Data Management Consultation report

Researcher: uses the data

Our Manager: identifies the resources require.

* What data will be created or collected (type, size, format, etc.)
  + Marine wildlife, measure ocean conditions, signals from animal tags.
  + NetCDF (use and storage format)
  + 300mb raw sensor data per day, converted to 500 mb of NetCDF formatted data, both versions are kept.
  + Minas Basin data, not currently kept.
  + Citizen scientists.
  + Field notes (animals captured and tagged, marked recapture population estimation experiments, observational studies conducted during classes)
  + 2 members of team run simulation models, predicting animal population and movements based on buoy data, zipped CSV files
  + Citizens science data downloaded as TSV
* What licenses apply to the data
  + Open data and data created by Prof. Periwinkle
* 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
  + Team of researchers, students and staff
  + Visual data made available to GP.
* Which data will retain value after the life of the project
  + Data collected used to write papers and create exercises for grad classes.
* What metadata and linked open data strategies will be employed
  + Darwin core field notes ( can be xml, csv, rdf), extension of dublin core, suggest CSV?
  + They will be told that DC is used in all fields on IOSBIS.org website
* How will its reuse be enabled and long-term preservation ensured after the original research is completed
* How much will the storage of this data cost (cloud and/or hard drives)

Data Lifecycle

* Plan consent for sharing
  + No current sharing agreement with other research groups
  + In contrast NOAA in the US has the IOOS system (uses NetCDF) European EMOD systems which both span their country, Canada needs a similar system but for international sharing.
* Collect data (experiment, observe, measure, simulate)
  + Collected since 1998, including personal shelf of data in different formats, reflecting evolution of tech. All needs to be converted and amalgamated into one source
* Capture and create metadata
  + Metadata for language of data

1. Processing Data
   * Enter data, digitize, transcribe, translate
   * Check, validate, clean data
   * Anonymize data where necessary
   * Describe data
   * Manage and store data
     1. External drives/USB keys, not currently encrypted., Storage of all data types together?
     2. TSV and CSV files can be stored on github,

# Saving netCDF data in ArcGIS

* + - 1. You can display netCDF data as a layer or table in the map, and save it in the project. The netCDF data displayed on a map is not saved with it. NetCDF layers or tables reference the netCDF file on disk.
      2. You can also save netCDF layers outside your map as a file on disk. When you save a netCDF layer to disk, you also save the symbolization. When you add the saved layer file to another map, it will draw exactly as it was saved.

1. Analysing Data
   * Interpret data
     1. Document on share drive (help document) needs to found and updated
     2. Training staff and students to analyze and interpret data, and how to use data analysis systems
   * Derive data
   * Produce research outputs
   * Author publications
   * Prepare data for preservation
2. Preserving Data
   * Migrate data to best format
   * Migrate data to suitable medium
   * Back-up and store data
   * Create metadata and documentation
   * Archive data
3. Giving Access to Data
   * Distribute data
     1. Via email request, by dropbox
   * Share data
     1. Open data?
     2. Access to other data through relationships with other institutions
     3. External hard drives/usb keys
     4. As students graduate they share their data with current students in the lab or take with them on personal devices
     5. Uniform data sharing method, must be justified to professor.
   * Control access
   * Establish copyright
   * Promote data
4. Re-using Data
   * Follow-up research
   * New research
   * Undertake research reviews
   * Scrutinize findings
   * Teach and learn

Data Management Plan

Data description: A description of the information to be gathered; the nature and scale of the data that will be generated or collected

* Marine wildlife, measure ocean conditions, signals from animal tags.
* Field notes (animals captured and tagged, marked recapture population estimation experiments, observational studies conducted during classes)
* Minas Basin data, not currently kept.

Existing data: A survey of existing data relevant to the project and a discussion of whether and how these data will be integrated.

Format: Formats in which the data will be generated, maintained and made available, including a justification for the procedural and archival appropriateness of those formats.

* NetCDF (use and storage format)
  + 300mb raw sensor data per day, converted to 500 mb of NetCDF formatted data, both versions are kept.
  + Cloud bucket storage?

Metadata: A description of the metadata to be provided along with the generated data, and a discussion of the metadata standards used.

Storage and backup: Storage methods and backup procedures for the data, including the physical and cyber resources and facilities that will be used for the effective preservation and storage of the research data.

