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

**Data Collection**

In this case, four data collection tools are used during the experiment:

1. **Text Documents:**

383 individual documents in different formats—PDF, MS Word, and plain text describing the team, their outcomes, and practices obtained from healthcare organization and transcription of the interview in MS Word format.

1. **Spreadsheet:**

Quantitative data regarding these documents is saved in an Excel spreadsheet.

1. **Audio File:**

15 interviews are saved in an MP3 format, which will be transcribed into a Word document.

1. **TVS File:**

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

**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 word “Original” followed by the project name and date. For example: “Original\_HealthCare\_20180212.” The subfolders should include the country name to clarify where the data came (Canada, Germany, or France). The Canada folder will have two subfolders labelled “Nova Scotia” and “Calgary,” which will have different subfolders as based on the institutions’ name. File names will include the type of file, participant’s ID (if applicable), then a language code, like “DE” for Germany, “En” for Canada, and “Fr” for France. For example, “audio\_P1\_En.”

As explained above, the naming convention:

1. Should be descriptive.
2. Should distinguish between the original data and the translated data.
3. Should have the date of collecting the data (YYYYMMDD) to clarify when the data was collected.
4. Should identify the type of data attached to participant’s ID, if applicable.

We believe that these methods for naming and structuring folders and subfolders will help other researchers more easily access and use the data.

**Documentation and Metadata**

The metadata will be written in a text file labelled “README,” which includes descriptive data information to ensure future users understand the data. The “README” file will include information from the table below.

|  |  |  |
| --- | --- | --- |
| **General Information** | Creator | Professor Green |
| Title | Teamwork in Hospital Environments |
| Date | April 10, 2018 |
| Funding Agencies/Period |  |
| Keywords | Hospital, Teamwork, Stress |
| Coverage | A few hospitals in Nova Scotia, a few in Calgary, one hospital in France, and one in Germany. |
| Funding | CIHR (Canadian Institutes of Health Research) |
| **Access Information** | Access Restrictions | Project members can access notes taken under the supervision of the main author (Professor Green). |
| Copyright | Exist |
| **Technical Details** | File Format | Text documents (PDF, Docs, plain text)  Excel Spreadsheet (quantitative data regarding these documents)  Audio MP3 files (interview) |
| File Count | 383 individual text documents  15 audio files |

**Ethics and Legal Compliance**

In terms of accessing sensitive information, only the selected group members will be given access to the notes created by Professor Green. Regarding participant privacy, each participant will be given an ID for identification purposes to ensure anonymity and protect personal privacy. The participant’s ID will also be used in all materials. Furthermore, the anonymity of textual data will be preserved by using an ID in any presentation or publication.

**Storage and Backup**

This research should include a storage system that meets the project’s requirements. It is important to employ multiple methods for backup and copying project data, and we must consider three aspects of data storage: space, cost, and security.

Each document’s file size ranges from a few hundred kilobytes to 25-30 MB. Each interview is an hour in length, or 128 kbps, and encoded into MP3 format. We anticipate the maximum size for all documents to be approximately 11.5 GB. Moreover, we have 15 interviews and each audio’s length is 128 kbps, so the complete size for audio files will be 823.5 MB.

Regarding the first aspect—storage space—using a 64 GB USB drive is sufficient for the data size; the usage space is 24 GB, so a remaining 40 GB is available. The data will expend within 10 years, however, in addition to translated data, so will require more space to save data.

Regarding cost and security, a USB drive is an inexpensive and simple solution, but is not secure and can be easily destroyed or lost.

When considering the upcoming years and future translated data, we anticipate the total data size to be more than 72 GB. Consequently, data needs to be stored in large and secure spaces, and sensitive data should not be stored on any repositories.

We recommend using multiple forms of storage:

1. **The Dataverse:** An open source research data repository provided by Dalhousie University. This software allows researchers to deposit and share data either openly or privately. The data is hosted on Dalhousie's servers and the service is primarily for those affiliated with Dalhousie.
2. **pCloud:** A server offering 500 GB with one payment of $175 for a lifetime of use. With pCloud, researchers can upload any file types—regardless of size—directly to his or her account. This option is more secure because information is encrypted using TLS/SSL when transferred from a device to the pCloud servers.
3. **Sync.com:** A system with 1 TB of secure file storage for $5 per user per month. The system uses end-to-end encryption and no third-party tracking to ensure privacy protection

We also recommend using these storage systems because of the ease of access and available control outside the office. Furthermore, researchers should maintain a master copy of data on an external locale, such as The Dataverse, or an external remote, such as pCloud or Sync.com.

**Selection and Preservation**

Interview data cannot be easily recreated or reproduced; thus, its long-term value must be preserved. Moreover, other data documentations are necessary to validate research findings, so data must be kept. Under the control of Professor Green, data will be preserved and available for at least 10 years, and can be stored in the following data storage systems:

* The Dataverse
* pCloud
* Sync.com

**Data Sharing**

The audio files and transcriptions cannot be publicly shared because they contain potentially identifying information regarding participants. Other quantitative data cannot be released either until the research is published; this ensures protection of the researcher’s intellectual property rights. Data will only be shared with a research team by Professor Green, and this will only occur if they require additional information while preparing the data and analyzing the data with participant anonymity via Dropbox (audio files), Google Docs (transcriptions), and Zotero (documents).

A CC BY-NC-ND license will be applied to the project. It will not allow others to make modifications on the original data or use the data for commercial purposes. It will allow other researchers, however, to copy and utilize original copies of the work with Professor Green's permission.

**Responsibilities and Resources**

Our consult group will be responsible for managing the data after Professor Green approves our plan. After project completion, the new consults will be responsible for managing project data if Professor Green receives new funding.