INFO 6540

Elvira Mitraka

JP Pineo

B00522848

Group assignment-case 2

# April 10th, 2018

## Case 2: Dr. Green Research Plan

### Admin Details

**Project Name:** Dr. Green Research Plan

**Principal Investigator / Researcher:** Dr. Green

**Institution:** Dalhousie University

### Data Collection

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

Our team will collect 383+ digital text documents (PDF, Word, plain text), tabular data (Excel), and 900 minutes of mp3 files. We will convert all text documents to plain text format.

#### 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?

Our data will be collected in plain text, CSV, and mp3 form. These formats are industry standard and therefore should allow for easier data re-use, sharing, and long-term access.

#### 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?

Our team recommends using GitrKraken for version-control.

### Documentation and Metadata

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

Our team recommends researchers to keep detailed notes regarding: reflexivity, methodology, methods, and explanations of coding procedures. Versioning software such as GitKracken is has a user friendly GUI while keeping track of any changes to the data and how it occurred. GoogleDrive will also record who has edited what and when. We recommend Dublin Core (DC) for the creation of metadata records due to its broad accessibility.

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

GitKraken and GoogleDrive both have automated documentation features to ensure consistent documentation of all researchers’ contributions. In addition, researchers will be expected to maintain detailed notes and to provide weekly progress reports to the team.

The researchers' contributions will be automatically documented by GitKraken and GoogleDrive. Otherwise, researchers will be expected to maintain detailed notes and to provide weekly progress reports to the team.

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

Metadata will be documented using Dublin Core (DC).

### 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?

Factoring in for file versioning, backups, and growth over time, our team anticipates requiring at least 200 GB of storage for at least 5 years.

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

Our team recommend creating three copies of the data, with one of the three as an external storage source. Dr. Green’s practice of carrying a backup on a USB key is good practice, however we recommend switching to an external hard drive located offsite. We recommend storing the data on both Dalhousie's networked drive and on a cloud storage server such as GoogleDrive ($10 per month for 1TB). GoogleDrive is especially useful as a platform for collaborative work so choosing Dalhousie’s networked drive as another backup is recommended.

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

GoogleDrive will allow researchers to collaborate, documents the changes made and who they were made by, and makes this information accessible to the researchers.

### Preservation

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

At the end of the research project, the data will be provided to the funding provider, CIHR (Canadian Institutes of Health Research), for long-term preservation and CIHR will choose who may access it.

#### 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.

Our team will ensure our data is preservation ready by converting the data format to non-proprietary and industry-standard formats: .txt, .mp3, and .csv. We will include all field notes and metadata records. We will also ensure that all interview data has been properly anonymized and codified to prevent identification of study participants. All format conversions will be documented.

### Sharing and Reuse

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

Due to privacy concerns, we will not share any raw data. However, we will share the processed data once it has been thoroughly analyzed ensuring it is anonymized and codified. Afterwards, we will freely share all analyzed and final data.

CIHR has the following requirements for sharing of data resulting from projects funded by them:

* "ensure that all research papers generated from CIHR funded projects are freely accessible through the Publisher's website or an online repository within 12 months of publication;
* deposit bioinformatics, atomic, and molecular coordinate data into the appropriate public database (e.g. gene sequences deposited in GenBank) immediately upon publication of research results;
* retain original data sets for a minimum of five years (or longer if other policies apply);
* and acknowledge CIHR support by quoting the funding reference number in journal publications". (CIHR, http://www.cihr-irsc.gc.ca/e/32005.html)

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

We recommend using an Open Data Commons license, as endorsed by the CIHR.

"CIHR believes that greater access to research publications and data will promote the ability of researchers in Canada and abroad to use and build on the knowledge needed to address significant health challenges. Open access enables authors to reach a much broader audience, which has the potential to increase the impact of their research. Only when research findings are widely available, enabling open scrutiny, will this evidence be translated into policies, technologies, health-related standards and practices, and new avenues of research that will benefit the health of Canadians and others. From a knowledge translation perspective, this policy will support our desire to expedite awareness of and facilitate the use of research findings by policy makers, health care administrators, clinicians, and the public, by greatly increasing ease of access to research". (CIHR, http://www.cihr-irsc.gc.ca/e/32005.html).

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

We recommend publishing in an open access journal, such as Sherpa/Romeo, to increase accessibility

### 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.

While the study is ongoing the data will be stored by the three previously mentioned repositories (DalSpace, Google Drive, and Dr. Green's external hard drive). Our research team of four people will be responsible for transferring, converting, anonymizing and codifying the data. Our team will also produce the accompanying metadata records for this process. This should be completed within one year; the staff will be trained in all programs and file formats involved.

#### 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?

Of the four researchers on this team our positions are as follows: one will be the principal investigator, and another will be designated co-investigator who will receive the same training and information as the principal investigator. This should ensure minimal disruption if there are any unforeseen personnel changes.

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

Resources required will include employing a staff of four people for approximately one year, $10 per month for GoogleDrive's 1TB plan, 1 T2B external hard drive. Our team estimates roughly $230,000 in total cost.

### 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?

All sensitive data will be anonymized and codified by researchers to protect the identities of our participants. The raw data will not be made publically accessible.

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

All participants in our research have been/will be required to fill out a standard consent form, which can be viewed here: <https://web.stanford.edu/group/ncpi/unspecified/student_assess_toolkit/pdf/sampleinformedconsent.pdf>