## **IDV (Integrated Data Viewer)**

* Unidata's [Integrated Data Viewer (IDV)](https://www.unidata.ucar.edu/software/idv/) is a Java application (for Java 1.4 or later) that can be used to display a variety of netCDF files, particularly well formatted, geolocated datasets. Features include:
* Access to local and remote netCDF files and a variety of [other data formats](https://www.unidata.ucar.edu/software/idv/docs/userguide/data/DataSources.html)
* Slicing and probing of multidimensional data
* Support for netCDF conventions (CF, COARDS, NUWG, AWIPS)
* InstallAnywhere installers for easy download and installation
* Save display state to a bundle for easy recreation of views
* Support for non-gridded data through the [Common Data Model (CDM)](https://www.unidata.ucar.edu/software/netcdf-java/CDM/)
* The IDV uses the [VisAD Java library](http://www.ssec.wisc.edu/~billh/visad.html) for interactive and collaborative visualization and analysis and the [netCDF Java library](https://www.unidata.ucar.edu/software/netcdf-java/) for reading and manipulating netCDF files.

Security: A description of technical and procedural protections for information, including confidential information, and how permissions, restrictions, and embargoes will be enforced.

Responsibility: Names of the individuals responsible for data management in the research project.

Intellectual property rights: Entities or persons who will hold the intellectual property rights to the data, and how IP will be protected if necessary. Any copyright constraints (e.g., copyrighted data collection instruments) should be noted.

* What licenses apply to the data
  + Open data and data created by Prof. Periwinkle

Access and sharing: A description of how data will be shared, including access procedures, embargo periods, technical mechanisms for dissemination and whether access will be open or granted only to specific user groups. A timeframe for data sharing and publishing should also be provided.

* Open data
* Plan consent for sharing
  + No current sharing agreement with other research groups
  + In contrast NOAA in the US has the IOOS system (uses NetCDF) European EMOD systems which both span their country, Canada needs a similar system but for international sharing.

Audience: The potential secondary users of the data.

Selection and retention periods: A description of how data will be selected for archiving, how long the data will be held, and plans for eventual transition or termination of the data collection in the future.

Archiving and preservation: The procedures in place or envisioned for long-term archiving and preservation of the data, including succession plans for the data should the expected archiving entity go out of existence.

Ethics and Privacy: A discussion of how informed consent will be handled and how privacy will be protected, including any exceptional arrangements that might be needed to protect participant confidentiality, and other ethical issues that may arise.

Budget: The costs of preparing data and documentation for archiving and how these costs will be paid. Requests for funding may be included.

Data Organization: How the data will be managed during the project, with information about version control, naming conventions, etc.

Quality Assurance: Procedures for ensuring data quality during the project.

Legal Requirements: A listing of all relevant federal or funder requirements for data management and data sharing.

**Case 2 - Professor Green**

* What data will be created or collected (type, size, format, etc.)
  + Interdisciplinary primary care teams, team work in hospital environments
  + Textual content analysis, 383 documents (teams, outcomes and practices) word, pdf, plain text and other. All are digital
  + Conducting 15 interviews with key informants, 60 mins each, recorded as mp3 files, transcribed in microsoft word, asked q’s about how care teams work, concerned with loss of jobs, reputations.
* What licenses apply to the data
  + Professor Green wants copyrights in place
* 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
  + Professor Green
* Which data will retain value after the life of the project
* What metadata and linked open data strategies will be employed
  + Interviews conducted over the phone
  + Languages: English, German, French
* How will its reuse be enabled and long-term preservation ensured after the original research is completed
* How much will the storage of this data cost (cloud and/or hard drives)

Data Lifecycle

1. Creating Data

* Design research
* Plan data management (formats, storage etc.)
* Plan consent for sharing
* Locate existing data
* Collect data (experiment, observe, measure, simulate)
* Capture and create metadata

