|  |  |  |  |
| --- | --- | --- | --- |
| **0. Project title, author, version and date** | | | |
| *Project: Understanding and predicting the behavioural footprint of older adults in the UK* | | | |
| *Author: Thabsheera PK* | | *Version: 1* | *Date: 01/06/22* |
| **1. Description of the data** | | | |
| **1.1 Type of study**  *Up to three lines of text that summarise the type of study (or studies) for which the data are being collected.*  *The given data is a diary data with activities recorded for every 10 minutes of the day. Along with the diary data, household interview and individual interview data is available. All these data are used to design and run a time series classification model to predict the behaviour of older adults*   * 1. **Assessment of existing data**   *An explanation of the existing data sources that will be used by the research project, with references. Or an analysis of the gaps identified between the currently available and required data for the research.*  *The United Kingdom Time Use Survey, 2014-2015 (UKTUS) is a large-scale household survey that gives data on how people aged 8 years and above in the UK spend their time. The source for the mentioned data is ‘uk data service’ and the data is publicly available. (*Sullivan, O., Gershuny, J. (2021). *United Kingdom Time Use Survey, 2014-2015*. [data collection]. UK Data Service. SN: 8128, [DOI: 10.5255/UKDA-SN-8128-1](http://doi.org/10.5255/UKDA-SN-8128-1) )  *The study provides the time use of individual over on a weekday and a weekend. The time diary provides information about the activities, location, presence of computer or phone, level of enjoyment etc. The questionnaire data provides the information of characteristics of the household and individual such as employment, household type, demographics information etc.*    **1.2 Types of data**  *Types of research data to be managed in the following terms: quantitative, qualitative; generated from surveys, interviews, administrative records, simulations, images. Include the use of any code or software if it aids the creation or processing of the data.*  *Majority of the data fields are categorical or qualitative data. Since the data is generated from surveys, most of the fields are qualitative such as demographics, employment, activities etc.*  **1.3 Format and scale of the data**  *File formats, software used, number of records, databases, (in terms that are meaningful in your field of research). Do formats and software enable sharing and long-term validity of data? How large is the data going to be?*  *The data is available in spss format, Used R to read and visualise the data. The diary data is available in two data type of data sets, one is the long format where each row is for a household- person- day of the week- time interval. The other format is wide, where each time interval is transposed into columns. Both have the same data of 4000 households and 8000 individuals.*  *The questionnaire data is of the same households and individual but their characteristics. There are of\_\_\_\_ records in household data and \_\_\_\_\_ in individual data* | | | |
| **2. Data collection / generation** | | | |
| *Focus on the good practice and standards for ensuring new data are of high quality and processing is well documented.*  **2.1 Methodologies for data collection / generation**  *How the data will be collected/generated and which community data standards (if any) will be used at this stage.*  *Since the data is a publicly available UK time use survey for the period of 2014-15 , the data collection was effortless. There was only a need to create an account in UK data service website and fill the form stating the purpose of the data.*  **2.2 Data quality and standards**  *How consistency and quality of data collection / generation will be controlled and documented, through processes of calibration, repeat samples or measurements, standardised data capture or recording, data entry validation, peer review of data or representation with controlled vocabularies.*  *The data has gone through lot of cleaning up and calibration before getting uploaded to UK data service. Weights have been given to household and individual so that it will be comparable to the entire households and population of UK. Some of the missing data has been replaced by the average values of the existing records or some of them were removed.*  *Currently, there’s no duplicate records in the data.* | | | |
| **3. Data management, documentation, and curation** | | | |
| *Focus on principles, systems and major standards. Focus on the main kind(s) of study data. Give brief examples and avoid long lists.*  **3.1 Managing, storing and curating data.**  *Briefly describe how data will be stored, backed-up, managed and curated in the short to medium term (*[*https://www.ncl.ac.uk/library/academics-and-researchers/research/rdm/working/*](https://www.ncl.ac.uk/library/academics-and-researchers/research/rdm/working/)*)*  *The data, which was publicly available in the UK data service, is downloaded and saved in the local system. The codes and generated small tables are getting backed up in the git hub.