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,
University of Edinburgh

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

NHS artificially generated hospital data with fictional patients at ten hospitals, including LOS, age, and death number for skill development. The data format is CSV format. The generated data will be kept and added to the NHSR dataset server in the Edingburgh DataStore. The DataStore is backed-up daily.

Documentation & Metadata

6) Documentation & Metadata

We will adhere to the FAIR Principles. Documentation will be available as part of the project's Notebook repository. The notebook will also contain information on R libraries and packages used and a folder with scripts and raw data as CSV format file.CSV file is usually common to load to R programming and Python and other software for analysis.

Ethics & Legal Compliance

7) Ethics & Legal Compliance

Regarding Ethics & Legal Compliance, data protection impact assessment (DPIA) is firstly used to detect the risks. We follow general data protection regulation (GDPR). We should keep in mind the extra sensitive nature of health data. Due to the artificially generated data, personal data will not be captured and stored.

Storage and Back-Up

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

All the datasets should be stored digitally in the Git hub as working datasets for daily routine work. The backup is saved daily in the UoE's DataStore facility as primary back-up and on One Drive for Business as secondary back-up biweekly. We have to do synchronization manually at the end of the working session.

Selection and Preservation

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

For long-term preservation and storage of your data, they deposit it in a recognized research data repository of DataShare as primary location and DataVault as secondary location(https://www.ed.ac.uk/information-services/research-support/research-data-service/after) at the UoE. For data sharing for future uses, researchers obtained consent from participants joining research projects when they are required to share strategies to minimize restrictions on sharing.

10) Which data will be retained long-term?

Preservation is a crucial component of effective data management. Once the project ended, data will only be made available to outside parties after patient data is anonymised. Therefore, we do not keep personal information like names, or contact details, including telephone numbers and

addresses and essential information is variable to measure the achievements of the research. Our team will keep the original and gathered data and the R scripts and structure (library/packages) and add an explanatory R markdown file for documentation purposes.

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 are vital 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?

B210624, principal researcher, B210624@ed.ac.uk.

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

We can find online training on file formats by completing the MANTRA module 'File formats & transformation':

http://mantra.edina.ac.uk/fileformatandtransformation/

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.

  Members of the University can access many free data management services, particularly, DataStore, DataSync, and DataShare. We incur a charge when using more than we allocated quota for DataStore.