1. Processing Data
   * Enter data, digitize, transcribe, translate
   * Check, validate, clean data
   * Anonymize data where necessary
     1. Interviews encrypted, anonymized, limited access
     2. Promised confidentiality to participants, small community, numeric identifications and legend stored separately. Secure database.
   * Describe data
   * Manage and store data
2. Analysing Data
   * Interpret data
   * Derive data
   * Produce research outputs
   * Author publications
   * Prepare data for preservation
3. Preserving Data
   * Migrate data to best format
   * Migrate data to suitable medium
   * Back-up and store data
     1. Storing data under zotero database
     2. Quantitative data stored in excel spreadsheet
     3. Suggest backup on secure cloud, USB not ideal to be travelling with Prof. Green
     4. Choose 3 options for data storage
   * Create metadata and documentation
     1. 15 interviews, each interview is an hour long, and I am encoding them in mp3 format, 128kbps = each file would be 57.6 mbs x 15= 864 mb
   * Archive data
4. Giving Access to Data
   * Distribute data
   * Share data
     1. With masters students and colleagues
     2. Zotero for documents, dropbox for audio, google docs for transcriptions
     3. Project meetings
     4. Binder of printed excel sheets -- MUST be digitized by MASTERS students w/ knowledge of health data, determine relevance.
     5. "The data is not shared with the public. As I mentioned, it can identify people. Share with my team are the notes I make (recordings, spreadsheets, text files) after I have gone through the data. And that is only shared with my project members. If they want more information, they have to ask me."
   * Control access SEE ABOVE, only Prof. Green has full access to data
   * Establish copyright
   * Promote data
5. Re-using Data
   * Follow-up research
   * New research
   * Undertake research reviews
   * Scrutinize findings
   * Teach and learn

Data Management Plan

Data description: A description of the information to be gathered; the nature and scale of the data that will be generated or collected

Dr. Green is currently working on a project with two Masters students. In this project, Dr. Green and his team has two methods of data collection. The first is a textual analysis of 383 documents and the second is conducting interviews. This project will collect textual data which consists of a mix of document formats as well quantitative data stored in an excel spreadsheet. The project will produce data in the form of interviews will be conducted over the phone, requiring MP3 files and transcribed to Microsoft word. Dr. Green also uses open data sets regarding healthcare in his research.

Existing data: A survey of existing data relevant to the project and a discussion of whether and how these data will be integrated.

There is currently a binder of printed excel spreadsheets, it is our recommendation that Dr. Green's trained Masters students go through to find any relevant data to the current research. We will then provide the students with step by step instructions of how to digitize relevant data.

Format: Formats in which the data will be generated, maintained and made available, including a justification for the procedural and archival appropriateness of those formats.

Data collection of the documents for textual analysis have been obtained from healthcare organizations and quantitative information regarding these documents will be recorded in a Microsoft Excel spreadsheet using DDI XML metadata documentation. Phone interviews will be recorded as MP3 file format.

To insure preservation of data, it is suggested that the document files and transcripts of interviews be converted to one file format of PDF. This will ensure that documents can be accessed in the future and will preserve the data as it is, meaning that it cannot be altered. PDF provides a high level data encryption and offers a secure way to transfer data over the internet. Dr. Green can customize the user access level, setting it so he is the only one who can access the files and add a digital signature. A digital signature confirms the authenticity of the text and protect Dr. Green's work.

Metadata: A description of the metadata to be provided along with the generated data, and a discussion of the metadata standards used.

Metadata for Dr. Green's project will be tagged in XML using the Data Documentation Initiative format. This is a widely used international standard for describing data from both social and science fields. Metadata about data will be recorded to give Dr. Green context about what the file contains.

Storage and backup: Storage methods and backup procedures for the data, including the physical and cyber resources and facilities that will be used for the effective preservation and storage of the research data.

To ensure that the data created by Dr. Green's project is accessible for future use, data will be deposited in (**repository)** and will be backed up on an external USB key. Dr. Green's Masters students will be tasked with migrating research data to new formats, platforms, and storage as required for the continuation of his research. A master copy of each file will be contained in the USB Key, as well a backup on the repository to ensure digital preservation.

Security: A description of technical and procedural protections for information, including confidential information, and how permissions, restrictions, and embargoes will be enforced.

The data files from Dr. Green's research project will be managed, processed, and stored in a secure environment by the researcher himself. To ensure security, we will install a lockable computer system on his laptop that has a strong firewall system with software to protect from viruses/malware. The folder containing files for his project will be encrypted and password protected. Any identifying information about the informants will be anonymized.

Responsibility: Names of the individuals responsible for data management in the research project.

