A Data Management Plan created using DMPonline

Creator: James Marshall

Affiliation: University of Sheffield

Template: University of Sheffield

Grant number: EP/P006094/1

Last modified: 12-03-2018

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customise it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

**Admin Details**

**Project Name:** Professor Green Data Management Plan

**Principal Investigator / Researcher:** Professor Green

**Project Data Consult:** Data Masters

**Institution:** Dalhousie University

1. **Data Collection**

In this case, three data collection tools were used during the experiment:

1-    Text documents:

383 individual documents in different format PDF, MS word, and plain text describing the teams, their outcome and practices. It is obtained from healthcare organization and transcription of the interviews in MS word format.

2-    Spreadsheet:

Quantitative data about these documents is saved in Excel spreadsheet.

3-    Audio file

15 Interviews is saved in mp3 format which are transcribed in word document.

4-    TVS file:

The open data set regarding healthcare, including healthcare outcomes, healthcare expenditures, healthcare staffing, enrolment by discipline.

**1.1. File Naming Convention**

This project will have two folders of the datasets. First folder will have the original datasets and the second one will have the datasets after translation. The naming convention for the original main folder of datasets should have the “Original” word, then project name and the date, for example “Origenal\_HealthCare\_20180212”. The subfolders should have the country name to recognize where the data come from. For example, Canada, Germany, and France. The Canada folder will have two folders, named "Nova Scotia, and Calgary", which will have different folders based on the name of institutions. Files' names will have the type of the file, participant’s ID (if applicable), then language codes such as Germany “DE”, Canada “EN” and French” FR”. For example, an audio file should be named “audio\_P1\_EN”.

As we explained above:

1-    The naming convention should be descriptive.

2-    It should distinguish between the original data and the translated data.

3-    It should contain the date of the data collection, in the format (YYYYMMDD).

4-    It should show the type of data attached to the participant’s ID if applicable.

We believe this way of naming convention of the folders, subfolders and files, will help other researchers to easily access and use of the data.

**2. Documentation and Metadata**

The metadata will be written in a text file called “README”, that includes descriptive information of the data to ensure that future users will understand the data. The README.txt file will include the information in the table below:

|  |  |  |
| --- | --- | --- |
| General information | Creator | Professor Green |
| Title | Teamwork in Hospital Environments |
| Date | April 10, 2018 |
| Funding agencies/period | 10 Years |
| Keywords | Hospital, Teamwork, Stress. |
| Coverage | A few hospitals in Nova Scotia, one hospital in France, one in Germany and a few more in Calgary. |
| Funding | CIHR (Canadian Institutes of Health Research) |
| Access information | Access restrictions | Project members can access the note taken under the supervision of the main author (professor Green). |
| Copyright | Exist |
| Technical details | File format | Text documents (Docx, plain text, PDF).  Excel Spreadsheets (quantitative data about these documents  Audio MP3 files (interview) |
| Count of files | 383 individual text documents.  15 audio files. |

**Ethics and Legal Compliance.**

In terms of accessing the sensitive information, the only selected group members will be given the right to access to the notes that is created by Prof. Green.

For the privacy of the participants, each one will be given ID for identification, to make them anonymous and keep them confidential. Also, the participant’s ID will be used in all materials. The anonymity of textual data will be preserved by using ID in any presentation or publication.

**Storage and Backup**

This research should have a storage system that will meet the requirements of the project. It is important to employ multiple methods to backup and copy the project data. We need to look into three aspects of data storage which are storage space, cost and security.

Each document file size ranges from a few hundred kilobytes to 25-30MB and each interview is an hour long, and is encoded them in mp3 format, 128kbps. So, we anticipate the maximum size of the whole documents will be around 11.5 GB. Moreover, we have 15 interviews and the length of each audio is 128kbps so the size of the whole audio files will be 823.5 MB.

In term of the first aspect which is storage space, using USB drive with 64 GB will fit the data size right now because the usage space is 24 GB which means 40 GB is free. However, the data will expend for upcoming 10 years in addition to translated data so we need more space to save the data.

In term of cost and security, using USB drives is cheap and easy solutions, but it is not secure and can be easily destroyed or lost.

We anticipate that size of the data will be more than 72 GB if we consider the upcoming years and translated data for these years. Consequences, data needs to be stored in large and secure space. Sensitive data should not be stored on any repositories

We recommend that you use multiple forms of storage:

1.     The Dataverse: it is open source research data repository software provided by Dalhousie university that allows researchers to deposit and share data openly or privately. The data is hosted on Dalhousie's servers. The service is primarily for those affiliated with Dalhousie University (DalLibGuides, 2018).

2.     pCloud:  it offers 500 GB with one payment “175$ “for lifetime. With pCloud researcher can upload any type of files directly to his/her account regardless of the size. It is more secure because information is encrypted using TLS/SSL when it is transferred from the device to the pCloud servers (Pcloud.com, 2018).

3.     Sync.com: it offers 1 TB of secure file storage by 5$ / user /month. it is used end-to-end encryption and No third-party tracking for privacy protection (Sync, 2018).

We also recommend using these storages because it is easy to access and control outside the office.

Researcher needs to maintain a master copy of the data on external local such as Dataverse or external remote such as pCloud and Sync.com.

**Selection and Preservation**

Interview data cannot easily be recreated or produced so it has a long term value that needs to be preserved. the other data documentations are necessary to validate research findings so it is required to keep them. Data will be preserved and available for at least 10 years under the control of Professor Green. The data can be stored in the following data storage:

·      The Dataverse

·      pCloud

·      Sync.com:

**Data Sharing**

The audio files and transcriptions cannot be publicly shared because it contains potentially identifying information of participants. Other quantitative data cannot be also released until this research is published to protect intellectual property rights of the researcher. Data will only be shared with research team by Professor Green if they need more information during preparing the data and analysis the data with anonymity of participants using Dropbox (for audio files), Google Docs (for transcriptions) and Zotero (for documents).

CC BY-NC-ND license will be applied to the project. It will not allow anyone to make any modification on the original data or use it for commercial purposes. However, it will allow other researchers to copy and utilize only the original copies of the work, if they have Professor Green's permission.

**Responsibilities and Resources**

Our consult group will be responsible to manage the data after Professor Green is satisfied with what we have come up with. After the project is completed, the new consults will be responsible to manage the project data if Professor Green has new funding.