University of Fort Hare
Together in Excellence

HEALTH RESEARCH ETHICS COMMITTEE

DATA MANAGEMENT COLLECTION AND STORAGE

1. DOCUMENT HISTORY

Date	Version No	Reason for revision
11 July 2018	1	Newly formulated
31 January 2022	2	Updated

2. PURPOSE OF THE SOP

To provide guidelines on the management and storage of data. Data management includes the design, collection, cleaning, and management of all participants and other information, observations, and measurements of the research project. Through efficient and effective data management risk data, errors can be addressed as data management should be considered during the planning, design, and finalization of the research protocol. It should be fully understood that data management addresses data collection and data capturing and includes the preparation of data, data analysis and publication, data archiving, and lastly, the destruction of the data:

1.1 Appropriate scientific data-gathering instruments should be used to provide plausible and reliable data
1.2 The recorded data should be durable and appropriately referenced by the researcher
1.3 The data must be retained for the period specified in the Ethics Research Policy of UFH, which is five years or as required by policy or legal frameworks.
1.4 Data reported in publications should be available for discussion and interrogation without breaching the confidentiality or anonymity of the respondent(s).

3. SCOPE

This document establishes the procedures to follow when initiating data management during a research project and the subsequent procedures to follow when data is stored and destroyed or data banking thereof.

4. RESPONSIBILITIES

All UFH HREC members, researchers in UFH issued with ethics approval, and Ethics Administrators should be aware of the procedure during the continuous review and recertification process.

5. PROCEDURES(S)

5.1 Identification and description of data

5.1.1 Identifying the data the researcher wants to work with is important as it will address questions such as what type of data will be collected, why it is needed, and how it will be used. Figure 1 shows the decision route for the identification of the data.

5.1.2 During the decision about the data that is required, the lifecycle of the data or lifespan of the data should also be taken into consideration.

5.1.3 Identification of the data will also include consideration for the types and formats in which the data should be obtained such as Numeric data, verbal transcripts etc.

- All questionnaires or questionnaire guides to be used should have a professional representation of UFH and a standard format
- All questionnaires or questionnaire guides should have been approved by the supervisor or research group
- The different types and formats of the data that will be obtained have to be considered in terms of the sample and the ability of the selected participants to complete the instruments with truthful and reliable data that captures the situation at that moment in time.
- The ability of the researcher to execute a particular research instrument should also be taken into consideration as the inability to correctly administer the instrument will lead to unethical data-capturing sessions, wasting respondents' time and resources.

5.1.4 Further consideration should be given to what the data will be needed for in particular who will need access to the data and what restrictions are there in terms of using the data.

- It should be made very clear in the informed consent form what the data will be used for, and the researcher should not go beyond this stipulation without further permission to do so.
- It should also be indicated who will be working with the data, and no free access to the data by anyone should be possible.
- Should the data come from a particular source, the time limitation on using the data should be adhered to if the data should be returned to the source.

5.1.5 Lastly in identifying the data, consideration should be given to the necessary permission to gather such data but also to consider who owns the data and if the data will be shared in the future with other researchers.

- Informed consent should be obtained from each and every participant who takes part in the research. The Policy on Research Ethics should be followed for all research projects where human participants are involved in projects of UFH Researchers.
- UFH HREC SOP 6 should be clearly followed in terms of informed consent procedures.
- The Policy on Research Ethics should be clearly followed in terms of consent where gatekeepers or an organisational structure involved in the research is concerned.

- Clear procedures should be followed in terms of the intellectual property (IP) resulting from research conducted with UFH funds or the use of its facilities.
- Any other concerns about the IP of the study should be taken up with the Directorate: Govan Mbeki Research and Development Centre (GMRDC).

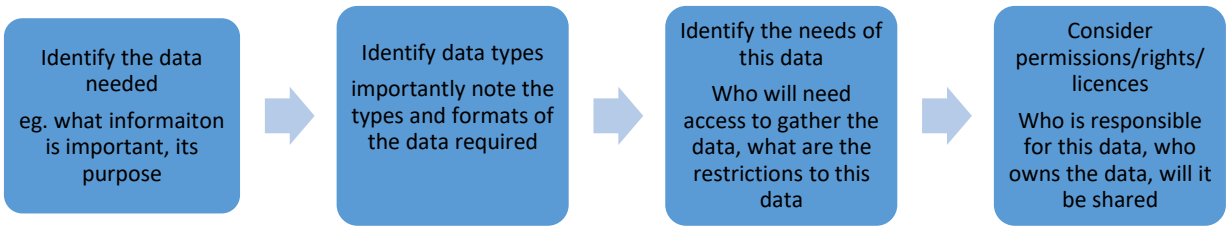

Figure 1: Identifying the data (Source: SOP-4.0.0-DM-121203 from Psycho-oncology Co-operative Research Group, The University of Sydney, Australia)

5.2 Identifying the mechanisms to capture the data

5.2.1 The data collection methods should be outlined in detail the methodology of the protocol. These methods should include a step-by-step procedure of how the data will be collected and who will do what during this process. For example, the fieldworker should distribute and collect the questionnaires from the participants. The researcher will then collect them on the 1st Monday of each month.

5.2.2 The procedures for each data collection instrument should be clearly stipulated.

5.2.3 If more than one instrument is used in different study phases, the same procedure should be followed as described in 7.1.

5.2.4 The sequence of data gathering and instrument completion should be clear to form a picture of how the data gathering will be executed in the study.

5.3 Outline the infrastructure and mechanisms to store the data

5.3.1 A researcher should have a clear set of guidelines regarding how variables for numeric data were coded, which is part of the data storage system applied by the researcher.