Dr. Green has assigned us to be the data manager to act as steward for the data throughout the data lifecylce. At the conclusion of the current grant, we will give all data management plan documentations to the new assigned data manager.

Intellectual property rights: Entities or persons who will hold the intellectual property rights to the data, and how IP will be protected if necessary. Any copyright constraints (e.g., copyrighted data collection instruments) should be noted.

Dr. Green will hold intellectual property rights for all research data and information directly created by the project.

Access and sharing: A description of how data will be shared, including access procedures, embargo periods, technical mechanisms for dissemination and whether access will be open or granted only to specific user groups. A timeframe for data sharing and publishing should also be provided.

Copyrights are in place, so that Professor Green has control, owns the data.

Anonymize data where necessary

Interviews encrypted, anonymized, limited access

Promised confidentiality to participants, small community, numeric identifications and legend stored separately. Secure database.

Use one sharing space for all file types - **Standard Dropbox**: it is something that you are familiar with, and limited users(3). 2TB (2000GB) of space more than suffices for files.

Professor Green can have a key (on an encrypted usb, stored in a safe) that indicates a numerical code for each interviewee. All the identifying information will be limited to access with Professor Green. The rest of the info can be accessed by masters students by being give permission by professor Green. They can have access to the Drop box data.

Data can be stored on an encrypted laptop that can be brought to presentations as need be. Professor Green will still feel that he has control of the information because it is in his possession

Audience: The potential secondary users of the data.

Selection and retention periods: A description of how data will be selected for archiving, how long the data will be held, and plans for eventual transition or termination of the data collection in the future.

Dr. Green's project makes use of 383 documents from healthcare organizations, once the textual analysis is complete it is recommended that these files are archived and retained for 1 year after the textual analysis part of the project is completed. Data generated from phone interviews will be recorded and transcribed. Transcriptions of interviews will be retained by the repository as part of the permanent research data; MP3 files will be archived after being transcribed and will remain as part of the master copy of data.

Archiving and preservation: The procedures in place or envisioned for long-term archiving and preservation of the data, including succession plans for the data should the expected archiving entity go out of existence.

Ethics and Privacy: A discussion of how informed consent will be handled and how privacy will be protected, including any exceptional arrangements that might be needed to protect participant confidentiality, and other ethical issues that may arise.

To protect the privacy of informants, Dr. Green will use an informed consent form. The form states that the researcher will not identify the participant by name in any information resulting from the interview and that confidentiality will be secure. Data will be subject to the standard practice of anonymizing data to ensure privacy. During analysis and storage, only Dr. Green and approved members of his research team will be allowed to access data. The research team will then remove any direct identifiers before it is deposited into the storage repository.

Budget: The costs of preparing data and documentation for archiving and how these costs will be paid. Requests for funding may be included.

Costs for creating, maintaining, and storing data will be covered by a grant Dr. Green has received from CIHR. Budget resources will go towards paying for a digital repository and ensuring version control for the long term plan of the research project.

Data Organization: How the data will be managed during the project, with information about version control, naming conventions, etc.

A private GitHub account will be purchased with the budget to manage version control during the course of Dr. Green's research. Variables will use a standardized naming convention which will consist of a determined prefix, root, and suffix system. This will ensure that Dr. Green and his team know which is the most recent version of their data analysis, and will allow the team to go back to previous versions at any time. Separate files will be managed for the different type of data produced.

Quality Assurance: Procedures for ensuring data quality during the project.

Legal Requirements: A listing of all relevant federal or funder requirements for data management and data sharing.

