Assignment 03 – Notes

* 3 to 4 pages maximum
* Any diagrams, any software
* Convince me why I should use your plan
* Ask questions within the discussion threads if you have any questions
* Show me that you understand the basics of what you’ve learned here
* You are able to use any answers you find in the discussion board
* **You need to decide if all of the information is important enough to keep after the project is over. It’s a matter of cost.**
* Do not forget to do include licensing in your data plan

**Professors 1, 2 , 3**

Amy

Cassandra

Jillian

Kristy

1. What data will be created or collected (type, size, format, etc.)’

•What licenses apply to the data

1. •What facilities and equipment will be required (hard disk space, backup server, central repository, off-site repository, etc.)

•What data management practices (backups, storage, access control, archiving etc.) will be used

1. •Who will own and have access to the data
2. •Which data will retain value after the life of the project

•What metadata and linked open data strategies will be employed

•How will its reuse be enabled and long term preservation ensured after the original research is completed

•How much will the storage of this data cost (cloud and/or hard drives

Lib guide – research data management

Class notes

* determine what is the best format to have the information converted to, then write into your data management plan that from this point on all information should be saved in that format.

# Data Management Plan Proposal Draft

**Project Name:** Professor Green

**Plan created by JACK Data Consultants**

**Principal Investigator of the Research**: Professor Green

### Data Collection

The project entails collecting data from healthcare organizations, open data, and conducted interviews. The types of data being collected, created and acquired are text, numeric, and audio data. The file formats currently collected are Microsoft Word .doc, .pdf, and .txt files. The audio formats are in the .mp3 file format, and open data collected in is .tsv file format.

**File naming conventions**

The recommendation is to implement a standard file naming system for all files. This will assist in organizing files, allow organization by date, and to maintain version control.

The naming convention may be based upon this format: QEII**\_**Staffing**\_**20140409**\_**v01.docx

* Use YYYYMMDD for the date.
* For each hospital or organization use a unique identifier, and be consistent when naming files.
* Keep track of versions using v01, v02, etc.

When using folder systems ensure folder hierarchies are as simple as possible. Select which organization method is preferred: the recommendation is to file by type of file (e.g., Interviews), then by hospital (e.g., IWK Hospital).

E.g.,  Healthcare\_Expeditures>QEII>datasets

**Documentation and Metadata**

Maintain full and complete documentation for the study to ensure that it can be interpreted correctly in the future. The following elements should be included:

* Research topic (e.g., teamwork in hospital environments, specifically interdisciplinary primary care teams)
* Description of method of data collection
* Definitions of variables (e.g., how the interviews were conducted/if they took place in-person or via video chat; do the companies engage in team-building exercises)
* Explanation of data coding practices
* Format and file type of data that is being collected and created (including changes to file formats)
* Details of contributors and responsibilities

Standards for data coding must be created and used throughout data collection and during the research project.  Determine short codes for the institutions (Hospitals, e.g., Halifax Infirmary = QEII). Ensure strict confidentiality practices are maintained when describing interview participants (e.g., used coded names for each participants and remove/change identifying features).

**Metadata standards**

Consistently capturing information will lead to consistency and completeness. During the research project use a template when conducting textual analysis from outside sources. JDC can assist with creating this template. This will ensure the consistent collection of important variables going forward with research conducted by the Principal Investigator or other collaborators.

JDC recommends using the the DDI metadata standard (<http://www.ddialliance.org/training/why-use-ddi>). This standard is machine-readable and interoperable, encourages comprehensive data description, and enables the reuse of metadata.

Some of the metadata elements include title, summary, subject terms, geographic coverage, time period, data of collection, data type and restrictions.

