**Case-2:turki**

**Brains on Board**

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

**Data Collection**

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

1- Text documents:

383 individual documents in different format PDF, MS word, and plain text describing the team, their outcome and practices obtained from healthcare organization and transcription of the interview 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 will be transcribed in word document.

4- TVS file:

The open data set regarding healthcare, including healthcare outcomes, healthcare expenditures, healthcare staffing, 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 “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 French. the Canada folder will have two folders named "Nova Scotia, and Calgary" which will have different folders based on the name of institutions. Files' name will have the type of files, participant’s ID (if applicable), then language codes such as Germany” DE”, Canada” En” and French” Fr”. For example for files' name is ” audio\_P1\_En”.

As we explain above:

1- The naming convention should be descriptive.

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

3- It should content the date of collecting the data (YYYYMMDD)to know in which date the data was collected.

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

We believe this way of naming and structure of folder and subfolder will help other researcher to easy access and use of the data.

**Documentation and Metadata**

The metadata will be written in a text file called “README” that includes descriptive information of the data to ensures that future users will understand the data. The README text 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 |  |
| 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). |
| Copy Right | Exist |
| Technical details | File format | Text documents (Docs, plain text, PDF).  Excel Spreadsheet (quantitative data about these documents  Audio MP3 files (interview) |
| Count of files | 383 individual text documents.  15 audio files. |

**Ethics and Legal Compliance.**

in term of accessing the sensitive information, the only selected group member will be given the right to access to the notes that is created Professor Green.

In term of participant’s privacy, each participant will be given ID for identification to make them anonymous and protect their privacy. 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. [1](http://dal.ca.libguides.com/rdm/terms)

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. [2](https://www.pcloud.com/cloud-storage-pricing-plans.html)

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. [3](https://www.sync.com/pricing/)

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.

**Case-3: Nazia**

**Brains on Board**

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

**Data Collection**

Neural and behavioural data on behaving honeybees in closed-loop VR flight simulator

Behavioural data on free-flying bees over short and long ranges Neural network simulation data (input, output, internal state)

Telemetry from various robotic platforms (ground-based, 3-d gantry, free-flying robot)

Virtual reality flight simulator with torque meter and single-cell electrophysiology apparatus

High-speed camera (short range bee flight) and harmonic radar (long range bee flight)

Neural network simulation on virtual and real sensor input, in control of virtual or real robot platform

Vicon motion capture data (ground-based and free-flying robots indoors), on-board robot sensors (e.g. IMU), controller input and output

**Documentation and Metadata**

For private data repositories, DataCite metadata will be generated via Figshare based on the compulsory fields, plus a brief textual description field.

For public data repositories, metadata will be generated as above, and accompanying documentation will be published in the appropriate form for the medium (e.g. Readme.md file for Github repositories, online supporting information file for publications associated with data respositories).

**Ethics and Legal Compliance**

Principal Researchers and their institutions hold the copyright for the research data they generate. By depositing with the university, investigators do not transfer copyright but instead grant permission to distribute the data and to transform the data as necessary to protect respondent confidentiality and improve utility.

***Informed consent***: For this project, informed consent statements, if applicable, will not include all those languages and terminologies that would prohibit the data from being shared with the research community and create ambiguity.

***Disclosure risk management*:** Once deposited, the data will undergo procedures to protect the confidentiality of individuals whose personal information may be part of archived data.

***Format – Submission***: The data and documentation will be submitted to the researcher according to the set criteria by the researcher.

Created using the DMPonline. Last modified 12-03-2018 2 of 4

**Storage and Backup**

The data got sensitive information as the researcher works on Information systems,For this purpose two type of data is created

1)Open Data : This will be

Research data will be stored on appropriate cloud storage services (e.g. Sheffield Google Drive for project data, Figshare (Sheffield and Sussex) and Open Science Framework (QMUL) for repositories), or institutionally-provided, safeguarded internal storage services

Sensitive data will be stored on private repositories required authenticated access

**Selection and Preservation**

Model and controller structures

Empirical robot data

Behavioural and neural data from animal experiments

Archiving of freely available data on recognised stable long-term repositories (e.g. GitHub, etc.)

Archiving of non-public raw data via institutional storage services.

Archiving of non-public processed data via private online repositories (Figshare (Sheffield and Sussex), OSF (QMUL))

**Data Sharing**

The data is categorized into two forms

1)Open Data : Thi will be share for all the users

2)Closed Data :The professor got a backup of lists.Non open data is shared on a case to case basis. The professor got spreadsheet to keep track of who has access to what.

