

DALHOUSIE UNIVERSITY  
INFO6540  
DATA MANAGEMENT

# Group Project

---

## Data Management Consultation Report

**Ghada Asiri – B00768421**

**Nazia Asad – B00760700**

**Turki Alqahtani – B00721827**

**Ychele De Freitas Braz – B00773857**

**Instructor: Elvira Mitraka  
4/10/2018**

The following three data management plans was created by our virtual company:

**Data Masters**

About our virtual company:

**Today, data is essential for success. Data Masters provides the most successful solutions for data management and analytics, for both individuals and organizations.**

## **1 CASE - 2**

### **1.1 Admin Details:**

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

**Principal Investigator / Researcher:** Professor Green

**Consultant Company:** Data Masters

**Institution:** Dalhousie University

### **1.2 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.
- 2. Spreadsheet:** Quantitative data regarding these documents is saved in an Excel spreadsheet.
- 3. Audio File:** 15 interviews are saved in an MP3 format, which will be transcribed into a Word document.
- 4. TVS File:** The open data set regarding healthcare, including outcomes, expenditures, staffing, and enrolment by discipline.

### **1.3 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 “Original\_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.

### **1.4 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:

<b>General information</b>	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, a few in Calgary, one hospital in France, and one in Germany.
	Funding	CIHR (Canadian Institutes of Health Research)
<b>Access information</b>	Access restrictions	Project members can access the note taken under the supervision of the main author (professor Green).
	Copyright	Exist
<b>Technical details</b>	File format	Text documents (PDF, Docx, plain text) Excel Spreadsheet (quantitative data regarding these documents). Audio MP3 files (interview).
	Count of files	383 individual text documents. 15 audio files.

*Table 1: case-2 documentation and metadata.*

## **1.5 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.

## **1.6 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 64GB 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. 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 (DalLibGuides, 2018).
- 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 (Pcloud.com, 2018).

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 (Sync, 2018).

These services are recommend because they offer easy access and control, even outside the office. In addition, researcher should maintain a master copy of the data on an external local such as DataVerse, or external remote, such as pCloud and Sync.com.

### **1.7 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:

- DataVerse
- pCloud
- Sync.com

### **1.8 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 (while keeping the 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.

## **1.9 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.

## **2 CASE - 3**

### **2.1 Admin Details**

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

**Principal Investigator / Researcher:** Professor Pinkerton

**Consultant Company:** Data Masters

**Institution:** Dalhousie University

### **2.2 Data Collection**

The research project involves basically two forms of data

- Excel spreadsheets (around 17,384 Spreadsheets) exists in different versions of Excel.
- Excel CSV files.

These spreadsheets are basically composed from complex text files, quantitative textual analysis, and other data that are either created by the researcher, or collected from different sources, based on the researcher field of interest (e.g. government, private corporations, and other researchers). The data is being collected through variety of methods and it is roughly 10 years old.

### **2.3 Documentation and Metadata**

The below table includes descriptive information about the data included in the plan. This metadata can be used in future for the SQL data option or the cloud storage option. It will help in the naming convention of the tables in the database option, or files and folders in the cloud storage option:

<b>General information</b>	Creator	Professor Pinkerton
	Title	MIS wizard
	Date	April 10, 2018
	Funding agencies/period	Personal & institutional funding
	Keywords	Managing Information Systems
<b>Access information</b>	Access restrictions	Open data accessed publicly. Closed Data accessed by permission only.
	Copyright	Exist



<b>Technical details</b>	File format	Excel Spreadsheet (different versions). CSV files.
	Count of files	17384 Spreadsheets
	Number of Rows	1-750,000
	Median Row	1000

*Table 2: case-3 documentation and metadata.*

## **2.4 Ethics and Legal Compliance**

### **Copyright and Creative Commons:**

For this plan our team recommends that copyright and creative commons should be taking into consideration. As the research data can contain both copyrighted and non-copyrighted data, or creative commons materials. It will be up to professor Pinkerton professional eye to keep track of this. However, when in doubt assume it is copyrighted and contact publisher for permissions ("Research Data MANTRA", 2018).

