create a public metadata record of the deposit in there.

Responsibilities & Resources

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

Yoku Kato, principal researcher, s2272740@ed.ac.uk.

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

This could be RDM training, data storage capacity, data repository deposit fees, or other costs directly related to the excellent management and sharing of the research data. As a member of the University you have access to many free data management services, in particular DataStore and DataSync , and DataShare. However, you will incur a charge if you decide to use more than your allocated quota for DataStore, or choose to preserve your data in DataVault.