Case 2 - Professor Green

* you know we have a mix document files (PDF, Word and plain text). Could you tell us the average size of those documents?
* "the file size ranges from a few hundred kilobytes to 25-30MB."
* Could you please provide us the name or the code for each hospital that you are planning to examining
* "I work with a few hospitals in Nova Scotia, one hospital in France, one in Germany and a few more in Calgary."
* This case did not mentioned the size of some data collection , how we can know the data size?
* "You are the expert! There are 15 interviews, each interview is an hour long, and I am encoding them in mp3 format, 128kbps."
* Is the data that was collected in 2002 related to the data that was mentioned in this case?
* "There might something relevant in the 2002 binder with the printed spreadsheets. I would like for you to look through it. It does not contain anything that can identify participants. "
* Is there a specific size for the zotero database ? Also,how much space do you wish to have for your future research (How much do you think you'll use in 5 years from now?
* "the current size of all my files right now fits in my USB drive which is 64GB and it is not full. There are about 40GB left over. I predict that the data influx will increase in the near future. My research is very important! You say five years? It will be at least ten. I predict it will triple in size."
* There are 383 documents with several formats. Would you like to have them available on the original format or would you be willing to have one standard format for your database ?
* "Why would I want them in one format? Do you plan on changing my original data??"
* 'll make use of a specific data storage service. As a Dal faculty member you can make use of the university's space to store your data. Is Dal's repository ok, or do you have any specific preference ?
* "the Dal repository is fine for now, but I would like to know what other options are available. Not too many, **two or three** would be good."
* If you are required to share interview instruments or anonymized data to conduct your studies, how is this done? What is the language of communication?
* "I am not required to share the original files or interviews. Only information that does not allow for subject identification. That data needs to be translated to the official language of each collaborator's location."
* Please send us a copy of the participant consent form so that we can fully understand the terms of confidentiality and ensure that we don’t breach them.
  + <https://web.stanford.edu/group/ncpi/unspecified/student_assess_toolkit/pdf/sampleinformedconsent.pdf>
  + You mentioned earlier that you expect to get a new grant once your current funding runs out. It would be helpful to know who provides your research funding so we can ensure the data management plan we are developing for you adheres to that funder's policies.
  + my funding comes from CIHR (Canadian Institutes of Health Research).
  + What is the reason that you have your transcriptions stored on Google Drive? Are members of the team doing live collaboration to edit those transcriptions?
  + no, I simply like to have quick access to the information, even when I am not on my computer in my office.
  + The founding comes from CIHR. Are you working in the Dalhousie or in the Century University? because the consent form shows that you are from Century University and I would like to make sure about that .
  + I am working at Dalhousie. The consent form is a general template for consent forms.
  + We would like to confirm that the data from the data healthcare providers and internal networks is the same as the open data sets you have collected. What is the type of data and format of the open data sets that you use in your research?
  + yes, they are the same. The open data is in tsv files. The type of data is explained in my original briefing file.

**Case 4 - Professor Chartreuse**

* What data will be created or collected (type, size, format, etc.)
  + Currently 20GB in size and expanding. SciSci often deals with big data, so therefore it might be necessary to use machine learning.
* What licenses apply to the data
  + Maintain PubMed licensing
* What facilities and equipment will be required (hard disk space, backup server, central repository, off-site repository, etc.)
* Professor Chartreuse will be a backup server as he is worried about the possibility of his data being lost. Cloud storage might work best for him to share his data with colleagues and collaborators.
* What data management practices (backups, storage, access control, archiving etc.) will be used
  + Convince Professor Chartreuse that Excel files are difficult to version control and errors are difficult to notice, so another method of storing data should be implemented
* Who will own and have access to the data
  + Professor Chartreuse will own the data, but will be able to share his data will colleagues and collaborators. His students, colleagues, and collaborators will not have to request permission to view the data.
* Which data will retain value after the life of the project
* What metadata and linked open data strategies will be employed
  + Need for standardization; ontologies and controlled vocabularies. Possibly UniProt, as it can be searched using PubMed and maintains non-commercial use, and can be accessed online
* How will its reuse be enabled and long-term preservation ensured after the original research is completed
  + Perl Scripts
* How much will the storage of this data cost (cloud and/or hard drives)

Data Lifecycle

1. Creating Data

* Design research
* Plan data management (formats, storage etc.)
* Plan consent for sharing
* Locate existing data
* Collect data (experiment, observe, measure, simulate)
* Capture and create metadata
* date, name, keywords, publications that used the data, the authors.

