



Data Management Plan for Research Students

Project title	Worked example project
Author name	Student name
Supervisor	Supervisor's name
Contact email	The student's university email
Date of last edit	20 Jan 2024

Guidance on writing a Data Management Plan can be found at
<https://lshtm.sharepoint.com/sites/intranet-library-archive-and-open-research-services/SitePages/Research-Data-Management.aspx>

and <http://servicedesk.lshtm.ac.uk>

Advice and feedback can be obtained from:

researchdatamanagement@lshtm.ac.uk

DESCRIBE YOUR RESEARCH

1. What digital resources – data, code, collection tools, etc. - will you collect/obtain and use?

Collection tools

I will create the following items as part of the research planning process:

1. A survey questionnaire to collect data from patients attending scheduled appointment. This will cover mental health when they attend clinic, treatment adherence, and information provided.
2. A question list to guide focus group discussion. This will cover topic 1, topic 2, topic 3.
3. A question list for one-to-one semi-structured interviews with policy-makers, health professionals and patients. This will cover topic 1, topic 2, topic 3.

Data

Using the above, I will collect the following data:

4. Quantitative survey data will be collected from each patient when they attend scheduled appointments. Approximately 300 patient responses will be recorded.
5. Audio recordings of approximately 50 semi-structured interviews with policy-makers, health professionals and patients. Audio will be recorded and stored in MP3 format.
6. Transcripts of the above semi-structured interviews will be stored in MS Word format.
7. Audio recordings of 10 focus group discussions conducted with patients and health workers. Audio will be held in MP3 format.
8. Transcripts of the focus group discussion will be stored in MS Word format.
9. Unstructured notes of document reviews will be recorded. Notes are based upon public information and will be used for internal use only.
10. In addition, a secondary data analysis will be performed upon anonymised survey data made available at [URL]

Code

I will also write the following code during the analysis process

11. STATA .do files necessary to harmonise, anonymise and process datasets?

2. What hardware and software will be used in your research?

The quantitative survey will be developed using LSHTM's Open Data Kit instance (<https://opendatakit.lshtm.ac.uk/>) and collected using Android-based tablets. Resulting data will be imported into STATA for processing and analysis

One-to-one interviews and focus groups will be recorded using a tablet or laptop (if allowed – see 16 for questions) with a USB microphone. The [software name] voice-to-text software will be used to create an initial transcript, which I will manually review and correct using MS Word. Transcripts will subsequently be imported into NVivo for coding and analysis.

3. What data-related activities will be performed during the research?

Task	Description
Survey data	
Develop survey	Investigate similar studies and develop a protocol for data collection, including a set of questions and permitted responses (resource 1)
Trial draft questionnaire	Trial with key stakeholders and obtain feedback. Revise accordingly.
Setup questionnaire on tablet	Design question form using Open Data Kit and trial it on tablet
Perform survey	Perform survey in the 2 designated hospitals and capture data using ODK (resource 4)
Prepare data for analysis	Clean, anonymise and process data for analysis using Stata. Harmonise data with third-party data (resource 10). Generate STATA code (resource 11)

Analyse data	Perform analysis outlined in research protocol
FGD recordings	
Develop interview guide	Investigate similar studies and develop a set of questions to guide discussion (resource 2)
Trial draft questions	Circulate to relevant stakeholders and obtain feedback. Revise accordingly.
Organise and host Focus groups	Organise and host 10 focus groups. Record discussion.
Process FGD audio recordings	The [software name] voice-to-text software will be used to create an initial transcript, which I will manually review using MS Word. In addition, I will anonymise and code each transcript
Analyse FGD transcripts	Perform analysis using NVivo
1-to-1 interviews	
Develop interview guide	Investigate similar studies and develop a set of questions to guide discussion (resource 2)
Trial draft questions	Circulate to relevant stakeholders and obtain feedback. Revise accordingly.
Organise and host Focus groups	Organise and host 10 focus groups. Record discussion.
Process FGD audio recordings	The [software name] voice-to-text software will be used to create an initial transcript, which I will manually review using MS Word.
Analyse FGD transcripts	Perform analysis using NVivo

4. What quality checks will you perform to ensure resources are fit for purpose?

Interview & focus group:

- Before: Review question list, test recording device to ensure setup correctly, ensure recording device is charged, bring backup device 'just in case', confirm participant(s) will attend interview
- During: Use checkpoints to check relevant details are provided, ask question in different ways to confirm answers, check device is still working and recording.
- Following: Check participant consent has been obtained, check question list has been addressed, check audio has recorded correctly.

