**Data Management Plan For:**

**Professor Green**

**Project Description and Scope**

Dr. Green is focusing his research on evaluating how teams perform in high-pressure situations. He has selected hospital workers in primary care units as a case study for his analysis. Essentially the professor wants to know whether the high-pressure environment in which these individuals operate in affects how teams function as a cohesive unit.

The professor’s primary methodology includes interviews and written analysis. While he has only two streams of data collection, the manner in which his information is stored varies greatly. The file scope for Dr. Green is rather diverse. This is proven by the use of many file types such as Word, PFD, Excel and MP3. Essentially file types will either contain text or audio. In terms of volume the real burden will come from the text files, such as Word, PDF and plain text, as these documents number in the hundreds. The audio files, which house the interviews, are relatively few in comparison.

Dr. Green has sole ownership of the data collected through his research. Although graduate students sometimes help him, the professor is wary of disclosing his data to others. For this project he is employing two graduate students who will need to have access to some parts of the data.

The primary objective of the data management plan should focus on Dr. Green’s requirements for organization and security. Organization is needed for the classification and sorting of his text files. Additionally security is paramount to his data needs as there are confidentiality requirements for his audio files. Lastly it is important to note the professor does not want his data to be open sourced.

**Documentation, Organization and Storage**

As mentioned previously Dr. Green’s data is spread across a variety of file formats. While the text related documents out number the audio files, it can be assumed that the latter will be a higher burden in terms of file size. A MP3 file containing an hour of content, depending on bit rate quality, can range anywhere from four megabits to one hundred and forty-four megabits (Audiomountains). In comparison an average PDF/Word file is under one megabit.

Even if all of Dr. Green’s audio files were high quality recordings, at average length of an hour, the professor would only need just over two gigabits to house the audio files. This combined with the three hundred and eighty-three text files, the professor’s total storage needs are around three gigabits or less.

The professor is currently using three different types of storage platforms for each of the different types files he deals with. Text and excel documents are presently stored using Zotero, audio files with Dropbox and Google Docs for transcripts. Best practices towards data management go against what the professor is currently doing. Instead Dr. Green should consider consolidating his files. The first task should be converting any Word documents into PDF files. Not only would this reduce the number of file types but it would also prevent other people from editing the document. Secondly the professor could benefit from the use of file containers such as ZIP files. These containers can store similar-in-content files, whereby drastically reducing the number of files the professor has to sort through. An example of this would be grouping together a MP3 file, its associated transcript and any other supporting documents pertaining to the interview into a single ZIP file. Lastly ZIP files would also reduce the file sizes of the professor’s MP3 files, whereby reducing his total storage needs.

The use of ZIP files will also have a secondary benefit for Dr. Green. By utilizing this simple categorization method, the professor will be making use of metadata. Given the ZIP files will house directly linked content, each file makes use of metadata in the sense that the files will be categorized. The use of ZIP files, and by extension metadata, will ensure that the professor can easily identify and reuse the data at a later date.

Theoretically using files containers will reduce the volume of files, however Dr. Green will still need to house these files somewhere that’s secure but also accessible. It will be recommended that Dr. Green stores his file containers on a single cloud platform, rather than three that he is currently using. Based on his small storage needs the professor should utilize cloud storage from Box. The individual free plan for Box includes ten gigabits of storage with an upload limit of two hundred and fifty megabits. It was estimated earlier that the professor had storage needs that was under five gigabits so having access to double that ensures that any future requirements can be met. Additionally the file upload size will not be a hindrance as his audio files, his largest files, will be under the upload threshold. Additionally these audio files will be even smaller once compressed and placed into container files.

Occasionally the professor has a need to share his data with his research team. Box would allow him to do this via shareable URLs to the professor’s platform. Box also includes version controlling, something that will be important to the professor as his team starts to sift through and work with the data. Security is also paramount to Dr. Green as he is the only one who can view sensitive interview data. Box would allow the professor to implement access control measures. This would ensure that his research team has access to the data but not to the files containing sensitive information.

Another feature the professor is looking for from his storage solution is accessibility. Presently he keeps a USB drive with him at all times as a fail-safe option. The use of cloud storage would ensure that the professor has access to his data anywhere he goes that has a secure Wi-Fi connection. Box also allows users to access their files via mobile devices as well. If the professor uses a smart phone he can have access to his files at any time regardless of a Wi-Fi connection. This would remove the need for carrying around a USB drive.

**Access, Sharing and Reuse**