Megan Baker

99-519 B

Slayton

8 June 2021

Data Management Plan 1

Aside from the Box drive organizing the photos already, I will keep my data and information (such as this document) on my Google Drive in a folder specifically for this class. The documents will be labeled by assignment name or another appropriate name so I can easily identify it. The original owner/creator of the data is most likely multiple photographers hired by the school or a newspaper, or photos taken by students for personal use. The current owner of the data is CMU libraries or more specifically Katherine Barbera. The data was most likely originally collected by multiple photographers over the course of many days, maybe even years. For the CMU libraries, this data could have been collected through old newspapers, or printed photos saved from MMCC itself. Additionally, they could be personal photos collected from prior students and their families.

I think the most important metadata to include are the folders they are currently in. Further metadata for every photo could include the date the picture was entered in CMU archives, whether the photo is more black and white or more yellow or contains color (some include red), and if the picture contains people, buildings or handwriting/text. The data could then be broken down by folder and metadata could be used to describe just those sets of photos. For example, the athletics photos could be described as either active sport photos or a posed team photo. The building's photos are already separated into MMCC and Morewood. And, the research center photos can be separated into photos taken in a lab and photos not taken in a lab. I think it is important as final note to highlight that this is mainly qualitative data with one or two pieces of quantitative data.

Data Management Plan 2

After the project is finished, the data will live online in both the raw data format and the data we publish in the project. The raw data will be collected on a Google Spreadsheet which is where it will remain when the project is finished. Because we will most likely be using our school Google Drive, after three years (when we graduate) if we still want the data to be preserved we must transfer the data to a different Google Drive or download it to our individual computers. The published project will live on GitHub and OSF and remain available to the public.

To protect the data and avoid security risks, the data will be collected on our school Google Drive. This documentation will only be accessible to group members and supervisors during the data collection stage. When the data is published, it will be in the format of “view only” so no changes can be made without our group's consent. Another form of protection is downloading the file periodically in the event of a security issue. This would prevent total loss of work and we would be able to work from the most recently saved file.

To prevent losing the raw data file, each member should have the file on their own Google Drive as well as a downloaded copy of the file on their hard drive after each working day. This would help mitigate the risk of completely losing the data in the event one member's raw data file is lost or corrupted. Further mitigation strategies for preventing loss of data would be to have a back up to the hard drive of the computer. Planning ahead will help minimize risks to the data but they are bound to happen. With any unforeseen risk, it is important to stay calm and come up with a plan with your group that will help solve the problem.

Data Management Plan 3

To start, I would store my data in GitHub which I have explained above. However, GitHub is not a long term repository that I would want to store my data in. For long term storage, I would look to store my data in KiltHub. KiltHub is appealing because it is a repository for CMU affiliates (which we are). Furthermore, it is a long term storage repository (as opposed to GitHub) and has easy user accessibility which is important for storing and accessing the data in the future. The only restrictions on what data can be stored is research on human or animal subjects, and material containing personal or administrative information. However, these restrictions do not apply to our research so they can be disregarded for this project.

When documenting and storing the metadata about the images, the specific metadata schema, PREMIS, could be used. PREMIS allows for the preservation of digital metadata so that long term usability of the metadata is ensured for future use. This would be the right metadata schema because the photos because they are stored digitally and should be preserved. PREMIS uses a data dictionary with the following five entities: Intellectual Entities, Objects, Events, Rights, and Agents to document data, all of which could be applied to the photos.

Data management Plan 4

For my current visualizations, I used Google Sheets to analyze and visualize my data. Google Sheets is a more simplistic and beginner tool compared to most other analytic tools; however, I only needed simple analytics and visualizations for my data thus far. The advantages of using Google Sheets is that it is connected to my student account, I can share my visuals with my group members (without having to put it in GitHub) so that I can get immediate feedback on my analysis and visuals, and they could potentially use my analysis to create their own visualizations. In the future, if I feel that I need more complex analysis and visualization than Google Sheets can provide, I will use Tableau. While I am less familiar with Tableau, it is a great tool for analysis and is user friendly. Furthermore, it is licensed by the school so I have free access to it. Tableau also allows for many file upload types which is important because I also have visuals (photos) that I would potentially need to upload. Tableau will require me to upload to GitHub to obtain feedback from group members, but that is not a huge drawback as I am already uploading my documents to GitHub.