### **Consent Form:**

A letter explaining the purpose, approach and distribution strategy (including plans to share data) of the research, and an associated consent form, will be prepared and translated into relevant languages. This letter will be formatted by our team, and that can be used by the researcher team in order to carry any future researches.

### **Intellectual Property Rights:**

Principal researchers and their institutions hold the copyright for the research data they generate. By depositing the data with the university, investigators do not transfer copyright, but instead grant permission to distribute the data, and to transform it as necessary. That is to protect respondents' confidentiality. Online and archival sources will be cited and clearly acknowledged in the database and research outputs.

### **Confidentiality:**

This project contains sensitive data (students performance information). Commitments to ensure confidentiality will be maintained, by ensuring details that can be used to identify participants, are removed from transcripts or concealed in write-ups. Our team will avoid transmitting unencrypted personal data electronically.

**Disclosure risk management:**

Once deposited, the data will undergo procedures (screening processes) to protect the confidentiality of individuals, whose personal information may be part of archived data.

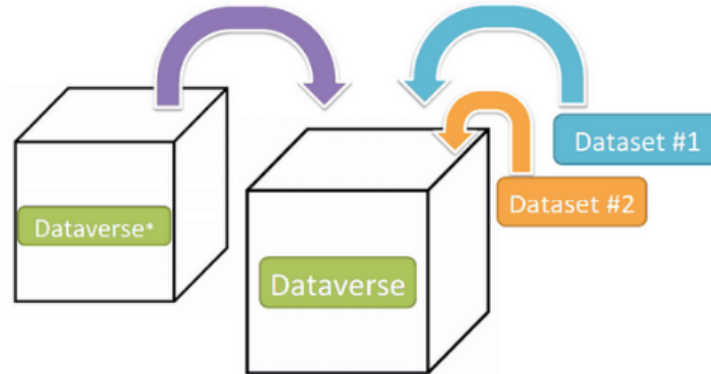
**2.5 Storage and Backup**

This project's research data is considered to be very important to professor Pinkerton, and our team recommends that multiple backups are in order. Storage and backups using both online clouds and hard drive copies should be considered.

First of all, it is recommended to use hard drive as first line of backup, and to keep it up to date as much as possible. One way to do that is by adding and updating the files on the hard drive every month. This would ensure the master copy is safe against any corruptions that could happen on the cloud. High quality products should be used, from reputable manufacturers.

Furthermore, taking into consideration the sensitivity of some of the data in the project, the following cloud services are proposed:

**Plan 1 - DataVerse:** It is an open source repository for research data that can be setup for individual or group use. It can contain different types of data sets, such as research data, codes, metadata, etc. as well as other DataVerses (Figure 1). The (v.4.7) version of DataVerse is offered freely by Dalhousie University to researchers from faculty members and students. It allows researchers to store and share data privately or publically. That would help professor Pinkerton in classifying her data into either public or private, to make sharing process easier for her, in a way that she will not need to keep using emails. The data will be hosted on Dalhousie's servers, which will ensure a good security as well. In addition, DataVerse allows researchers to discover data, help them obtain web visibility, academic credit and increased citation counts. To set up DataVerse, Dalhousie Libraries Research Data Management can help in doing that, and they will also make sure to grant the administrator rights to the owner. An administrator to a DataVerse can manage all the setting related to that DataVerse from themes and widgets to permissions, roles, access and more (DalLibGuides, 2018). To conclude, DataVerse meets almost all professor Pinkerton data management requirements, that is why it is highly recommended by our team.



