DATA MANAGEMENT PLAN - TEAMWORK WITHIN INTERDISCIPLINARY PRIMARY CARE TEAMS IN HOSPITAL ENVIRONMENTS

Consultants: F.A.M.E DATA MANAGEMENT – for data that are **F**indable, **A**ccessible, **M**emorable, & **E**nduring. Team members: Jenny Farbstein, Lisa Nardecchia, Victoria Roberts, Carlye Stein

Principle Investigator: Dr. Green

Funder(s): Canadian Institutes of Health Research (CIHR)

Plan Details:
The following data management plan has been developed for Dr. Green's research data relating to teamwork in hospital environments, specifically within interdisciplinary primary care teams.

DATA COLLECTION

Data are currently collected using two methods. The researcher first conducts content and textual analysis of individual documents describing different aspects of the primary care teams. These documents have been obtained from healthcare organizations. Data are also collected from interviews conducted with key informants in both audio and text format.

It is recommended that Dr. Green adopts a file format that can reasonably be assumed to have longevity, i.e. one that is commonly used and accessible in various software packages or has an open format so that the data can be accessed, retrieved, and used in the future.

With respect to the interview transcripts, we highly recommend that text (.txt) files are used as opposed to Microsoft Word (MS Word). We recommend that MS Word documents be batch converted into text files and that all future files of this type are saved in .txt format. If MS Word is the preferred software to work in, it can still be used to modify files as long as the finished output is saved as a .txt file instead of the default. Text files are being proposed as they remove formatting, take up little space, and are easy to work with. Since PDF files are locked to prevent textual edits, it is recommended that they be batch converted to text files to ensure consistency of data types. It is appropriate to continue saving the audio interviews as MP3 files as this format meets both of the longevity and usability recommendations mentioned above.

All Excel files should be batch converted to comma-separated-values (.csv) files. This is a widely accepted and interoperable format that can opened using other programs aside from Microsoft.

DOCUMENTATION AND METADATA

Metadata relating to the textual analysis and interviews such as who created the data and when, how the data were created, their quality, accuracy and information to facilitate understanding and reuse are likely being captured in the research process. This information could then be set out in a table, ideally in .csv format, for reasons stated above, to enhance searchability, sharing, and reuse of the data. However, it is further recommended that a more formal process of capturing metadata be used so that management of the research data is consistent with the Tri-Agency Statement on Principles of Digital Data Management which states that "All research data should be accompanied by metadata to enable future users to access, understand and reuse the data."[1]

Data Documentation Initiative (DDI) is being suggested it is an international metadata standard for describing the data produced by surveys and other observational methods in the social, behavioral, economic, and health sciences. More information can be found here: http://www.ddialliance.org. The Canadian Data Research Centres program has adopted this

[1]Government of Canada. 2016. Data Management. Retrieved from http://www.science.gc.ca/eic/site/063.nsf/eng/h_83F7624E.html

standard and it is also recommended by the ICPSR repository which will be disussed in detail in upcoming sections.

ETHICS AND LEGAL COMPLIANCE

To ensure ethical compliance, it is recommended that the guidelines established by the institutional Research Ethics Board be followed.

Copyright and Intellectual Property Rights (IP/IPR)

As the recommended open access storage repository is openICPSR for the research data and finished publication, its licensing terms should be inherited. openICPSR uses Creative Commons Attribution 4.0 International (CC-BY) which requires that the original researcher be credited while allowing users to read, share, modify, enhance, etc. the data. If users make changes to the data, that must be noted so that the integrity of the original is maintained. Creative Commons licenses are reputable and have terms and standards that future researchers will recognize.

STORAGE AND BACKUP

In order to simplify the storage implementation and reduce associated overhead costs, we suggest that the researcher takes advantage of the institution's in-house infrastructure. Files are to be stored on the OneDrive and Dataverse environments of the institution. Although CIHR does not mandate a specific storage protocol for this type of data, discussions with internal IT personnel confirmed that OneDrive was initially selected for its adherence to Canadian data protection laws, therefore making it an appropriate choice. Leveraging OneDrive allows the data management team to rely on the existing university backup protocols, which write OneDrive data from the internal cloud storage to an institutional OneDrive server nightly.