Survey:

- Before: Train field workers on survey tools, ensure ODK form includes validation to flag required information, test recording device to ensure setup correctly, ensure recording device is charged, bring backup device 'just in case'.
- During: Ensure ODK form includes validation to flag required information
- Following: Check participant consent has been obtained

5. How will you address ethical & legal issues within your research?

I will obtain ethics approval from LSHTM, collaborating organisations, and country-specific ethics committees.

I will provide participants with information on the purpose of the research, how data will be managed and analysed, and plans for data sharing. I will discuss the process with them to collecting data and provide them with the opportunity to ask questions.

Participant identifiable information will be stored in the country in which it was collected. Data will be anonymised prior to analysis and sharing with others.

6. What documentation will be created to ensure resources can be understood?

A document will be written on the data collection workflow to be applied. This will provide details of the collection technique(s) applied (e.g. Computer-assisted Personal Interview), who will perform the task (if more than one person), as well as hardware and software to be used. This will be supported by collection forms that outlines questions to be raised and expected/permitted responses. This document will be written prior to the start of collection and updated/expanded over time.

Survey data will be supported by a codebook necessary to understand its content. This will provide details on the response type for each question (i.e. free text, controlled list), permitted values and their meaning (e.g. 0 = no, 1 = yes), measurements used, the meaning of abbreviations and acronyms, processing actions performed (e.g. as a result of cleaning and redaction), weighting applied.

Interview and focus group transcriptions will be supplemented by documentation that describes issues encountered and decisions made when translating content to a second language, labelling applied for redactions and other changes, the coding scheme developed and used to structure content for analysis, and non-verbal responses (e.g. tone of voice, physical behaviour).

STORAGE AND SECURITY

7. Where will resources be stored at key stages of your research?

Data Classification

Resources will be classified as follows under the LSHTM Data Classification and Handling Policy

- *Confidential*: interview and focus group discussion recordings (resource 5 and 7). The first 'raw' version of survey data (resource 4) and interview FGD transcripts (resource 6 and 8) will be Confidential, but derivatives may be classed as Internal if they can be sufficiently de-identified.
- *Internal*: De-identified survey datasets (resource 4) and interview transcripts (resource 6 and 8), subject to the above de-identification process. If transcripts cannot be de-identified, interview summary notes will be created.
- *Public*: Collection tools (resource 1-3) and document review notes (resource 9) are based upon public sources and can be made available accordingly. The anonymised survey data (resource 10) is currently available publicly.

Storage location

Survey data (resource 4) will be captured via the ODK survey tool and temporarily held on Android tablets deployed for use in the field. If an internet connection is available, captured information will be securely transferred to the Open Data Kit server operated by LSHTM. Following completion of survey work, the complete dataset will be exported and transferred to the LSHTM secure server for processing.

Audio recordings of interviews and focus group discussions (resource 5 and 7) will be captured on encrypted Android tablets. Following completion, they will be moved to LSHTM's secure storage facility for processing. This system uses user authentication to limit access to specific users only.

Audio recordings will be encrypted and sent to Organisation Y for transcription and translation into English. In advance of this activity, the organisation will be asked to sign a contract agreement, which outlines various conditions of use, e.g. they will store the data securely, will not share it with others, and will delete it once work has been completed. Following completion, transcripts (resource 6 and 8) will be encrypted and transferred to LSHTM for processing. These will be stored on the institution's secure server throughout the time they are held.

Anonymised survey data (resource 10) and document review notes (output 9) will be stored in the storage area allocated to the researcher by their university. These will be shared with my supervisor and other advisors.

8. What labelling conventions will you apply to manage your resources?

Filenames will be labelled using key characteristics that will allow relevant files to be located quickly, while protecting participants from identification. For instance, collectionMethod_IDnumber_resourceType_versionNo.fileformat. E.g. FG1_01_audio_v1.wav, FG_01_transcript_v1.1.docx.

Questions and recorded responses will be labelled using an identifier (e.g. q03exp) to enable link-up between the two.

