DMP Part 3: Standards and Documentation

Zach Butler

1. Design a plan for the data documentation of your research data. The metadata standards that we discussed in [Module 5: Data documentation](https://canvas.oregonstate.edu/courses/1871025/modules/2847853) may or may not be right for your data. It is possible that a metadata standard is not the best choice to document your data. If that is the case, state the fact and describe other strategies that you will use for data documentation. Make sure that you:

1. Describe strategies that you will use to keep the data organized (e.g. file names, structure).

I will keep my data organized by file names. I have six separate datasets and they are named the following: daily\_p\_d2H\_mean, daily\_p\_d2H\_mean, Stream\_d2H, Stream\_d18O, PrecipData, and StreamData. Each file is structured the same with Stations as the headers and date as the far-left column from 2013 to the present. These data files were made via python and R code that is uploaded on a GitHub repository, as well as the data files. The code for my data files is updated per different versions as well as readme’s with each one and the overall methods between each script and data file. The goal is to have one readme that will explain how to to use the code to generate associate data files. Each code has an explanation of it in the beginning to explain what it does.

1. Describe how you will keep track of the different versions of your data. If you are planning on using version control software, describe.

I am not worried about the data files versions. The only updates on the data files will be updating the time period to the closest to the present. This will only add more data to each file. I am keeping track of different versions of my code on our GitHub repository by naming it V1, V2, or V1\_B or V1\_C for example.

1. Describe the metadata that you will generate during the project. What kind of information will you record? This information should enable reuse (by you or by others). You don't need to describe exactly all the information that you will record, but explain the type of information that you will document and give some examples. For example, your documentation may include methodology, analysis, definition of variables and units, etc.

The metadata I will generate from this project will be related to the site data for each data file. Additionally, I will need to keep information regarding each of the six files I described in A. The “daily\_p” and “Stream” file pairs will be the same format. One is just d2H isotopes and the other is d18O.The metadata for the stations is something I already have, which is characteristics of the stations. This includes station id, location, latitude, longitude, elevation, mean annual precipitation, catchment area, etc. The metadata for each data files will include what each number in the file means. The data is simple so it will not be overly descriptive. The metadata for the code will just be a readme describing what order to use it to generate the data products.

1. Describe how you will capture this information (e.g. in a separate document, using specific software, etc) and at what point in the research process you will do that.

I am planning to keep track of this in a csv file, which I already have for station characteristics. I need to be doing this now for the other data files since I already have my complete dataset. I already have a readme for the code describing what it does and when to use it to generate associated data files.

2. Provide an example of one of the metadata files that will be created when you apply of the documentation strategy you have chosen. Give a short explanation about what this example is, and why it is a good illustration about how you will document your data. It does not need to be about your data or created by you, it can be an example already available, but it should be directly applicable to your data. For example, if you are going to use a data standard that already exists, provide a documentation file from a dataset that uses that standard. If the format of the documentation is not straightforward to use (e.g. a plain text is straightforward, a format that requires installation of non common and/or proprietary software is not straightforward), please document the example in some way that it will be easy to access and understand. For example, you could make a short video opening the file and describing it, or you could take screen shots. Another example: if you chose to create your own template, provide a draft of the template.

I will follow Hydroshare’s resources to create metadata files, which use DCMI. This is the link from Hydroshare, <https://help.hydroshare.org/creating-and-managing-resources/best-practices-for-describing-your-resource-with-metadata/>. Additionally, the data from NEON that I am using follow this metadata standard, <https://www.neonscience.org/node/4414>. I’ve uploaded an example readme from a past project using similar data that I will model for my research. I’ve also uploaded a metadata file for the station characteristics.