As CIHR requires that finished publications and datasets be placed in an open repository, OpenICPSR (www.openICPSR.org) should be used for the finished output. Dal's Dataverse institutional repository will be leveraged until the time of publication because it has similar advantages to those of OneDrive, along with customizable and easy-to-use backup protocols.

Access and security

As collaboration and communication with European contributors may be required, files can be shared securely using FileExchange. A secure upload link will be sent, with contributors only being given one-time access to receive the required content. A new share link will be used each time.

OneDrive's built in group management permissions will be used to ensure all relevant stakeholders internal to the insitution have read access, with additional permissions evaluated on a case by case basis.

Dr. Green is strongly encouraged to follow security best practices and transfer sensitive interview and survery data that could identify participants to a restricted OneDrive folder, which would resistct access to anyone else while reducing the risk of data loss through theft or human error.

SELECTION AND PRESERVATION

Research data resulting from CIHR funding should normally be preserved in a publicly accessible, secure and curated repository or other platform for discovery and reuse by others. CIHR policy also states that datasets must be retained for a minimum of 5 years after the end of the grant. This requirement applies to all data whether published or non-published. In addition, the university's research ethics board policies on retention and preservation of data must be followed.

As this data contain personally-identifiable information, it is recommended that multiple versions of the data are retained: one that is suitable for public release (de-identified), and one that is suitable for further research but that is available on a highly restricted basis.

It in order to prepare data for preservation, it is necessary to establish an interoprable file format standard: .txt and .csv files are suggested formats. Audio content can remain in MP3 format. The printed spreadsheets in the binder from 2002 will need to be scanned and converted to digital format for the purpose of analysis and data cleansing if possible. Any data that cannot be used to address research questions of interest could then be archived as appropriate.

Long-term preservation of the dataset(s)

In compliance with the requirements of CIHR, we recommend that the annonymized research data and the associated final article(s) be posted to openICPSR for long-term preservation. openICPSR is a searchable service and an extension of the Inter-university Consortium for Political and Social Research, of which Dalhousie University is a member (https://www.icpsr.umich.edu/icpsrweb/membership/administration/institutions). Therefore member subscription fees are already paid and there is no cost to users to access the information. openICPSR gives the contributor control over what is published, when it is published, and in what format. Research data is available to users within 24 hours of being posted.

DATA SHARING

As mentioned in other sections, data sharing will be facilitated through OneDrive with appropriate access given to identified stakeholders. Any other data sharing will be done through FileExchange in order to give external collaborators secure access to the content they require.

Since Dalhousie resources that are already in place and widely used provide the same functionality as Zotero, it is our recommendation to stay within Dalhousie as its internal infrastructure provides the benefit of added security and ensures compliance with confidentiality policies.

Data can be shared in the original language of the collection. Translation will be the responsibility of the collaborators in the target-language country.

Restrictions on data sharing

The F.A.M.E Data Management understand that Dr. Green’s primary concern is to protect sensitive data that may cause the study participants’ loss of reputation and other career-related risks. Therefore, any data that are shared will exclude sensitive data. Only the lead researcher

will have access to sensitive information. The use of USB drives, Dropbox, or Google Docs for data storage and sharing should be strictly prohibited and replaced with FileExhange as recommended in previous sections.

RESPONSIBILITIES AND RESOURCES

Responsibility for data management

Documentation will be created by a designated resource such as a data manager. Best practice and start-up assistance will be provided by F.A.M.E Data Management and transitioned after the end of their responsibilities.

Resources required to deliver the data management plan

The following resources will be required:

Internal Information Architecture Resources
OneDrive
FileExchange
Dataverse

Human Resources
Data Management Research Assistant (RA) / Intern

Financial Resources
RA $23/hour