5.3.2 Other data storage systems used such as spreadsheets, text documents, specific computer driver space (local drive), etc. should be specified.

5.3.3 When mechanisms for data storage are considered the following questions should be asked:

- For physically and digital data storage
  - Will data storage be centralised or on-site where data is being gathered?
  - What is the timeline for data collection and storage?
  - How much data storage is needed?
  - How is the system secured (locked filing cabinets, password protected, etc)?
  - Has a log system been created for cataloguing the movement of data?
  - Where is the data stored?
  - In what format will the data be stored, and why is this chosen format?
  - Is any specific software required to read, analyse or process the data?

- Who is responsible for the data?
- What are your institution's data management policies?

5.4 Describe data security

5.4.1 Build and maintain a secure network system in which confidential passwords and documentation of an audit trail to capture changes to the information are clear.

5.4.2 Protect participant information through the de-identification of personal information and the use of participant pseudonyms or participant IDs where necessary.

5.4.3 Maintain vulnerability management programs which include anti-virus software and ensure that regular backups of the research data are made.

5.4.4 Implement a strong access control measure that restricts access and uses unique IDS for each person who has permission to access the data and criteria for electronic signatures.

5.4.5 Regularly monitor and test networks.

5.4.6 Specific security and data management procedures for informed consent

- Signed participant consent forms are to be kept with the researcher (student at UFH) or with the supervisors of the student. Should the student complete the study, the informed consent forms should be handed to the supervisor, if not already with the supervisor.
- The signed participant consent forms should be stored separately and securely from the de-identified data for five years.
- If a participant gives verbal consent, it should be recorded on the consent form by the researcher or fieldworker.
- If no consent was given, but data was collected, it should also be noted on the database
- The person responsible for managing the informed consent records should be identified.

5.4.7 De-identification of participant data

- During the de-identification process, the researcher should remove all identifying information from the data to protect the anonymity and confidentiality of the participants, which may be necessary when data is published or shared.
- This can be done by recoding names, and addresses or using any other identifiers.
- The researcher should retain enough information to confirm who the participant is if necessary.
- The use of Unique participant numbers issued at recruitment will allow the researcher to re-identify the data if needed
- The researcher should retain a master file of names and other identifiable data to be stored securely and separately from the study data in locked/password-protected databases with passwords kept separately.

5.4.8 Data sharing

The Policy of Research Ethics should be consulted when data sharing is considered.

5.5 Standardising data entry, checking and validation

5.5.1 Data entry should be specific in terms of how missing variables are to be coded, how logic checks will be run, and how the process for querying any inconsistencies will be dealt with.

5.5.2 Specific details should be available on who will update the data, and how regularly this will be done (for example, once a week when new data is received, it will be captured).

5.5.3 The batch of data that will be captured should be dated and initiated on the front of each questionnaire to indicate it has been captured.

5.5.4 Cleaning and validation of the data is an important part of quality assurance and control of the data. Checks should be run for the following:

- Invalid character values
- Invalid values for Likert scales
- Has a valid range been entered?
- Univariate statistics should be checked for continuous variables
- Repeated IDS.
- Check and validate the dates if required where an intervention took place.

5.6 Strategy for backing up data

5.6.1 The strategy for backing up data should be clearly stated such, as once a week or daily.

5.6.2 The researcher should ask how frequently data will be backed up on the systems or manually.

5.6.3 The researcher should also ask how disaster recovery will be dealt with should anything go wrong.

5.7 Auditing data

5.7.1 Audits may be completed to determine if the research is being conducted as specified in the protocol approved by the ethics committee.

5.7.2 It is valuable to include the frequency with which audits will take place.

5.8 Data analysis

5.8.1 The data cleaning decisions might influence the analysis and should be considered and checked.

5.8.2 Revision of missing values should be considered in numeric data to determine if a pattern of missing data can be identified.

5.8.3 Member checking should be considered in any qualitative study.

5.9 Archiving and destruction of data

5.9.1 Data should be stored for five years as per the Policy on Research Ethics at UFH.

5.9.2 Storage of data should be done so that it is easily retrieved.

5.9.3 Data should be kept de-identified, and consent forms should also be kept.

5.9.4 Preferably, the information should be boxed and labelled with the project title, the date on which the data was transferred to storage, the name of the researcher who was responsible for the data, and the date for destruction and number of boxes for this particular project.

5.9.5 When the data is destroyed it should be done in such a way that the information is completely destroyed.

5.9.6 Confidential data and records in paper format should be shredded.

5.9.7 Electronic format confidential data and records should be destroyed by reformatting or overwriting. A delete instruction is not sufficient to ensure that all system software has been destroyed.

5.9.8 Audio-visual tapes should be degaussed through a magnetic field bulk eraser

5.9.9 When destroying confidential data and records, the researcher should ensure that the most effective method is applied.

5.9.10 Data that may be permanently kept includes but is not limited to:

- Controversial or of high public interest.
- Would be costly or impossible to reproduce.
- Relates to the use of or supports the development of an innovative intervention.
- Supports a patent application or other services.
- Has long-term heritage, historical or cultural value.
- Is of significant value to another researcher.

6. REFERENCE DOCUMENTS

The UFH HREC adopts as its guiding reference the following documents t:

- SOP_4.0.0-DM-121203, Study documentation and data management, Psych-Oncology co-operative Research Group, The University of Sydney, Australia
- Ethics in Health Research: Principles, Processes and Structures, 2nd Ed, 2015
- UNISA Policy on Research Ethics, 2016.