INFO 6540

Elvira Mitraka

JP Pineo

B00522848

Group assignment-case 1

April 10th, 2018

Case 1: Professor Periwinkle

References:

Question 6: 2018. Dublin Core Metadata Element Set, Version 1.1: Reference Description. Dublin Core Metadata Initiative. http://dublincore.org/documents/dces/

Question 17: 2018. Data Manager Salary (Canada). Pay Scale. https://www.payscale.com/research/CA/Job=Data\_Manager/Salary

# My plan (Portage Template)

### Data Collection

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

Raw spatial data, images, and text will be collected from: remotely-operated marine vehicles (ROMV), tags that are surgically implanted in captured and released animals, static sensor buoys that measure ocean conditions, and communication lines that passively listen for signals from animal tags. At a later date, data from the public will be connected as well.

#### 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?

Collected data will be converted to NetCDF formatted files.

#### 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?

For version control, our team suggests using GitKraken as it is considered especially user friendly. GitKraken’s user friendly qualities extends to organization efforts as its design helps users understand where specific forms of data are being stored. The content files will be labeled according to what kind of content is being used in them. For example, they will be labelled based upon specimen, location, date, and how and where it was collected from (e.g. 20180410\_seal\_thunderbay\_eartag)

### Documentation and Metadata

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

We recommend using Dublin Core metadata, because it gives preliminary and necessary metadata to ensure interoperability while being easy to read. To ensure no important parts are overlooked, the data management team will need information about where and when the data was collected and the significance of each piece of data. This will help the data management team create effective and consistent metadata records in order to keep the most necessary information and increase ease of access.

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

After each important section of information is put into the metadata, our team will incorporate a secondary review process to ensure accuracy and consistency. For consistency, our team will create a standard method of document creation. The data managers will go through each documentation of data and create a new document that is needed for each one before creating an official document. Our team will use GitKraken to take advantage of its version control capabilities. Our team will upload each document as we work on them so that the changes from each version can be assessed for consistency with our created method for metadata.

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

In part because it is machine readable, our team recommend using Dublin Core for a metadata standard.

Attached is a link to the different metadata elements that will be needed to create the metadata scheme for Dr. Periwinkle. http://dublincore.org/documents/dces/.

### 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?

We recommend moving Dr. Periwinkle’s data from Dropbox to another cloud storage platform with greater storage capabilities. This data storage needs to be capable of expanding by 500MB a day and other cloud platforms are capable of this level of storage scalability. We recommend staying with a cloud storage method because Dr. Periwinkle’s requires her data to be open to anyone who desires to use it. For all these reasons, we recommend Google Drive as an appropriate solution. Although it requires login credentials, which may hinder the open data goals of Dr. Periwinkle, our team recognizes that the added security of determining who has access to this data is a worthy pay-off.

In Google Drive there can be multiple files and folders shared within a master folder, this would make collaborating multiple people on the research team easier. It will also allow each researcher and any grad students to have their own folder within the main research folder. Anyone can add their own data into a shareable file and have as many folders as needed for organizing different parts of their study.

This format is appropriate considering the high volume of student-collected data contributing to Dr. Periwinkle’s collection. Because of the ever-changing nature of the relationship Dr. Periwinkle’s students have to her research goals, putting data in a cloud will allow any data that is collected to remain in the cloud even after students potentially leave. In addition, Dr. Periwinkle and her staff need consistent access to this storage system regardless of location and employment.

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

Google Drive automatically backs up and saves all work. For additional back up, external hard drives will be used and updated once a week to include any new data and information. Each eternal hard drive will be 2TB and will be used to store all collected data and information as a failsafe in case the cloud storage experiences data loss.

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

Collaborators can add or change aspects of their contributions in Google Drive. Google Drive automatically records who performed changes, what they changed and when, and makes this information available to everyone using it.

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

At the end of Dr. Periwinkle’s study, all data will be given to the Department of Fisheries and Oceans Canada (DFO) where it will be kept for long term storage. DFO will receive all raw and analyzed data. DFO will grant access to whoever needs this data through a simple request process.

#### 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.

For preservation purposes, all data will be converted into a text document, in case difficulty arises from technology not being able to read NetCDF format. These changes will be noted, and if there is any data lost in the conversion it will be noted, and the data that is missing will be noted as to what was lost. These changes will be noted because all files will be kept in NetCDF format in addition to a word document. By storing the data in two formats, research members can compare the two files for discrepancies that indicate data loss.

### Sharing and Reuse

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

Shared data includes all data forms (raw, processed, analyzed, and final) collected from ROMVs, tags that are surgically implanted in captured and released animals, static sensor buoys that measure ocean conditions, and communication lines that passively listen for signals from said animal tags.

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

Our team recommends a simple Open Data Commons licenses, specifically Attribution License (OCD-By). This license allows all data collected into Dr. Periwinkle’s database to be accessed by those who need it which is in line with her research goals.

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

Our team recommends using emails to institutes and programs with similar research projects and work in the same field of research.

### 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.

During the duration of the study Dr. Periwinkle’s data will be stored with our research team. A small group of three people will be assigned to look after the data by ensuring it is kept up to date with the latest findings from Dr. Periwinkle's team and carefully organized and saved in appropriate files. Working closely with Dr. Periwinkle and her team will allow the data management team to understand the data being collected, and in turn, this will help organize it. Each of the three team members of data managers will have their own specific job on the team; one person will collect the data from Dr. Periwinkle and her team of researchers, one will organize it and upload it onto the new Google Drive files and onto the external hard drives, and the third person will ensure the consistency and completeness of each file. After the study is complete all data will be collected and delivered to DFO. The data will be accessible to the public and other marine wildlife research teams

#### 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?

All data is added to Google Drive by Dr. Periwinkle’s team of graduate students and our team of data managers will add this data to an external hard drive. Adding Periwinkle’s data to an external hard drive is beneficial for when students inevitably leave Dr. Periwinkle’s team, and decide they want to take their data with them, it can be retrieved and re-uploaded to Google Drive.

#### 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 include a group of 3 data managers, 4 2TB external hard drives, and a means of storing the hard drives offsite. An estimation of the costs for implementing this five-year data management plan is $907, 250.00.

### Ethics and Legal Compliance

#### legal, ethical, and intellectual property issues?

Since Dr. Periwinkle has grad students collecting data for her, a contract should be created stating that the research and data they collect is the intellectual property of Dr. Periwinkle and they consent to allow others to use and store this collected data. This contract will ensure no legal discrepancies regarding intellectual property rights occur between Dr. Periwinkle and her student researchers.