●      Title, Alternate Title

●      Study Number

●      Principal Investigator

●      Funding

●      Bibliographic Citation

●      Series Information

●      Summary

●      Subject Terms

●      Geographic Coverage

●      Time Period

●      Date of Collection

●      Unit of Observation

●      Universe

●      Data Type

●      Sampling

●      Weights

●      Mode of Collection

●      Response Rates

●      Extent of Processing

●      Restrictions

●      Version History

**Storage and Backup**

#### Storage Requirements

The current amount of data collected is 24 GB and is projected to triple over the next 10 years, to a total of approximately 72 GB of data. JDC therefore recommends obtaining 100 GB of total storage to allow for growth over time. This will allow data to be stored for an indefinite amount of time.

#### Data storage and backup plans

Standard data practice is to ensure that data is stored in three separate places in the event of natural disaster, file corruption, or human error. It is also important to store data in multiple formats. Thus, it is recommended that the research data be stored in 3 locations to avoid the loss of your important research.

1. **Secure cloud storage.** **Box**is a secure cloud storage service that is located in Canada. The Box Starter package (2018 version) would facilitate storage and access needs, and costs $7/user/month. Box is the best option for cost, security, and maintaining Canadian privacy. The plan has a minimum of three users, but allows the owner to provide granular access to specific persons; the Principal Investigator maintains complete control. This software provides both storage and backup features, and offers up to 100 GB of storage.

For a larger budget, the Box Business package (2018 version) has additional security features such as Advanced user and security reporting and monitoring. Given the sensitivity of interview data, this may suit the researcher better. The cost is $21/user/month and provides unlimited storage and the option to add more collaborators.

2. **Network Drive.** Another copy of the data will be kept on the Dalhousie Network Drive that Prof. Green has access to as a member of the Dalhousie faculty.

3.  **External Hard Drive.** The third copy of data will be stored on an external hard drive that can be kept in a secure off-site location. Although it is good practice to perform backups after every change made to data, JDC recommends creating a backup schedule and ensuring that files are backed up weekly, or monthly, depending on how often the data is updated. Data should be encrypted on the hard drive to ensure complete control over any data that contains personal or sensitive information.

#### Data access, modification, and team communication

Currently, the research team communicates through Zotero (for documents), Dropbox (for audio files), and Google Docs (for transcriptions). Box may be used in lieu of these three programs to streamline team communication. Box offers features such as large file sharing, real-time editing and automatic version control. The password protection feature will also allow Prof. Green to share only anonymized data with his graduate students, or other specified collaborators. Additionally, team members should refrain from discussing confidential matters or divulging identifying information in Box or any other means of digital communication, including email.

**Preservation**

#### Long-term preservation and access

#### **Upon completed of data collection, and at the end of the research project the final data should be shared with colleagues and other researchers. The data and research will have enduring value to future researchers. The information can be entered into the Dalhousie DataVerse and the Federated Research Data Repository (FRDR).  DataVerse is an open-source data repository where Dalhousie researchers can deposit and share data. The DataVerse will hold data files, documentation, and descriptive metadata. FRDR is a part of that platform to support the preservation and sharing of Canadian research data.**

#### **By entering data into the Dalhousie DataVerse repository and FRDR, the research can be spread to a wide audience of interested parties and researchers. The team also recommends publishing the  data, analysis, and findings to provide the widest audience.**

#### Preservation-friendly file formats

Ensure that preservation-friendly formats are used in long-term storage (.txt and .csv). To guarantee that information is anonymized, remove any identifying information from interview data. This includes direct identifiers, such as names, employment numbers, or place of employment; and indirect identifiers, such as gender, salary or age.

Include supporting documentation: describe study, method, metadata and backup processes.

### Sharing and Reuse

#### Format of data to be shared

The principal investigator may decide what form of data to share, and to base the level of data upon the mandates of the funding body, the CIHR. JDC recommends not sharing the raw data to preserve anonymity of participants; instead share data with identifying features removed or coded.

#### End-user licensing

For the data that is collected through the textual analysis and interviews, the recommendation is to  obtain the following licensing for the data that you release. This helps you assert your rights as the original creator and ensure your data is used in the way you intend.