**Figure 1:** DataVerse in 4.0 (DalLibGuides, 2018).

**Plan 2 - OneDrive:** Dalhousie University offers both students and faculty members the advantage to use their institutional funded OneDrive Cloud. Every Dalhousie faculty member, staff member, student, or different committees (e.g. Computer Science Graduate Society) is allotted their own 1TB of space for storage. Additionally, they have the benefit of using Microsoft Office 365 for collaborative work, and each group is provided with 1TB of storage (separate from individual user storage). My team recommends that research data should be stored on the University's server, because it is a high quality, enterprise-class storage with guaranteed backup and resilience (Dalhousie, n.d.).

**Plan 3 - Amazon Web Services (AWS):** This plan entails inserting the Excel data into a structured relational database. Now, taking into consideration the future potential growth of the data, our team recommends using Amazon Aurora MySQL relational database, with Amazon Web Cloud Services. Amazon Aurora database offers many advantages ranging from being cost-effective, combines speed and reliability but most importantly it offers open source and data sharing capabilities. Furthermore, it offers a more well-rounded structure, and consistency in data organization than the other plans. This database also gets automatically backed up regularly, and performs failure detection and repairs. In addition, Amazon always ensures to make security their main priority. However, this option is not free, but it offers many solution to many of the problems on this project. If this option is selected a good naming convention strategy is required, such as an ER diagram will need to be generated, while taking into consideration all the important metadata for the tables and columns relations (Amazon Aurora, 2018).

## **2.6 Selection and Preservation**

Professor Pinkerton has been collecting data for the past 10 years, this dedication proves that preservation is essential part of the plan. Thus, this data has a long term preservation time period and needs to be preserved for at least another 10 years. A good recognized stable long-term repository should be considered such as DalSpace. DalSpace is an institutional repository that collects, preserves and distributes digital content produced by members of Dalhousie University. It is suitable for documentation related to research datasets, such as journal articles. It can also be used to preserve final datasets (those that will not require changes or editing). It is also another appropriate means for sharing these saved data online (DalLibraries, 2018).

Another good stable repository to consider is one that offers curated services such as Dryas Data Repository. This is a good service, because the owner will be able to pass on their data into the hands of a professional data curator, where their main job is to continue to preserve the previous owner's data (Dryad, n.d.). The data is selected based on professor Pinkerton's interest and all the data should be retained.

Therefore, once professor Pinkerton decides it's time to pass on her data she can choose one of the above options, but until then the data will be continue to be preserved by both professor Pinkerton and her chosen colleagues.

## **2.7 Data Sharing**

The data is categorized into two categories:

- **Open Data:** This will be shared with all the users.
- **Closed Data:** This will be shared upon request with colleagues only.

Access control can be done using any of the plans offered previously. If DataVerse is selected, the owner will have complete control in deciding which data is private, and which is publically shared. This will help to eliminate answering and keeping track of all data sharing requests received through email (DalLibGuides, 2018).

On the other hand, OneDrive allows data sharing with explicit granted permissions to desired team members. There is a "Share" button for every user, that when it is clicked a share request is sent, and the shared data will be seen in the "Shared with me" section (Dalhousie, n.d.). As for the AWS plan, it offers a complete access control for the owner. In addition, the owner's public data sets can be shared via sharing programs such as Amazon EC2, Amazon Athena, AWS

Lambda and Amazon EMR. Users can download either one of these programs and use the search interface to locate their desired data they are searching for (Amazon, 2018).

## **2.8 Responsibilities and Resources**

Our team is responsible for transferring the data. Then, after the completion of this project the responsibility is delegated to the researcher team (postdoctoral).

### **2.8.1 Budgeting**

#### **Hardware:**

This project needs a hard drive and a laptop. The cost for hardware for data management plan is covered by the researcher funding.

#### **Human Resources :**

As stated in the case, Neil Gaiman is a colleague of professor Pinkerton, and he is helping in creating the file structure for this project. This project needs an IT consultant as well, to connect the database to the cloud server, and help in moving the data properly.

