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.

# Professor Green

## DMP title

### Admin Details

**Project Name:** Professor Green

**Institution:** Portage

### Data Collection

#### What types of data will you collect, create, link to, acquire and/or record?

Text and audio.

#### What file formats will your data be collected in? Will these formats allow for data re-use, sharing and long-term access to the data?

Current file formats of text files: Word .doc, .pdf, and .txt.

Audio files: .mp3

#### 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?

Standard file naming system for all files.

Use YYYYMMDD

For each hospital or organization use a unique identifier, and be consistent when naming files.

Example:

FileNm**\_**Guidelines**\_**20140409**\_**v01.docx

Keep track of versions using v01, v02, etc.

Ensure folder hierarchies are as simple as possible.

### Documentation and Metadata

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

Two different methods of data collection for studying teamwork in high-stress environments:

1) textual/content analysis (methodology) of 383 individual documents describing

2) conducting interviews with key informants

Needs to define variables. E.g., face-to-face meetings versus Skype, do the companies engage in team-building exercises.

Data coding:

-Hospitals, e.g., Halifax Infirmary = QEII

-Interviews for confidentiality (identifiers)

Who has worked on the project: himself, the masters students

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

Use a consistent template to determine which components you wish to extract from your research, to facilitate the consistent collection of important variables going forward with research conducted by either yourself, or potentially graduate students.

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

 We suggest the metadata standard of DDI, http://www.ddialliance.org/training/why-use-ddi

It is ...

Some of the metadata elements include:

* 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

#### 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?

 RIght now you have 24 GB of data and are going to triple that over the next 10 years.

We expect that to grow to about 72 GB in 10 years. We recommend having 100 GB of total storage.

You can store your data for an indeterminate amount of time.

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

We recommend that you keep your data stored in 3 locations to avoid the loss of your important research.

We recommend using a secure cloud storage. Box is a secure cloud storage service that is located in Canada. The Box Starter package would facilitate your needs, at $7/user/month. Box is the best option for cost, security, and maintaining Canadian privacy. The plan has a minimum of three users, but you can provide granular access to your students, so they can access only what you want them to.

This provides storage and backup in one.

We also recommend using two other forms of backup.

One copy will be kept on your secure Box account, another on your Dalhousie Network Drive that you have access to as a member of the Dalhousie faculty, as well as a hard copy kept on an encrypted hard drive kept in a secure off site location. We recommend that you form a backup schedule in which you will backup your files once a week or month depending on how often you update your research.

The plan provides up to 100 GB of storage, and can grow if needed.

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

 Team members may communcate through email, but must not transfer files or discuss confidential matters in email. Team members may use Box to transfer and access files.

### Preservation

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

 After you have completed your data collection and analysis we recommend depsiting the information you wish to share with collegues and fellow researchers into the Dalhousie Dataverse.

#### 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.

Use preservation-friendly format in the dataverse. .txt .csv

Anonymize information from interviews:

A person's identity can be disclosed from:

* **Direct identifiers** such as names, employment number, place of employment, gender
* **Indirect identifiers** which, when linked with other available information, could identify someone, for example information on occupation, salary or age

Include supporting documentation: describe study, explain method, describe metadata

### Sharing and Reuse

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

You can decide what form of data you would to share. We recommend not sharing the raw data to preserve anonymity of participants.

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

 When data collection and analysis is complete, we recommend obtaining the following licensing for your data you release. This helps you assert your rights as the original creator and ensure your data is used in the way you intend.

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

Others may 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 your data, they must use the same license.

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

 Dataverse repository. We recommend publishing your data to provide widest audience.

### 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.

Primarily, Dr. Green is responsible for carrying out the data management plan. 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.

#### 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?

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.

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

With your CIHR funding, the major costs are the Box storage -- the total for 3 users is $252/year. We recommend purchasing a new, secure external hard drive - costs are $50 to $150, add in encryption software to ensure security.

### 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?

Once interviews are conducted and the applicable data is extracted into our recommended template, the MP3 files should be destroyed as well as any documentation that identifies the individuals beyond the participants coded identifier to ensure ongoing protection of the identity of the participants.

#### If applicable, what strategies will you undertake to address 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

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

Question not answered.