License: **Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0)**

This license allows others to share and adapt your work, but they must give appropriate credit and indicate any changes that were made. The data may not be used commercially, and if anyone builds on the data, they must use the same license. The open data sets that are used could have varying licenses, which means that the end-user license will depend on the individual data set. It is recommended that the licensing of the sources from which the data is collected be identified and maintained.

### Responsibilities and Resources

#### Project Data Management Responsibility

Primarily, Dr. Green is responsible for carrying out the data management plan, and holds responsibility for any modifications to the plan. He also will manage the data after the conclusion of the research project.  As the project grows, it may be beneficial to designate a graduate student to maintain the plan. The student would be fully trained to learn the backup methods, metadata, etc.

In the unlikely event of a change in Principal Investigator, we recommend designating a co-investigator in your department (choose a trusted colleague) to obtain access.

#### Data Management Plan Resources and Cost

With your CIHR funding, the major costs are the Box storage -- the total for 3 users is $252/year, or $756/year for the Business package. The purchase of a new secure external hard drive will cost between $50 to $150, with the addition of encryption software to ensure security.

### Ethics and Legal Compliance

#### Security of sensitive data

Once interviews are conducted and the applicable data is extracted into the recommended template, the MP3 files should be destroyed as well as any documentation that identifies the individuals beyond the participants’ coded identifiers to ensure ongoing protection of the identity of the participants.  
  
Though it is recommended the raw data be destroyed as noted above, should Dr. Green anticipate the need to continue to use the raw data, or wish to keep the interview recordings with identifiers included, JDC recommends the purchase of an additional encrypted external hard drive for its storage. As it contains sensitive raw data, the device should only be accessible to Dr. Green and stored in a secure area. Further, it would be recommended that additional safeguards such as password protection be used.

#### Secondary uses of sensitive data

A consent form will provide participants with information about how data will be shared and note that their data will be anonymized.

**Legal, ethical, and intellectual property issues**

JDC’s recommendation is that the primary researcher adhere to and pay close attention to the licensing of the public domain data that is gathered, as it will vary according to source. This will help to ensure that data is being used, modified, and shared appropriately in accordance with the appropriate licensing.

**Project Name:** Professor Pinkerton

**Plan created by JACK Data Consultants**

**Principal Investigator of the Research**: Professor Pinkerton

### Data Collection

The project entails collecting data from internal and external sources and inputting it into Excel. The types and formats of data that will be collected will vary according to source. The types of data that will be created are numeric and text, and should be saved in XML or CSV format, which are open file formats that are widely used in the research community and willl enable efficient data sharing and long-term access.

(...text can be saved in these formats as a standard and then open in Excel)

**What conventions and procedures will you use to structure, name and version-control your files to help you and others better understand how your data are organized?**

JDC recommends that the project be structured by distinguishing files based upon their origin, (external versus internal) or by subject (student data, job description, etc.). Before proceeding with the DMP or cloud storage, it will be necessary to access the structure of the files and to discuss with Neil Gaiman the current file structure and naming convention being utilized. For optimal navigability Gaiman’s metadata standards should be adopted, or an agreed upon convention will be established and utilized by both parties.

Files should be named by following a standardized file format. Records will be renamed so that the subject is clear, allowing for ease of retrieval. When adding dates to file metadata, the standard YYYYMMDD format must be followed. If there is more than one copy of the same file or an updated data set exists, a careful assessment of whether it is necessary to keep the files should be undertaken. JDC recommends discarding any redundancies of files to assist in the reduction of file backlog and to eliminate the unnecessary use of hard drive space. If it remains necessary for the researcher to maintain multiple versions of the file, then clear indication of version number in a standardized format should be included in the filme name (e.g., V01, V02).

Assessing content, sorting, and renaming will be a lengthy process. The principle researcher may wish to consider reducing the number of files first before commiting to the renaming and uploading process.