#### **Cloud :**

Dalhousie OneDrive, and DataVerse are institutionally funded, but as it might not support structured data at times, for which Amazon Web Services is recommended. The cost for AWS varies with slight change of the size of data. There is no upfront commitment with Amazon Aurora; you simply pay an hourly charge for each instance that you launch ("Amazon Aurora Pricing – Amazon Web Services", 2018):

<b>Storage Rate</b>	<b>Cost in Canadian Dollars</b>
<b>First 50 TB / Month</b>	\$0.025 per GB
<b>Transfer Rate</b>	<b>Cost in Canadian Dollars</b>
<b>1GB per month</b>	Free for one year
<b>Next 9.99TB per month</b>	\$0.09 per GB

*Table 3: Amazon Aurora Pricing ("Amazon Aurora Pricing – Amazon Web Services", 2018).*

## **3 CASE - 4**

### **3.1 Admin Details**

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

**Principal Investigator / Researcher:** Professor Chartreuse

**Consultant Company:** Data Masters

### **3.2 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, by pass the Excel sheet format and directly insert (query) the JSON files format into a SQL relational database (Contributors, 2018). By doing this professor Chartreuse can keep querying his own inputs as he like with less confusion, and the ability to share data. Furthermore, JSON file format will allow for the data to be reused, inserted into the database, easily shared with access control in the SQL Server, and preserved for a long time in the SQL relational database Server Cloud.

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

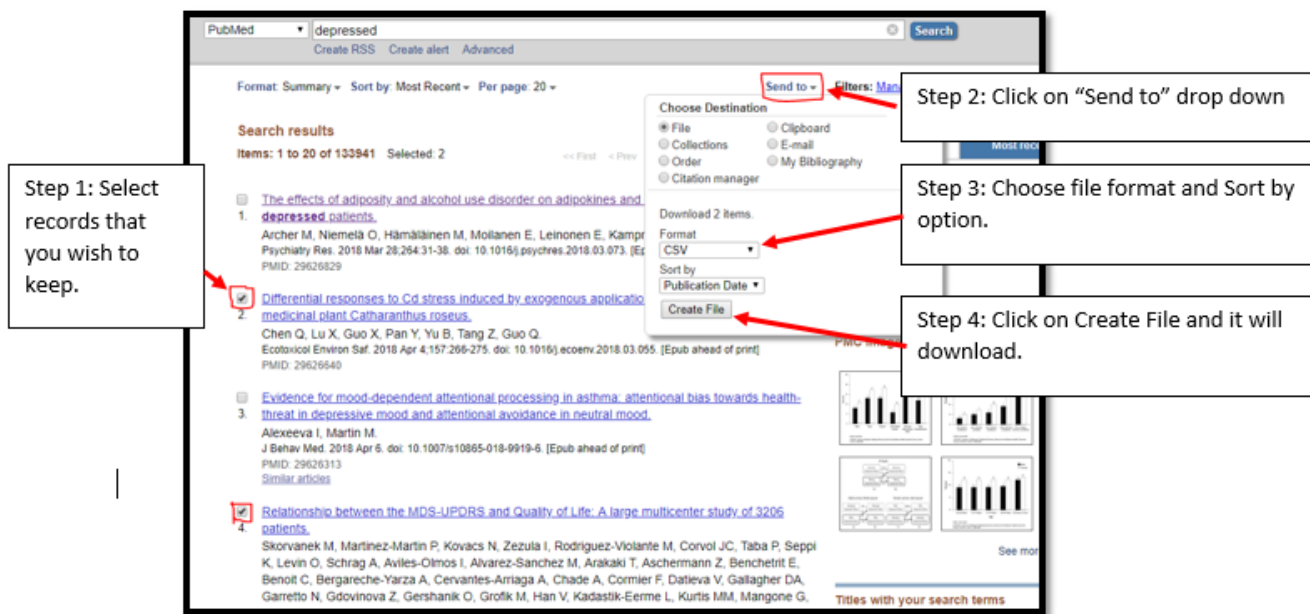
Another option is using the University cloud OneDrive. It does not offer the same advantage of the structured organization of the SQL database, but it is free. The institutional (university) OneDrive cloud offers storage, collaborative work through Office Documents, access space via the web, personal storage with syncing, data 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 (Dalhousie, n.d.).

