Professor Chartreuse

**Introduction & Recommendations**

This data management plan is intended to provide context to ‘science of science’ research conducted at JCU. Research in this area is performed by Professor Chartreuse and his lab. Specifically, the plan is meant to outline data management practices of the lab for the purpose of external stakeholders’ consideration.

As it currently stands, Professor Chartreuse uses Excel files to store data he has collected using PubMed and other research search platforms. He is seeking to update his data management practices to be able to share his research files with other colleagues more easily, to improve version control, to consolidate files, and to combine research information with other colleagues in his lab.

Based on these wishes, it is recommended that Professor Chartreuse and his teammates register for an Open Science Framework account. Open Science It is a secure, well known data repository. The OSF permits academics to create and manage research projects on the cloud. Different levels of access to others can be granted by the project creator.

Each OSF project can hold up to 5 GB of data. Additional data can be stored by linking another repository, such as Google Drive, One Drive, Amazon, etc. to the OSF account. Projects can be assigned metadata and keywords for searching purposes. This will assist in the findable and accessible guiding principles of data management. Files can be shared publicly via NCBI through OSF. Before sharing files, the creator applies a license to the files to be shared to restrict the use of the data based on the creator’s desires.

Beyond the reasons outlined above, OSF is recommended because it is a free resource. It is also recommended that Professor Chartreuse use Google Docs as an additional data repository as this is also a free resource for less than 15 GB of data. Professor Chartreuse may also take comfort in using Google Docs since he is already comfortable using Excel files. Beyond 15 GB, it would cost Professor Chartreuse’s lab CAD 2.79 per month for 100 GB of data storage (Google, 2018). Alternatively, Professor Chartreuse could use a data repository that JCU already subscribes to an funds (ex. OneDrive, Sharepoint, etc.)

With this being said, for the purpose of the data management plan below, it is assumed that Professor Chartreuse will adopt the recommendation of using the OSF. It is also assumed that he will employ Google Docs as a linked data repository for additional data storage.

**Data Management Plan for the ‘Science of Science’ Research Lab at JCU**

*Introduction*

The ‘Science of Science’ lab is a research team at JCU. It works to understand how researchers work to generate new knowledge. The lab is head by Professor Chartereuse.

* Data Collection
  + What data will you collect or create?

Surveys and interviews are created and conducted to collect primary data for research purposes. The surveys and interviews are conducted by lab researchers and work to collect data from other researchers to gain insight into research practices and the generation of new knowledge.

Secondary data is also collected by the lab through readings of peer-reviewed research papers and articles. This information is collected by using a web API and JSON text to search relevant papers using keywords. Keywords, date of publication, name of publication, authors, areas of study, and insights from the publication are all noted. Secondary data is stored in Excel files (ex. xls, xmlx, csv).

* + How will the data be collected or created?
* Documentation and Metadata
  + What documentation and metadata will accompany the data?

Each document is title according to date at the present time. However, the lab may consider switching its naming practice to be project-specific. Metadata included in each documentation includes:

* + - Keyword searched
    - Keywords relevant to the file
    - Date created
    - Date of last update
    - Name of document
    - Names of articles noted in the Document
    - Authors of document
* Ethics and Legal Compliance
  + How will you manage any ethical issues?

The research lab works to follow the ethical standards as designated by JCU. For each research project that includes the collection of primary data, the research lab gains the Ethics Approval of the JCU Ethics Department before proceeding with conducting surveys or interviews. The research lab also gains written consent from all research participants to confirm that the participants understand their rights, permit the collection of their personal information, and permit their answers be used for research purposes. Additionally, participants may answer anonymously if desired.

* + How will you manage copyright and Intellectual Property Rights (IP/IPR) issues?

All information created and collected by the research lab is the property of JCU.

* Storage and Backup
  + How will the data be stored and backed up during the research?

Data will be stored on One Drive and on Open Science Framework. As well, the working copies or older versions of Excel sheets may be stored on the hard drives of the research lab’s work computers. It will automatically be backed up by Google Drive and by Open Science Framework. Both platforms offer automatic version control.

Open Science Framework stores up to 5GB of data per project. The system is backed up daily and past versions of files are kept on the Amazon for 60 days. These files can be retrieved by communicating with Open Science Framework.

* + How will you manage access and security?

Access will be permitted to research lab teammates via Open Science Framework. Security will occur via access through a username and password. Information will be shared with the public through Open Science Framework. Specific files to be shared will require the permission to be shared through Open Science by Professor Chartereuse. Only information that does not include personal information will be shared, unless respondents have specifically given permission otherwise.

Information will be shared via Open Science Framework under a designated creative commons license of: BY-NC-SA. This license permits the redistribution of the work so long that the original work is attributed. It also allows the user to select their own license type for their own adaptations of the work. However, it does not allow users to modify or adapt the original work or to use the work for commercial purposes.

* Selection and Preservation
  + Which data are of long-term value and should be retained, shared, and/or preserved? What is the long-term preservation plan for the dataset?

The primary data collected by the lab in the form of surveys and interviews have long-term value and should be retained and preserved. It may also be shared via Open Science Framework so long as the lead scientist wishes it to be made publicly available.

Secondary data will be shared openly under the creative commons license and described above. These data files should be kept for five years. As time passes, the files that prove to be most use to the public (ie. are the most accessed and cited) should be kept available. However, if files are not accessed or cited for a full year, they can be removed from the public platform.

The long-term preservation plan for the dataset is to train all incoming lab employees and students on the data management practices of the lab. Each individual will be responsible for their own file management and for conforming to the practices of the lab. Once every six months, Professor Chartereuse or another designated individual will manage the Open Science portal for each ongoing project and remove any old versions of files that are no longer required.

The long-term preservation plan of the data is to keep relevant data available on Open Science Framework so that other institutions may access the data long-term. This research is expected to be ongoing for years to come.

* Data Sharing
  + How will you share the data? Are any restrictions on data sharing required?

Data will be shared via the open source portal of Open Science Framework. Only files with information relevant to the public will be shared. Additionally, only the most up to date version of each research files will be shared. Based on the creative commons licensing to be assigned to the shared files, users outside of the Chartreuse lab will be permitted to view the data files, but will not be able to edit the original content of the work.

* Responsibilities and Resources
  + Who will be responsible for data management?

Professor Chartreues will be primarily responsible for the data management of the lab. Additionally, Professor Chartreuse may choose to delegate his responsibilities to another trained member of the lab.

* + What resources will you require to deliver your plan?

To manage the data, continued access to and use of the Microsoft Office Suite will be required. Additionally, an Open Science Framework account will be required. The Open Science Framework