### Documentation and Metadata

#### What documentation will be needed for the data to be read and interpreted correctly in the future?

Documentation for the internal data that has been collected solely by the Principal Investigator can be added to ‘Excel properties’, and the elements that should be documented in this way include the meaning of the data, how the data was created, the original type and format, and the method of transforming the data from one format to another. Documentation for the external data (from other researchers, various governments and private corporations) that is kept as local copies should also be added to ‘Excel properties’, and should include elements such as the source of retrieval, the original type and format, and the method of transforming it from one format to another.

#### 

#### How will you make sure that documentation is created or captured consistently throughout your project?

JDC recommends creating a template for the data collection process. This will ensure that important variables are described, there is metadata consistency for all records, and these elements are properly extracted from the documentation via the excel properties.

#### If you are using a metadata standard and/or tools to document and describe your data, please list here.

JDC recommends using a metadata standard for all of the data that is collected to ensure that it is navigable. At the very least, the following metadata elements should be included, whether the data sets are internally or externally sourced:

* Title
* Subject/description
* Author
* Source
* Category
* Comments

Additional metadata elements can be added as needed to make the data retrieval process as efficient as possible.

### Storage and Backup

#### What are the anticipated storage requirements for your project, in terms of storage space (in megabytes, gigabytes, terabytes, etc.) and the length of time you will be storing it?

Professor Pinkerton currently has approximately 60 GB of data and will continue to collect data sets at a high rate. A storage system that accommodates the 60 GB and is able to grow as needed will therefore be suitable for the project. The data will be able to be stored for as long as needed.

#### How and where will your data be stored and backed up during your research project?

Standard data practice is to ensure that data is stored in three separate places in the event of natural disaster, file corruption, or human error. It is also important to store data in multiple formats. Thus, it is recommended that the research data be stored in 3 locations to avoid the loss of your important research.

1. **Data.world.** Data.world is a collaborative cloud platform used for organizing and sharing data. Using data.world would eliminate the need for Professor Pinkerton to store data on her laptop, and would enable her to create data sets in the program itself instead of storing data in Excel files, should this appeal to Professor Pinkerton.
2. **Dalhousie OneDrive.** OneDrive would be a suitable second cloud location to store and backup the Excel files.
3. **External hard drive.** JDC recommends storing a third copy of the data on an external hard drive that is kept in a secure off-site location. Although it is good practice to perform backups after every change made to data, JDC recommends creating a backup schedule and ensuring that files are backed up weekly, or monthly, depending on how often the data is updated.

You can use the online storage site: data.world

For secondary backup and storage, you can use the University cloud option - OneDrive.

If you desire a physical backup, I recommend using a large format USB flash drive.

We recommend using a USB flash drive to save all information securely.

#### How will the research team and other collaborators access, modify, and contribute data throughout the project?

Data.world enables secure data sharing and collaboration, which will eliminate the need to email spreadsheets back and forth with colleagues. Data.world also allows members to invite others to view or contribute to their data and manage access controls by making data public or private. Professor Pinkerton currently wants colleagues to obtain permission to view data sets but not be able to edit them. Should she want to allow colleagues to contribute in the future, Professor Pinkerton will be able to change the access controls.

### Preservation

#### Where will you deposit your data for long-term preservation and access at the end of your research project?

With the recommended subscription to data.world, collected research and data sets will be protected and maintained through the software’s built in monitoring practices. By utilizing Amazon web services and Google Cloud storage to house data, multiple safeguards and encryption procedures protect customer data on data.world. Additionally, a variety of granular access controls allow the primary data contributor to determine who may view which data. JCD recommends granting the postdoctoral fellow full access, while allowing varying degrees of access to others who wish to view your acquired data sets.

Any research that is designated by the primary data contributor as open will be available online through data.world for viewing, collaboration, and sharing. Data.world adheres to the findable, accessible, interoperable and reusable principles of research data, but offers the subscriber secure sharing features. The primary subscriber is in control and may grant access as deemed appropriate.

#### Indicate how you will ensure your data is preservation ready. Consider preservation-friendly file formats, ensuring file integrity, anonymization and de-identification, inclusion of supporting documentation.