#### **3.2.1 Public Data Collection Optimization**

There are two ways to optimize collection of public data, especially in PubMed, instead of inserting the data manually into Excel. The first method, is to learn the process of Web API public

data downloading in JSON format, then to save the JSON files, back them up and input them directly into Azure SQL Database via querying. This is a more technical way and will take our team a while to implement, but it offers a more efficient way of handling the data using a database, and this will lead to a more consistent and structured organization.

The second way is much simpler; in PubMed, after keyword search, a list of results appears. Next to the articles, a checkbox is located, and the user can select which articles' index information to download, then the user should go to the 'Send to' drop down menu, 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 simpler than doing it by hand; however, this might still lead to a messy excel sheet if CSV file is chosen as an output. If XML format is chosen, it will be easy to query it into the database just like the JSON file format.



**Figure 2:** Visual Process of simpler method of public data collecting in PubMed (NCBI, n.d.).

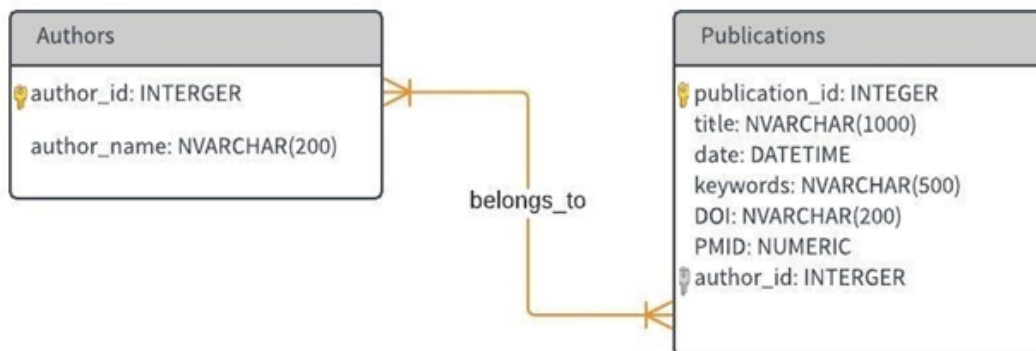
### **3.2.2 ER modeling Naming Convention**

In order to structure, name the files, and to better understand how the data needs to be organized, an Entity Relationship modeling schema naming convention and procedure should be used. This model helps to give a visual layout on how to organize and label the Metadata in the database. The Metadata labels used are date, name, keywords, publication, authors, DOI (digital object identifiers) and PMID (unique document identifier).

### **3.2.3 Version Control**

As this project takes into consideration future collaborative work opportunities with other colleagues, version control is needed. However, in databases it is different; for SQL databases it is called SQL source control. SQL source control is a software, where it is installed and connected with the user's database (Dave, 2016). After installation, the source control keeps track of the database, the transient - which is a local copy of the 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 (Redgate, 2016). Then a three way comparison is performed and a list of changes is then generated. Flyway and Visual Studio database project are free recommended SQL source version control that is compatible with Azure. There are also many commercial source controls such as, Red Gate, ApexSQL source control, and DB Ghost Change Manager (Dbmstools, n.d.). Depending on professor Chartreuse's needs our IP's will install the one that fits them perfectly.