**Responsibilities and Resources**

PI and Project Manager (brainsonboard-coordinator@sheffield.ac.uk)

Resources as already provided by open-source providers and institutional partners

Created using the DMPonline. Last modified 12-03-2018 3 of 4

**Case-4: Ychele**

**Brains on Board**

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

**1. Data Collection**

The type of data that is collected in this case is open data, such as data about science of science, which is mostly available from public domain resources. The important collection of data (i.e. medical data) which is collected from database called PubMed is available via web API and is served and collected in JSON file format. The plan is to keep the JSON format and by pass the Excel sheet format and directly insert (query) the JSON files format into a SQL relational database (…). By doing this you can keep querying your own inputs as you like and offers less confusion as the whereabouts of the data type, and facilitates the data sharing. Furthermore, JSON file format will allow your data to be reused, and with the database it is easily shared with access control in the SQL Server, and is able to be preserved for a long time in the SQL relational database Server Cloud.

For this plan we recommend Microsoft Azure SQL Database, which is an intelligent relational cloud database service. The main advantages this database offers, as it can handle big data, it supports JSON objects and XML objects, Cloud storage and backup for file loss prevention, and version control when data sharing will become necessary. The fees for this option is displayed in the budget plan.

There is a free option by using the University cloud drive, it does not offer the advantage of using SQL database, but it is a institutional OneDrive cloud that offers storage, collaborative work through Office documents, access space via the web, personal storage with syncing, sharing and reviewing the version history of a file. Furthermore, it offers group storage which is separate of the users own storage, sharing via permission and collaborative editing. This will help with storage and backup, as it offers up to 1TB space for every faculty and student member of the university.

1.1. Public Data Collection Optimization

There are two ways to optimize collection of public data, especially in PubMed, instead of doing it by hand into an excel sheet. The first way is to learn the method of Web API public data download in JSON format, then to save the JSON files, back them up and input them directly into Azure SQL Database by querying. This is a more technical way and will take our team a while to familiarize with, but it offers a more efficient way of handling the data into a database, and this will lead to a more consistent and structured organization.

The second way is much simpler. In PubMed, after keyword search and a list of results have appeared. Next to the articles a check box appears, the user can select which index information of the articles they want, go to the Send to tab up in the right corner and select the appropriate output file format they desire. Choices ranges from, TXT files, CSV files, XML files, etc. This is much easier than doing it by hand; however, there is still the option of a messy excel sheet if CSV file is chosen as an output. If XML format is chosen it will be easy to query it as well into the database just like the JSON file format.



***Figure 1:*** *Visual Process of simpler method of public data collecting on PubMed*

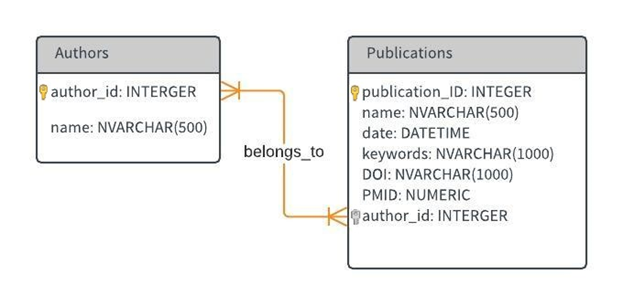
1.2. ER modeling Naming Convention

In order to structure and name your files and help you understand, how your data needs to be organized an Entity Relationship modeling naming convention and procedure is used. This model helps to give a visual layout on how best to organize and label your Metadata in the database. The Metadata labels that is used are date, name, keywords, publication, authors, DOI and PMID. As many of the important sources originates from PudMed, we recommend that their unique document identifier label is used and each of these PMID. DOI (i.e. digital object identifiers) which helps to identify journal articles and scientific articles, is also a Metadata standard that we recommend that is used, as this meta area of research is comprised of science of science.

1.3. Version Control

As this project takes into consideration, future collaborative work with other colleagues, version control is needed. However, for databases this is different. For SQL databases this is called SQL source control. SQL source controls are software’s, where it is installed and connected with your database. After installation the source control keeps track of the database, the transient which is a local copy of your latest revision of database and the working base which is a copy of the database the last time the user ran a commit or pulled from the database (…). Then a Three way comparison is performed and a list of changes is then generated. Here is a list of free recommended SQL source version control that is compatible with Azure; Flyway and Visual Studio database project. Here is a list of commercial priced source controls; Red Gate, ApexSQL source control, and DB Ghost Change Manager (…). Depending on your needs our IP’s will install the one that fits perfectly with your needs. All pricing is discussed in the budget plan.