A standardized file naming systems will facilitate easier navigability among files, and the present consistency of .csv and .xml file formats should be maintained. This will facilitate the easiest retrieval of data across all recommended platforms (Data.world, OneDrive,  and the external hard drive). Further, Data.world offers a document keyword or tagging system to create more robust metadata descriptions, and allows for fellow researchers to find data by subject or topic.

As 95% of Professor Pinkerton’s data is open source, issues of anonymity are not a concern. However, if the student files or entry level job descriptions are kept as data sets, then a coding system should be employed to protect any names or designators of identity within the research data. Once coding occurs, JDC recommends destroying the original raw data which contains any identifiers.

### Sharing and Reuse

#### What data will you be sharing and in what form? (e.g. raw, processed, analyzed, final).

We recommend sharing the raw data that is open data through data.world.

#### Have you considered what type of end-user license to include with your data?

JDC advises licensing data that has been created and is going to be shared and distributed. JDC recommendsusing the *Attribution-NonCommercial-ShareAlike 4.0 International* license, which allows others to share and adapt the data, provided they give appropriate credit and  indicate any changes that were made. Additionally, this license prevents the data from being used commercially, and if anyone builds on the data, they must use the same license.

#### What steps will be taken to help the research community know that your data exists?

JDC recommends using data.world to share the data.

### Responsibilities and Resources

#### Identify who will be responsible for managing this project's data during and after the project and the major data management tasks for which they will be responsible.

The client, Professor Pinkerton, and the postdoctoral fellow, Neil Gaiman, will be responsible for managing the data and following the structure of the data management plan outlined by JCD. We recommend that Gaiman’s initial task involve renaming and uploading files to the Data.world repository, and that this online software remain the primary cloud storage system as data collection proceeds.

Data.world has step-by-step tutorials to assist with data uploading, and a user-friendly interface. JDC will guide the client through the initial phases of the data uploading process to ensure ease of software use, and that best practices are established.

#### How will responsibilities for managing data activities be handled if substantive changes happen in the personnel overseeing the project's data, including a change of Principal Investigator?

To ensure continued access to valuable research files and data sets, JCD recommends that the Principal Investigator allow the postdoctoral fellow or another trusted faculty member, full access to the Data.world online repository as well as the secured external hard drive. This will provide an additional safeguard, and ensure that the data remains available to fellow researchers in the unlikely event of a change in project personnel.

#### What resources will you require to implement your data management plan? What do you estimate the overall cost for data management to be?

The Data.world package JCD recommends will cost $50 USD/month. This would be $600 USD annually, which would be between $750-800 CDN per year at current (April 2018) conversion rates. An external hard drive would range between $25 and $100 CDN.

### Ethics and Legal Compliance

#### If your research project includes sensitive data, how will you ensure that it is securely managed and accessible only to approved members of the project?

Should the Principal Investigator decide to retain student files or entry level job descriptions as data sets, then a coding system should be employed to protect any names or designators of identity within the research data. Once coding occurs, JDC recommends destroying the original raw data which contains any identifiers.

#### If applicable, what strategies will you undertake to address secondary uses of sensitive data?

Once the student and job description data is coded and all identifiers have been removed, there should be no concern with sharing the data sets via the online Data.world repository for secondary uses, should the Principal Investigator wish to share this information.

#### How will you manage legal, ethical, and intellectual property issues?

JCD recommends undertaking stringent citation practices to ensure that the intellectual property of the creators of the data is respected, and to protect those who seek these data sets from mistakenly attributing the wrong person to the work. Following JDC’s recommended use of a template to create metadata standards will protect the Principal Investigator and ensure that consistent records are kept regarding the source of each dataset.

**Project Name:** Professor Chartreuse

**Plan created by JACK Data Consultants**

**Principal Investigator of the Research**: Professor Chartreuse

**Data Collection**

The project will collect data from healthcare organizations, open data, and conducted interviews. The data being collected, acquired, and created are text and numeric. The data should be saved in .xml or .csv format for data being collected in Excel. Textual data should be saved in .txt format. These formats will allow for the data to be easily shared, accessed, and re-used by the researcher and anyone else given access to the data.