1. Processing Data
   * Enter data, digitize, transcribe, translate
   * Check, validate, clean data
   * Anonymize data where necessary
   * Describe data
   * Manage and store data
2. Analysing Data
   * Interpret data
   * Derive data
   * Produce research outputs
   * Author publications
   * Prepare data for preservation
3. Preserving Data
   * Migrate data to best format
   * Migrate data to suitable medium
   * Back-up and store data
   * Create metadata and documentation
   * Archive data
4. Giving Access to Data
   * Distribute data
   * Share data
   * Control access
     1. Control access for editing, permission access for viewing data
   * Establish copyright
   * Promote data
5. Re-using Data
   * Follow-up research
   * New research
   * Undertake research reviews
   * Scrutinize findings
   * Teach and learn

Case 4 - Professor Chartreuse

* Is the "science of science" considered part of social sciences? ( I need to know as to select the proper metadata standard for your research)
* "The science of science (SciSci) is based on a transdisciplinary approach that uses large data sets to study the mechanisms underlying the doing of science—from the choice of a research problem to career trajectories and progress within a field"
* What's the size of the data that you have so far? and how long do you think your research is going to last?
* "My data collection is about 20GB is size, and I plan on continuing until the day I am no longer here. "
* What type of metadata would you like to see happen with the files? Metadata is how you label your files. Example: title of the file, the date of the creation of the file. the subject of the file and so on.
* "I would like the metadata to include relevant information for me to be able to query it later on. So, definitely date, name, keywords, publications that used the data, the authors. Those a mandatory, if you think that you can add more, so that I can make more complex queries, let me know."
* What type of openness would you prefer to see for the data? For example, the database could be close to only your view or it can be open to everyone.
* "Since I collect my data from public sources (mainly Pubmed), I would like to keep their licensing and definitely make it available for non commercial use only."
* What data do you consider to be the most relevant in retaining?
* "Any type of data that will lead to a publication."
* Where is the data being stored?
* "Right now the data lives in Excel files in my hard drive."
* How do you organize your data in excel?
* "Simple excel files with a maximum of 3 worksheets."
* Is there a specific format structure in the way you store your data in your external hard drive, that you would like to keep (i.e Different Files structure)?
* "I store it by date. I have found that it makes it hard to go back and search for things. So I am not adamant in keeping it."
* Is there a specific budget range you are looking to maintain? Or spend no money at all?
  + “there is some money allocated for this. But if you can give me an option that would be free, I would be more than willing to take that.”
  + For how many years have you been collecting data? I am trying to gain a better understanding of how much data storage on average you require per year. Also, do you have data in your files that are from a previous area of study? ie. Is there data in your files that may not be necessary to include in a database centred around resources on the science of science?
  + I have been collecting data for four years. All the data is for this particular area of study
  + How far will the collaboration extend to, such as will you give colleagues access to editing your database if you were to adopt a database management system (i.e. SQL database)?
  + I would like for them to be able to view the data, but not be able to edit it. And they need to get permission to access it first.
  + Anyone of your intended audience (i.e graduate students, colleagues) can view the data at any time without permission, but need permission to access such as to use the data, like download it? Also, i know in the future there are plans for colleagues to add to the database, that means they will add to your own personal database? and they will have to be granted access to add to your database to do that? Correct me if I am wrong.
  + Yes, your assumptions are correct
  + I was hoping you could elaborate a little bit more on your issue with the keyword search. Why would you return the results by hand into an excel sheet?
  + I do it by hand because that is what works best for me and that's the way I know how to do it. It has nothing to do with PubMed or its infrastructure.

Data Management Consultation Report for Professor Chartreuse

This project generates SciSci research data that will be gathered by Professor Chartreuse, distributed to his colleagues for collaborative purposes and to his graduate students at JCU for educational purposes. The data are predominantly from public domain resources and will maintain the licensing of PubMed. The data will require different levels of accessibility due to their sensitive nature. Professor Chartreuse and his colleagues require full access to the data and students should be able to download the data without altering the original dataset. Professor Chartreuse is responsible for data storage and providing access to other parties. The 20GB of previously generated data will be converted from JSON to CSV files, while new data that is produced will be saved as CSV in order to be machine readable. Perl will be used as the programming language, and Professor Chartreuse will be able to extract from the dataset using a graphical user interface. All saved data files will be deposited in cloud storage using Open Science Framework.

Open Science Framework will allow for secure access to the dataset. Professor Chartreuse has the ability to control who has access to the data and can tailor the privacy to his team of collaborators while sharing data with his graduate students. Open Science Framework is open source, free to use, and has version control capabilities which will reduce risk in losing valuable data. OSF allows contributors to tag the research data using keywords to increase its findability and to be able to locate data quickly. This function will provide ease of data sharing between the research team and assists with reproducibility amongst other researchers.