### **3.3 Documentation and Metadata**



*Figure 3: ER Diagram Schema with Primary and Foreign Key restraints.*

### **3.4 Ethics and Legal Compliance**

The most important records are derived from PubMed. Hence, PubMed licensing agreement was reviewed to make sure no policy or laws are broken. This data management plan will follow these summarized copyrights and policy and guidelines:



**Copyright and Restrictions (PubMedHealth, n.d.):**

- There are both copyrighted, 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, PMC Open Access Subset can be checked.
- Materials produced by U.S. government employees as part of their official duties in not copyrighted within the US.
- 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 (NCBI, n.d.):**

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

**Endorsement (NCBI, n.d.):**

- No commercial use.

**3.5 Storage and Backup**

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

For the research data, which is in JSON format, it should be stored and backed up before it is transported into the database, and multiple backups is needed for this type of project. First the JSON files should continue to be backed up on professor Chartruese's external hard drive. However, this is not secure enough against file loss prevention. Therefore, another appropriate cloud storage services is required; Open Science Framework (OFS) cloud service seems to be a good choice, as it is free, supports JSON objects, can easily archive data, offers controlled access and easy collaboration, simplify the workflow, offers version control, and offers friendly user interface for project management (OFS, n.d.). OFS also offers openness to view the work of fellow researchers. The user can also choose which data to share and which to maintain private. Another

option to consider is the University's cloud servers. It offers more storage than OFS, but it might not offer the same openness as OFS.

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

Both database and research data should be backed up on a weekly basis by professor Chartreuse. Both OFS/OneDrive and Azure offers sufficient storage for that. In Azure, the first 12 months offers a free package size of 250GB for the database. As for OFS, its 5Gb on initial registration. Then, much larger size can be obtained by emailing their support team. If this cannot be done, there is always OneDrive, which offers 1TB of space.

### **3.6 Selection and Preservation**

Professor Chartreuse is very dedicated to his research and plans to continue his work until the day he is no longer here. Accordingly, this database and research data has a long-term value of a minimum of ten years and needs to be preserved well. Professor 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 curation services, such as institutional repositories or Dryad Data Repository (Dryad, n.d.). The data that must be retained are those that leads to publication, are selected based on the records of interest to the professor.

### **3.7 Data Sharing**

As the data is collected from public sources such as PubMed, its policy license must be upheld. In this case, the main policy about data sharing, is that it may not be used for commercial purposes. The 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 professor Chartreuse. The professor decides about to whom the access is given. Data sharing will be possible through both Azure SQL Database Cloud Server, OSF cloud server and their search engines. If colleagues wishes to collaborate with professor 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 colleague in Azure, with his

chosen set of restriction parameters, and then the colleagues will be prompted to choose their usernames and passwords.

### **3.8 Responsibilities and Resources**

Our consultant team is responsible for the Web API method of downloading JSON format files, in order to input directly these JSON files into to SQL database by querying the input. Then this information will be appointed to professor Chartreuse. In addition, all the copyrights and licensing policy will be followed while working on this project.

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

#### **3.8.1 Budget Plan Proposal for Azure SQL database**

Single Database Model (Standard Plan per month):

List	Description	Prices	Discounts
DTU's	Included storage 250GB	\$21.48	12 month free access trial
Backup Storage fee per GB	Long-term retention storage	\$0.0608	free until June 30th 2018
Threat Detection (Optional)	Security Center Standard Tier	\$18.24	60 days free trial
<b>Tentative Total</b>		<b>\$39.78</b>	

*Table 4: Monthly Budget Plan (Microsoft, n.d.).*

Azure SQL Database offers either a standalone database or elastic database pool, and these are priced in 3 tiers; Basic, Standard and Premium. DTU (Database Transaction Units) is defined as a measure of CPU, memory, data I/O, and transaction log I/O in ratio. This standard plan is recommended by our team, which is the go-to option for cloud application. It includes support for concurrent queries, collaboration and web application options. Lastly, Azure SQL database offers advantages such as bulk data migration, migrate SQL-based applications, simple user interface, security, data visualization, controlled access and collaborative work, big data handling, supports JSON and XML objects, Cloud server, storage and backup for file loss prevention. Many of these advantages are not all found in one free database, and with the provided discounts and free trials, it is the best chance to try this plan out now while it's still free.