**File Organization, Naming, and Structuring**

JDC has created an Entity Relationship Diagram (ER Diagram) that provides a visualization of how the data should be organized. This diagram is located at the end of this document.

Files should be named according to a standard naming system to allow for enhanced document searching and retrieval. You had mentioned that you find it difficult to determine what files are which when you look back upon your work, By utilizing the design suggested in the ER diagram, issues around which type of file contains which type of data should be avoided. This design will allow for search by keyword, date published, date retrieved, author, format, etc. This system will elevate the organizational structure of the data and allow for better retrieval of the desired files.

**Documentation and Metadata**

To ensure that the data can be properly interpreted in the future, the following elements should be included in the documentation for all materials collected and acquired:

* Record titles, DOI, volume, edition, page numbers
* Type of licensing on the record
* Date retrieved
* Record date
* Authors
* Other sources which use this record
* Keywords
  + Note: It is important to maintain consistency with the keywords chosen; i.e., if the keyword “doctors” is assigned to a record, continue to use “doctor” for similar records and do not interchange with “physician” as this will assist with future searches within the data)

**How will you make sure that documentation is created or captured consistently throughout your project?**

JDC recommends using the the DDI metadata standard (<http://www.ddialliance.org/training/why-use-ddi>) for social science research such as that being conducted here. This standard is machine-readable and interoperable, encourages comprehensive data description, and enables the reuse of metadata. Some elements of the standard are listed below:

* Title, Alternate Title
* Study Number
* Principal Investigator
* Subject Terms
* Geographic Coverage
* Time Period
* Date of Collection
* Data Type
* Mode of Collection
* Extent of Processing
* Restrictions
* Version History

Metadata collection should remain consistent. The metadata elements being used could be placed into a template to allow for sharing with other researchers or colleagues. This will ensure that any data collected by others will follow the same labelling and metadata practices and allow for easy organization and of the valuable information being collected and shared.

**Storage and Backup**

**What are the anticipated storage requirements for your project, in terms of storage space (in megabytes, gigabytes, terabytes, etc.) and the length of time you will be storing it?**

Professor Chartreuse currently has 20 GB of data that was collected over approximately 4 years. If Professor Chartreuse’s data collection continues at this same rate, the total amount of data could increase to 70 GB within the next 10 years. Additionally, given that the Professor is looking to expand his personal research, a storage system with upwards of 70 GB would be most suitable. Professor Chartreuse will also want to store his data for the length of his entire lifetime.

**How and where will your data be stored and backed up during your research project?**

Standard data practice is to ensure that data is stored in three separate places in the event of natural disaster, file corruption, or human error. It is also important to store data in multiple formats. Thus, it is recommended that the research data be stored in 3 locations to avoid the loss of your important research.

**Network Drive.** Another copy of the data will be kept on the Dalhousie Network Drive that Professor Chartreuse has access to as a member of the Dalhousie faculty.

We recommend using Box, the Dalhousie Network Drive, and an external hard drive for storing and backing up data. The Box Personal Pro plan is a suitable choice for individual researchers who would like to share their data with other researchers. The Personal Pro plan is $14 per month, and offers 100 GB of storage, which accommodates the Professor’s storage requirements. The backup and version control features of Box will ensure that data is consistent and protected. We would also recommend

Further, we would advise keeping the external hard drive offsite, as another layer of data protection.

A backup schedule is recommended. Professor Chartreuse may wish to get into the routine before he leaves his office every Friday, or following a substantial addition to his data set, to upload the updates onto his online research depository, Box, and copy the new data onto the external hard drive that you have in your office. This will take about ten minutes and will greatly reduce any retrieval measures in the unlikely event of lost data.

