Data Management Plan

Presented to Professor Pinkerton

**Executive Summary**:

Outlined in this report is a data management plan, which includes recommendations for database management, data storage and data sharing. It is understood that you have created and collected a significant amount of data.

Best practices and policies of government funding agencies call for data to be made available as *open data*, and follow the FAIR (findable, accessible, interoperable and reusable) guiding principles for data management. This is not only part of the requirement for government sourced funding but it also is a growing part of data management best practices. It meets the university’s policy on research data, is often required by peer-reviewed journals, and is shows accountability for public funds.

We have prepared this report - based on requirements outlined by the university librarian and input from the research project manager, Dr. Pinkerton - to be thorough and of benefit to the research team.

**1. Project Description**

While most of Professor Pinkerton’s research is related to Management Information Systems, we understand she has many interests and an insatiable appetite for data. Pinkerton currently has a significant amount of Excel spreadsheets saved to her computer. She has a total of 17, 384 spreadsheets. The median size of the spreadsheets is 1,000 rows of data by 10 columns. While this seems like a lot of data, it needs only 1.16 GB of storage. Pinkerton gathers data from literature reviews, job descriptions, student performance data, and several files from other professors, as well as open data portals. Pinkerton is a resource to her peers and is often asked to share files, she her too busy to reply to all of the data requests she receives. Pinkerton is willing to share her collection of data, and is waiting on a postdoctoral fellow to finish with the folder structure in the cloud.

**2. Documentation, Organization, and Storage**

Professor Pinkerton only has roughly 1.16GB of data worth of Excel files (XML), and very little data in other formats. While she is constantly adding to her data set, all of her data eventually ends up in excel (XML), usually in numerical values, which doesn’t need much storage. Most cloud options provide 5-10GB for free, which should cover her storage needs for a long time if she adds to her collection at a constant rate. Once she exceeds this limit, additional storage is extremely cheap. In the meantime, I suggest using the university cloud as it provides more than enough storage space as well as allows you to share your data with whomever you wish.

While the size of the collection is not large enough to warrant purchasing cloud services, the documentation and organization of Pinkertons collection is very important. In order to maximize the societal benefit of Pinkerton’s collection of data, it is important that it is organized in a way that she and her fellow academics can quickly and easily locate the data set they need. For this reason, we recommend creating Metadata. Metadata is essentially data about other data. There are three types of metadata that exist, descriptive, structural and administrative, and we suggest you use all three in order to organize your data efficiently.

Descriptive metadata helps discover and identify the data, this can include the title, author, and any keywords that can help readers locate the correct file. Structural metadata describes, as the title suggests, the structure of the data including the types, versions, and other important information. Finally, administrative metadata provides information such as how, why, when, and where the data was generated as well as the file type and who can access the file.

We also suggest organizing the data into two seperate folders within the cloud.

* Two seperate databases
  + One on the cloud that is open and accessible for the majority of Pinkerton’s data
  + One on a hard drive for proprietary data

**3. Access, Sharing and Re-use**

Professor Pinkerton has full discretion over the potential availability of the majority of her collection as 95% of the data is from external sources and therefore not subject to funding agency requirements.

* Open Access to data
  + Accessible to everyone to read and be shared
  + Academic research can have a greater impact on the world
  + Communication of results important to research process
  + Maximize value of investment in research
  + No funding institution
  + Cost saving and sharing expertise
  + wants control over who she shares it with, so be able to share link

Should save data in an open format whenever possible

* Maybe grade data, and private corporation data in proprietary
  + Make sure data is available to you later

**4. Archiving**

**5. Conclusions and Recommendations**

**What data will be created or collected (type, size, format, etc.)**

· Median size is 1, 000 rows of data and ten columns

· Oldest spreadsheet is 9 years old and created in an earlier version of excel

· XMV-based file format for Excel 2010 and 2007 (.xlxs)

· Prefers excel but accepts CSV

· 1.16GB in Excel workbooks

Always ends up in excel

**What licenses apply to the data**

·

**What facilities and equipment will be required (hard disk space, backup server, central repository, off-site repository, etc.)**

·

**What data management practices (backups, storage, access control, archiving etc.) will be used**

·

**Who will own and have access to the data**

· Willing to publicly share collection of data

· Share with those people she wants to

· She is a well known resource in her department

· 95% of the data is from external sources

o Other researchers

§ Probably cant be public?

o Various governments

o Private corporations etc..

§ Cant be public?

Wikidata\*\* - Might not work because it can be edited by everyone

**Which data will retain value after the life of the project**

* All data

**What metadata and linked open data strategies will be employed**



**How will its reuse be enabled and long-term preservation ensured after the original research is completed**