## References

- Amazon Aurora – Relational Database Built for the Cloud - AWS. (2018). Amazon Web Services, Inc.. Retrieved from <https://aws.amazon.com/rds/aurora/>
- Amazon Aurora Pricing – Amazon Web Services. (2018). Amazon Web Services, Inc.. Retrieved 10 April 2018, from <https://aws.amazon.com/rds/aurora/pricing/>
- Cloud File Storage with Microsoft OneDrive for Business*. (2018). *Dalhousie University*. Retrieved 10 April 2018, from <https://www.dal.ca/faculty/computerscience/current/onedrive.html>
- Contributors. (2018, February 19). Work with JSON data in SQL Server. Retrieved April 06, 2018, from <https://docs.microsoft.com/en-us/sql/relational-databases/json/json-data-sql-server>
- Dalhousie. (n.d.). Cloud File Storage with Microsoft OneDrive for Business. Retrieved April 06, 2018, from <https://www.dal.ca/faculty/computerscience/current/onedrive.html>
- DalLibGuides. (2018). Research Data Management: DataVerse. Retrieved April 06, 2018, from <http://dal.ca.libguides.com/rdm/daldataverse>
- DalLibraries. (2018). Research Repository (DalSpace). Retrieved April 06, 2018, from <https://libraries.dal.ca/research/digital-initiatives/digital-publishing/repository.html>
- Dave, P. (2016, September 04). Team Database Development and Version Control with SQL Source Control. Retrieved April 06, 2018, from <https://blog.sqlauthority.com/2016/04/20/team-database-development-version-control-sql-source-control/>
- Dbmstools. (n.d.). 4 Version Control tools for Azure SQL Database. Retrieved April 06, 2018, from <https://dbmstools.com/version-control-tools/azure-sql-database>
- Dryad. (n.d.). Dryad Digital Repository. Retrieved April 06, 2018, from <https://datadryad.org/>
- Amazon. (2018). Large Datasets Repository | Public Datasets on AWS. Retrieved April 06, 2018, from <https://aws.amazon.com/public-datasets/>
- Microsoft. (n.d.). Your vision. Your cloud. Retrieved April 06, 2018, from <https://azure.microsoft.com/en-ca/>

- NCBI. (n.d.). PMC Copyright Notice. Retrieved April 06, 2018, from <https://www.ncbi.nlm.nih.gov/pmc/about/copyright/#publisher-links>
- NCBI. (n.d.). Policies and Disclaimers - NCBI. Retrieved April 06, 2018, from <https://www.ncbi.nlm.nih.gov/home/about/policies/>
- OSF. (n.d.). Open Science Framework. Retrieved April 06, 2018, from <https://osf.io/>
- Pcloud.com. (2018). *Cloud Storage Pricing & Cost, Cloud Storage Plans* | *pCloud*. [online] Available at: <https://www.pcloud.com/cloud-storage-pricing-plans.html> [Accessed 9 Apr. 2018].
- PubMedHealth. (n.d.). Copyright - National Library of Medicine - PubMed Health. Retrieved April 06, 2018, from <https://www.ncbi.nlm.nih.gov/pubmedhealth/copyright/>
- Redgate. (2016, May 18). SQL Source Control 5. Retrieved April 06, 2018, from <https://documentation.red-gate.com/soc5/reference-information/how-sql-source-control-works-behind-the-scenes>
- Research Data MANTRA. (2018). Mantra.edina.ac.uk. Retrieved 9 April 2018, from <https://mantra.edina.ac.uk/>
- Sync. (2018). *Plans & pricing* | *Sync*. [online] Available at: <https://www.sync.com/pricing/> [Accessed 9 Apr. 2018].