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. Data from the public will be linked because it will come into play later.

#### 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 to be especially user friendly. GitKraken will also help others understand where my data is coming from and why it went where it went. I will name the files according to what kind of content is being used in them, and something that is relevant to the study.

### 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 lay out, because it gives preliminary and necessary metadata to ensure interoperability and is also easy to read. I will need the information about where and when the data was collected, the importance of each piece of data so that I know how to create the metadata record and so that it will not be over looked. This will help to keep the information that will be needed in the future will be easy to access and look over, so the researchers will not waste time.

#### 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. I will go through each piece of documentation and create what I need for each one before creating an official document. Our team will use GitKraken in this process 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.

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/. Dublin care is machine readable.

### 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. Perifwinkle’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 platforms are capable of this level of storage scalability. We recommend staying with a cloud storage method because of Dr. Periwinkle’s requirement to have the data open to anyone who desires to use it. For all these reasons, we recommend Google Drive. 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 it this data is a worthy pay-off. -**INSERT RESPONSE TO COMMENT HERE-** 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.

Even though Dr. Periwinkle is already using Dropbox as her place of storage, I would like to shift it over to a cloud storage space that is not as limited as Dropbox can be, most of the stage will be in megabytes, roughly 500MB a day.  Also, because Dr. Periwinkle wants to be able to have her data be open to anyone who wants access to it, I think that best cloud sharing option would be Google Drive, even though it does require login credentials for those trying to access it, I feel this would be better and more secure way because then she could have some control over who is accessing it. Also, the Google Drive, can have more than file shared within it, so it can have as many folders as needed and there can be one for each part of her research. This would also allow for incoming and outgoing students to share their data, and when they leave it is still there on the drive and there will always be access to it. Since this is an ongoing study, there needs to be an unlimited amount of time that Dr. Periwinkle and her team needs to have access to a storage system and Google Drive has this capability.

#### 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 updated once a week to include any new information.

#### 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 this occurred and make this information available to everyone using it.

As well, it shows who was the last one to update something and you can see what and where they updated/changed, and it automatically saves so data and other information will not be lost if something was to happen to the computer/laptop they were working on.

### Preservation

#### 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 grant access to whoever needs this data through a simple request process. DFO will receive all raw and analyzed data.

Dr. Periwinkle's data at the end of the study will be given to the Department of Fisheries and Oceans Canada, there the data will be kept for long term storage and be accessible to anyone that may want access to the data, they will just have to request it from them. They will be given the analyzed data as well as the raw data that has been collected.

#### 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.

Choosing NetCDF as the format will aid goals of preservation as it is useable in databases.

Most of the documents are in NetCDF format, which is fine for the data to be kept in because this format is useable in databases.

However, there all the data will be converted into a text document, and 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. That is also why the NetCDF format will be kept, in case something is lost in the shifting over to a text document.

### 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.

This way other researchers can look at the raw data and see where the data analyzed comes from. Also, that way other researchers can use it how they want to for their studies. They might be focused on a different aspect than Dr. Periwinkle.

We will be sharing the final data as well because it will help show where all the information went to and how it supports Dr. Periwinkle research and findings.

#### 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), Attribution for the data/databases. This license allows all data collected into Dr. Periwinkle’s database to be accessed by those who needs it.

#### 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. Perinkle’s data will be stored with our research team. A small group of three people will be assigned to look after the data and to make sure that is it kept up to date with what the findings of Dr. Periwinkle's team and ensure that it is organized and saved in appropriate files. By working closely with Dr. Periwinkle, this team will understand the data being collected which will help in organizing it. 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. Our team of data managers will add this data to an external hard drive so when any students leave Dr. Periwinkle’s projects, and want to take their data with them, it can be retrieved and re-uploaded to Google Drive.

Since Dr. Periwinkle has grad students working for her on this project, the data that they collect will be added into the Google Drive documents and the data that they find will be added to the external hard drive and will be sorted by the team of 3 data managers from our company. as so that when they leave their data is still with the study.

#### 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, 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.

The estimate for this plan I suspect will be about $907, 250.00 in a five-year time span. This includes the cost of the team of 3 data managers, external hard drives, and the cost of storing those hard drives off site.

### 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.