# Professor Pinkerton

## DMP title

### Admin Details

**Project Name:** Professor Pinkerton

**Institution:** Portage

### Data Collection

#### What types of data will you collect, create, link to, acquire and/or record?

Numeric and text

#### What file formats will your data be collected in? Will these formats allow for data re-use, sharing and long-term access to the data?

Excel files should be saved in XML format or CSV (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?

When you look at structuring your files, you should consider whether you have already provided your assistant, Neil Gaiman, with a suggested file structure. If you have, you should consider whether this structure is currently working for you.

In looking at structures, you may wish to consider whether it makes sense to sort files based on their origin (e.g., external versus internal) or subject (e.g., student data, job descriptions, etc.).

Files should be named following a standardized file format. You will want to rename your files so that the subject is clear, as this will allow for ease of retrieval. When you add dates to file names, follow the standard YYYYMMDD format. If you have more than one copy of the same or a similar data file, you should assess whether you need to keep all of these files. You may wish to discard redundancies or files in which data is outdated. If you have more than one version of a file that you need to keep, you should indicate version number in a standardized format (e.g., V01, V02).

### Documentation and Metadata

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

For her own data that she collects, she would explain how the data was created, its meaning, and it’s general context through documentation via excel properties

For the data that she collects from external sources (other researchers, various governments, private corporations) and maintains local copies of, we would recommend that she add metadata via excel properties (such as retrieval source of data)

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

She would have a template for her own collected data to ensure that important variables are described, and she would include these elements via the excel properties

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

We recommend that you employ a metadata standard for all of your data that you collect 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

-And custom elements as needed

### 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?

 Storage space needed is approximately 61 GB for the current amount of data that you have. A storage system that can grow will allow you to continue to collect data sets. You may store data as long as you wish.

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

You can use the online storage site: data.world

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

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

We recommend using a USB flashdrive to save all information securely.

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

You can share access with your postdoc fellow.

For the data sets you share on data.world, you can select that each data set be private or public.

### Preservation

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

Through your subscription to data.world your information will be protected and kept secure through the companies monitoring practices. You have a variety of granular access controls which will determine who can access what. You may wish to grant your postdoctoral fellow full access, whereas you can grant varrying degrees of access to others who wish to view your acquired data sets.

#### 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.

As stated above, you will need to create a standardized file naming systems to facilitate easier seeking of your files, and also continue to consistently save as .csv and .xml. This will help you to find your information across all recommened platforms (OneDrive, info.world, and if desired, your jump drive).  Although data.world does allow for a tagging system to increase your metadata description and may help when people are seeking out specific topics or keywords on your profile.

As 95% of your data is open source, issues of anonymity are not a concern. However if you wish to keep your student files or entry level job descriptions as data sets, then you may wish to code any names or designators of identity so that they cannot be identified by others seeking out your research.

### 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?

 For your data: you may want to consider licensing the data you have created that you are keeping. We recommend  [Attribution-NonCommercial-ShareAlike 4.0 International](http://creativecommons.org/licenses/by-nc-sa/4.0/)

Others may 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 your data, they must use the same license.

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

 We recommend using an online data set sharing site, and we have selected data.world for you to try.

### 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 Pinkerton and her postdoctoral fellow, Neil Gaiman, will be responsible for managing the data. Neil Gaiman has been asked to create a folder structure and instead could work on uploading files to the data repositories we have suggested.   
  
The repository we have suggested is user-friendly and their are tutorials available. Our team can also provide assistance and guidance.  
  
Assessing content, sorting, and renaming may be a process that would take a large amount of time. You may wish to consider reducing the number of files you choose to continute storing.

#### 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 the files you have shared, you may wish to give access to your postdoctoral fellow or to another faculty member at your institution.

#### 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 we would recommend costs $50 USD/month. This would be $600 USD annually, which would be between $750-800 CDN per year at current conversion rates. A USB 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?

 It isn't indicated that any of the data the professor has currently received has sensitive data, as it is open source and publicly available there is no need for approval.

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

Question not answered.

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

We recommend that you undertake stringent citation practices to ensure that you are respecting the intellectualy propertly of thoe creators of the data, and to protect yourself and those who seek your data sets from mistakenly attributing the wrong person to the work.