9. How will you keep data safe and secure? (choose one or more)

Only anonymised data will be used - personal, sensitive, or otherwise confidential data is not needed for the research		Store personal details in a separate secure location & link it via an identifier	✓	Delete personal & confidential details at earliest opportunity (specify when below)	✓
Use digital storage that require a username/password or other security feature	✓	Physical security (such as locked cabinet or room)		Protect portable devices using security features, e.g. biometric	✓
Encrypt storage devices	✓	Encrypt during transfer	✓	Avoid cloud services located outside the relevant jurisdiction (e.g. GDPR compliant area)	✓
Take 'Information Security Awareness training'	✓	Ensure backups are also held securely	✓		
Notes:	Participant identifiable information will be deleted on completion of the analysis.				
Identify additional steps you will take to avoid, reduce, or eliminate risks that may affect your resources.					

ARCHIVING & SHARING

10. What resources should be kept as evidence of your research?

I will keep the following resources for 10 years:

- Collection tools (resource 1-3)
- Anonymised survey data (resource 4)
- Anonymised interview transcripts (resource 6)
- Anonymised focus group transcripts (11)
- Stata .do files (resource 11)

The following resources will be deleted when no longer needed (most likely following allocation of grade)

- Participant identifiable details captured during survey (resource 4) – deleted in compliance with GDPR
- Audio recordings of interviews (5) and focus groups (7) – deleted in compliance with GDPR
- Document review notes (9) – notes will have been incorporated into thesis and no longer required.
- Anonymised survey data (resource 10) – deleted in accordance with data provider agreement and GDPR. Primary copy will continue to be kept by XYZ university

11. Where will these resources be hosted?

The anonymised survey data, interview transcripts and FG transcripts will be uploaded to the university research data repository and made available for 10 years, in accordance with the institutional research data management policy. If transcripts cannot be de-identified, they will be stored on the institutional secure server and interview summary notes will be created and uploaded to the repository in their place.

12. When will the resources be made available? (choose one or more)

During the research life		At the same time as findings are published in an academic journal	✓	A set time after research end, e.g. 12 months. Specify below	
Resources already available (provide details below)		On completion of my thesis	✓	Other (provide details below)	
Further information / Other					

13. How will you make other researchers aware that the resources exist?

Publish a metadata record describing the resources in a repository or other catalogue	✓	Obtain a Digital Object Identifier (DOI) or other permanent ID	✓
Cite resources in future research papers, e.g. in the data access statement or reference list	✓	Cite resources in project reports	✓
Publish a description for the project website		Write and publish a Data Paper	
Add resources to a list of your academic outputs	✓		
Other measures / Further details			

14. What steps will you take to ensure resources are easy to analyse and use in future research?

Prepare a codebook or other documentation that provides an accurate description of content	✓	Store resources in open file formats such as CSV, Rich Text, etc. See https://ukdataservice.ac.uk/learning-hub/research-data-management/format-your-data/recommended-formats/	✓
Write a user guide that provides a high-level overview of research	✓	Apply a standard licence that allows a broad range of uses (e.g. Creative Commons, Open Data Commons)	✓
Designate a corresponding author / data custodian who will handle data-related questions	✓	Use domain-specific standards that make it easy to import and analyse data	
Other / Further information			

15. If resources can be made available, but not openly, what conditions on access/use must be met?

E.g. data can be used for specific types of research only. Leave blank if not applicable.

Requirement:	To be addressed by:
Interview/Focus group transcripts: It may not be possible to remove all personal details and participants may be identifiable as a result.	The following steps will be taken to protect participants being identified: [1] Data will be restricted and shared only with people that have ethical approval for their research. [2] Eligible researchers will be asked to sign a licence agreement indicating they will keep data secure, use it for the specified purpose only, delete it after use, not release it to others, etc.

RESOURCING

16. What are the primary data management challenges in your research?

I'm unsure about the data collection process as a whole and would appreciate advice on the following:

- Can I use a tablet to record the audio interviews and focus group discussions?
- Is there a charge for borrowing Android tablets from the university?
- How can I check that the tablet is encrypted? Can IT Services do this for me?
- How do I securely transfer audio recordings to a university server?

17. How can LSHTM & others help you to better manage your data?

- Advice on any tablet lending schemes available to students
- IT support on how to encrypt a tablet. Ideally, I'd also like them to check that it was performed correctly.
- IT support on how I can transfer audio recordings securely using university systems.