*    **3.2 Metadata standards and data documentation**  *What metadata is produced about the data generated from the research? For example descriptions of data that enable research data to be used by others outside of your own team. This may include documenting the methods used to generate the data, analytical and procedural information, capturing instrument metadata alongside data, documenting provenance of data and their coding, detailed descriptions for variables, records, etc* (see: [*www.dcc.ac.uk/resources/metadata-standards*](http://www.dcc.ac.uk/resources/metadata-standards) / <http://www.data-archive.ac.uk/create-manage/document>)  *The meta data and data dictionary for the data is available in the UK data service* [*website*](https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=8128#!/documentation)*. The data dictionary has the information about each column and what the value in that column means.*  *All the data manipulation that has been done as part of this project is properly commented in the codes and it is easily reproducible. The code changes are timely updated in the git log.*  *The entire procedure for the project has been properly documented and uploaded as a ReadME file in the github.* | | | |
| **4. Data security and confidentiality of potentially disclosive information** | | | |
| *This section should be completed if your research data includes* ***personal data relating to human participants in research****. For other research, the safeguarding and security of data should also be considered. Information provided will be in line with your ethical review. Please note this section concerns protecting the data, not any potential patients.*  **4.1 Main risks to data security**  *All personal data has an element of risk. Summarise the main risks to the confidentiality and security of information related to* human participants*, the level of risk and how these risks will be managed. Cover the main processes or facilities for storage and processing of personal data, data access, with controls put in place and any auditing of user compliance with consent and security conditions. (*[*https://www.ncl.ac.uk/research/researchgovernance/ethics/gdpr/*](https://www.ncl.ac.uk/research/researchgovernance/ethics/gdpr/)*)*  *The data is masked for the identity, the household number represented by a serial number. This serial number to exact household mapping is not available anywhere publicly, this is only known to the people responsible for the survey. There’s no description to understand the individual as well, the field pnum(person number) is considered as individual.* | | | |
| **5. Data sharing and access** | | | |
| *Identify any data repository(-ies) that are, or will be, entrusted with storing, curating and/or sharing data from your study, where they exist for particular disciplinary domains or data types. [Information on repositories is available here.](https://www.ncl.ac.uk/library/academics-and-researchers/research/rdm/sharing/)*  **5.1 Suitability for sharing**  *Is the data you propose to collect (or existing data you propose to use) in the study suitable for sharing? If yes, briefly state why it is suitable.*  *If No, indicate why the data will not be suitable for sharing and then go to 5.3.*  *The data is suitable for sharing within the organisation in this case it’s Newcastle University.*  **5.2 Discovery by potential users of the research data**  *Indicate how potential new users (outside of your organisation) can find out about your data and identify whether it could be suitable for their research purposes, e.g. through archiving into a research data repository. Will the repository create a persistent identifier (e.g. DOI) to be included in research outputs, including publications?*  *The data is saved in the local system, and which is originally a publicly available data, so anyone can download from the UK data service after submitting relevant information.*  **5.3 Data preservation strategy and standards**  *How long will the data be archived and shared for? Will any data be preserved but not shared? If yes, what is the plan for this long-term archived?*  **5.4 Restrictions or delays to sharing, with planned actions to limit such restrictions**  *Restriction to data sharing may be due to participant confidentiality, consent agreements or IPR. Strategies to limit restrictions may include data being anonymised or aggregated; gaining participant consent for data sharing; gaining copyright permissions. For prospective studies, consent procedures should include provision for data sharing to maximise the value of the data for wider research use, while providing adequate safeguards for participants. As part of the consent process, proposed procedures for data sharing should be set out clearly and current and potential future risks associated with this explained to research participants.* | | | |
| **6. Responsibilities and Resources** | | | |
| *Are there any resources (e.g. storage/ training) that you will require to fulfil the plan?* | | | |
| **7. Relevant institutional, departmental or study policies on data sharing and data security** | | | |
| **Policy** | **URL or Reference** | | |
| Data Management Policy & Procedures | <https://www.ncl.ac.uk/media/wwwnclacuk/research/files/ResearchDataManagementPolicy.pdf> | | |
| Information Security | [*https://services.ncl.ac.uk/itservice/policies/InformationSecurityPolicy-v2\_1.pdf*](https://services.ncl.ac.uk/itservice/policies/InformationSecurityPolicy-v2_1.pdf) | | |
| Other |  | | |