**2. Documentation and Metadata**



***Figure 2:*** *ER Diagram Schema with Primary and Foreign Key restraints*

**3. Ethics and Legal Compliance**

As the most important sources are derived from PubMed, our PI’s familiarized themselves with PubMed licensing agreement. This data management plan will follow these summarized Copyrighted, Policy and guidelines, which was gathered from PubMed website:

Copyright and Restrictions (...):

1. There are both copyrighted and non-copyrighted, and creative commons materials found on PubMed. It will be up to the user to figure it out. But, when in doubt assume it is copyrighted and contact publisher for permissions. To find out those that are covered by creative commons license can check PMC Open Access Subset.
2. Materials produced by U.S. government employees as part of their official duties in not copyrighted within the US.
3. Works are commonly protected by U.S. and international copyright laws, if they are created by organization other than the federal government.

Restrictions on Systematic Downloading of Articles (...):

1. The NCBI`s systems may not be overloaded, by numerous queries.
2. Bulk downloading is prohibited.
3. Crawlers and similar robotic method are not allowed to be used to systematically retrieve content from PubMed web site.
4. There are only two services allowed PMC OAI service and the PMC FTP service are permissible for automated downloading of articles.

Endorsement (...):

1. No commercial use.

**4. Storage and Backup**

The Azure Database package comes with cloud services for storage and backup for the database. Thus, the database will be stored and backed up with Microsoft cloud.

For the research data, which is in JSON format, should be stored and backed up before inputted into the database and multiple backups is needed for this type of project. First the JSON files should continue to be backed up on your external hard drive. However, this is not enough security against file loss prevention thus another appropriate cloud storage services is required, and Open Science Framework (OFS) cloud service seems to be a good choice, as it is free, supports JSON objects, can easily archive your data, controlled access and easy collaboration, amplify your workflow, version control, and can easily manage your project using their one dashboard (…). OFS also offers openness to view what fellow researchers are up to. You can also choose which data to share and which to maintain private.

As this is data collected from public repositories there are no sensitive data to worry about, but Azure Database and Cloud service offers the optimum security with required authenticated access. OFS offers strong security as well.

Both database and research data should be backed up on a weekly basis by Prof. Chartreuse and both OFS and Azure offers sufficient storage, for Azure the first 12 months offers a free package size of 250GB for the database and for OFS its 5Gb in initial registration however much larger size can be obtained by emailing their support team.

**5. Selection and Preservation**

Prof. Chartreuses is very dedicated to his research and plans to continue his work until the day he is no longer here. Thus, this database and research data has a long-term value of at least minimum of ten years and needs to be preserved well. Prof. Chartreuse will continue to control and preserve this data until he sees fit to archive his freely available data on a recognized stable long-term repositories such as Github, OFS, or data repositories that offers data curated services such as institutional repositories, Dryad Data Repository or the ICPSR (Inter-university Consortium for Political and Social Research) (...). The data that must be retained are those that leads to publication and records are selected based on the records of interest to the professor.

**6. Data Sharing**

As the data is collected from public sources such as PubMed, their licensing policy must be upheld. Their main policy about data sharing is that it may not be used for commercial purposes. Their copyrighted and creative commons policy must also be maintained as discussed. This data will be shared with graduate students and fellow colleagues in the same field as Prof. Chartreuse. The professor decides who to give access too. Data sharing will possible through both Azure SQL Database Cloud Server and OSF cloud server and their search engine. If colleagues wishes to collaborate with prof. Chartreuse on his Azure SQL database, then they will have to contact the professor first with their necessary contact information such as name and email, and then he will proceed to make a user credential for the colleagues in Azure, with his chosen set of restriction parameters, and then the colleagues will be prompted to choose their username and password.

**6. Responsibilities and Resources**

Our consultant team is responsible with familiarizing themselves with the Web API method of downloading JSON format file in order to input directly these JSON files into to SQL database by querying input, and relate it to prof. Chartreuse. In addition, need to be familiarized with all the copyrights and licensing policy while working on this project.

Resources are provided by open-course providers such as OSF. University funding for Azure SQL Database is conditional on well documented reason for the cost.

6.1. Budget Plan Proposal for Azure SQL database

**References APA format:**