**How will the research team and other collaborators access, modify, and contribute data throughout the project?**

Professor Chartreuse would like collaborators to be able to access files with permission but not be able to modify them. This will be facilitated through Box, which will enable  Professor Chartreuse to manage access and sharing controls, and invite collaborators to view, access and search his account for files. Alternatively, Professor Chartreuse will be able to share files of any size through individual links.

**Preservation**

**Where will you deposit your data for long-term preservation and access at the end of your research project?**

The Federated Research Data Repository (FRDR) will enable the data to be shared with researchers from other institutions throughout Canada. Plus, preservation and archiving is done automatically in FRDR.

<https://portagenetwork.ca/wp-content/uploads/2017/03/RDM-FRDRY1_Project_Update_January2017.pdf>   (A good description of the benefits of a FRDR - which speaks to a lot of Chart.’s concerns - seems like this is the direction Elvira wants us to go in?)

https://www.frdr.ca/docs/en/about/

We recommend a Canadian research data repository such as FRDR which is currently in their BETA version. It is a potential repository solution for researchers who do not currently have a practice in place.

Additionally, you are able to upload your files and maintain your metadata system from the dataset which can be created using the ER diagram we have provided you with. FRDR has made the process quite easy and there is no need to re-enter your metadata into the web form.

You can contribute up to 300GB of data which should be enough to accommodate your present needs and any future acquired records and data. Through FRDR your important research will also be stored longterm with Archivematica (version 1.6.1), a stable and reliable web archival software.

**Indicate how you will ensure your data is preservation ready. Consider preservation-friendly file formats, ensuring file integrity, anonymization and de- identification, inclusion of supporting documentation.**

Excel files should be saved as XML or .csv, and the documentation should be saved as a .txt file

**Sharing and Reuse**

**What data will you be sharing and in what form? (e.g. raw, processed, analyzed, final).**

Processed and analyzed data will be shared in Excel spreadsheets. Documentation will be shared in its final form.

**Have you considered what type of end-user license to include with your data?**

Since data collection will be conducted using primarily public domain resources, the type of end-user license included will vary based on the specific data. It is recommended that the licensing of the sources from which data is collected be identified and maintained. Additionally, if it is not already specified within a license, it is recommended that you make the data available for non-commercial use only (e.g., a *non-commercial* condition can be added to Creative Commons licenses).

**What steps will be taken to help the research community know that your data exists?**

Distribution on FRDR will help to ensure that the research community is aware of the data. The research may also be shared to a personal or institutional project website.

**Responsibilities and Resources**

**Identify who will be responsible for managing this project's data during and after the project and the major data management tasks for which they will be responsible.**

Professor Chartreuse, the Principal Investigator, will be responsible for managing the project's data during and after the project.

**How will responsibilities for managing data activities be handled if substantive changes happen in the personnel overseeing the project's data, including a change of Principal Investigator?**

You may wish to ask a colleague at your institution to be a co-investigator of your personal research data.

**What resources will you require to implement your data management plan? What do you estimate the overall cost for data management to be?**

Question not answered.

**Ethics and Legal Compliance**

**If your research project includes sensitive data, how will you ensure that it is securely managed and accessible only to approved members of the project?**

Since the data being collected is being retrieved from public domain resources, it is unlikely that it will include sensitive or personal information.

As mentioned previously, the Box storage option suggested includes settings that would allow for data to be accessible only to specified parties by the researcher should any sensitive data be included in the project in the future.

**How will you manage legal, ethical, and intellectual property issues?**

We recommend that you adhere to and pay close attention to the licensing of the public domain data that you gather, as it will vary according to source. This will help to ensure that you are are using and sharing the data appropriately. We have built into your data collection plan to consistently make note of the licensing of the data you acquire so that this information will be easily accessible to you and you will be taking steps to protect the rights of the original creators of the data. Additionally, our ER diagram for your research will be structured in such a way that you will be able to search your records for the type of licensing you are looking for. If you would like to retrieve records that are public domain, you would simple search for those and you will be returned a list of applicable results.

