LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Writing a
MRC / NC3R
Data Management Plan

Report Version Control

Version	Date	Author	Change Description
1.0	10 / 05 / 2015	Gareth Knight	First draft using University of Bristol & University of Leicester DMP guide as a basis.
1.1.	16 / 07 / 2015	Gareth Knight	First version
1.2.	17 / 07 / 2015	Simon Jennings	Error checking and minor amendments
1.3.	21 / 07 / 2015	Gareth Knight	First published version

Executive Summary

The Medical Research Council expects funded researchers to manage and share research data in a manner that maximizes opportunities for future research and complies with best practice in the relevant subject domain.

This document describes MRC requirements for data management and sharing, offers suggestions on how these obligations can be met by LSHTM researchers, and provides worked examples that may be used as a basis to develop your own Data Management Plan.

Applicability to other funders

The Medical Research Council's approach to data management and sharing follows recommendations established by Research Council UK. As a result, many of its recommendations will be familiar to applicants to other funders.

Some funders, such as the NC3R[1] (National Centre for the Replacement, Refinement & Reduction of Animals in Research) require funders to complete an MRC Data Management Plan and follow their recommendations. As a result, this guide may be helpful in completing applications to other funders.

Key Expectations for Data Management & Sharing

The Medical Research Council has published a considerable amount of information on their expectations for data management and sharing. Key messages that they communicate include the following:

- All MRC-funded research must possess a Data Management Plan (DMP):

  - Applicants must complete a DMP and submit it as an attachment (labeled "Data Management Plan" in their Je-S Proposal Form (https://je-s.rcuk.ac.uk/)

  - MRC Institutes and Units must complete a DMP and submit it as part of the Quinquennial Review (QQR) report. The Institute/Unit Director may develop a DMP for each programme if they wish.

  - The MRC encourages applicants to write for two audiences:
    1. Scientists within your broad subject domain
    2. Technical experts with data management expertise

- Data management and sharing requirements will vary, depending upon the funding call and type of research being performed:

  - Population and patient based studies must meet additional requirements for data retention[2] and access provision[3]. Established population/patient studies must already possess a Data Sharing policy which can be referenced in the DMP.

  - Joint funded research (i.e. that supported by the MRC and another funder) may impose additional requirements.

- Applicants are encouraged to claim costs for resources necessary to manage and share data; these should be clearly described and justified. See LSHTM Costing guidance for further details[4].

---

[1] http://www.nc3rs.org.uk/sites/default/files/documents/Funding/Handbook.pdf (p.28)
[2] http://www.mrc.ac.uk/research/research-policy-ethics/data-sharing/data-management-plans/
[3] http://www.mrc.ac.uk/research/research-policy-ethics/data-sharing/data-sharing-population-and-patient-studies/
[4] http://www.lshtm.ac.uk/research/researchdataman/plan/rdm_costs.html

- Research data should be made available in a form that can be shared and reused by others. If this is not feasible (e.g. due to legal, ethical, contractual, or other obligations), restrictions should be clearly described and justified.

- It is recognized that data management and sharing practices vary across subject domains and disciplines:

    o Researchers are encouraged to deposit data with a discipline-specific data repository, where feasible. If one does not exist, institutional or general repositories may be used.

    o Data formats should be appropriate for the content and accessible to the target audience.

- A limited period of exclusive use of data is reasonable to enable analysis and publication. However, you will need to state a specific time period when the data will be made available for access.

- The MRC encourage applicants to consult the following resources for information on appropriate standards:
    o UK Data Service
      < http://ukdataservice.ac.uk/manage-data.aspx >

    o Inter-university Consortium for Political and Social Research
      < http://www.icpsr.umich.edu/icpsrweb/datamanagement/ >

    o Australian National Data Service
      < http://www.ands.org.au/datamanagement/ >

    o National Statistics Code of Practice
      < http://www.ons.gov.uk/ons/guide-method/the-national-statistics-standard/code-of-practice/ >

Further details on MRC data management and sharing requirements can be found at
http://www.lshtm.ac.uk/research/researchdataman/plan/funder_mrc.html

Costing Data Management

Consideration should be given to the costs associated creating, managing and sharing research data. Traditionally, researchers have been wary of stating these costs in a research bid, believing that it will make their application appear uncompetitive. However, funders increasingly view the lack of data management costs as an area for concern, possibly signifying that the applicant has not considered the practicalities of their project.

In an RCUK briefing on data management and sharing costs[5], the MRC offered the following advice on resource allocation:

1. *Resources should be clearly described and justified in research applications:* Sufficient information should be provided to help the evaluator to identify resources allocated for data management, the reason that it is needed and the associated costs.

2. *Institutional services must not be claimed twice:* Institutional resources that are already claimed in indirect costs must not be included in direct costs.

3. *Post-project costs must be spent before project end:* Costs for long-term curation and preservation, such as POSF (Pay Once, Store Forever) storage, must be allocated before the close of the project grant. The MRC will not support unexpected costs incurred after the project end.

Specific RDM resources that the MRC encourage applicants to consider include:

---

[5] http://blogs.rcuk.ac.uk/2013/07/09/supporting-research-data-management-costs-through-grant-funding/

- *People and Skills:* Do project staff have the necessary time and expertise to prepare data for sharing? If not, do additional specialist staff need to be assigned to the project, e.g. to create metadata for the dataset.

- *Training:* Do staff need paid-for training to perform data management activities?

- *Technical Infrastructure:* Does the project have access to systems capable of storing, managing and sharing data? An overview of LSHTM storage systems can be found at http://www.lshtm.ac.uk/research/researchdataman/store/ and http://www.lshtm.ac.uk/its/staffservices/.

Further Information
- RCUK: Supporting research data management costs through grant funding
  < http://blogs.rcuk.ac.uk/2013/07/09/supporting-research-data-management-costs-through-grant-funding/>
- UK Data Service: Plan to Share
  < http://ukdataservice.ac.uk/manage-data/plan.aspx >

Format of a Data Management Plan

The Medical Research Council provides a structured Data Management Plan template[6] that they encourage applicants to complete and submit. This template consists of eight sections, the majority of which include two or more sub-sections:

1. Description of the data
    a. Type of study
    b. Types of data
    c. Format and scale of the data

2. Data collection / generation
    a. Methodologies for data collection / generation
    b. Data quality and standards

3. Data management, documentation and curation
    a. Managing, storing and curating data.
    b. Metadata standards and data documentation
    c. Data preservation strategy and standards

4. Data security and confidentiality of potentially disclosive information
    a. Formal information/data security standards
    b. Main risks to data security

5. Data sharing and access
    a. Suitability for sharing
    b. Discovery by potential users of the research data
    c. Governance of access
    d. The study team's exclusive use of the data
    e. Regulation of responsibilities of users

6. Responsibilities

7. Relevant institutional, departmental or study policies on data sharing and data security

8. Author of this Data Management Plan

Data Management Plans should be proportionate to the scale of the project: a simple project may require a DMP of just 500 words, whereas a complex project may require 1000-1500 words (3 pages). Longitudinal studies that involve several data collection activities may be longer, where appropriate.

You are encouraged to complete the provided template. However, the MRC will accept a DMP in a different format (e.g. as free text) if it addresses all of the requested areas.

The Data Management Plan will be reviewed as an integral part of the application; a poorly written plan can have a negative impact on an otherwise strong application.

Advice and Guidance

For tailored advice and guidance, please contact the LSHTM RDM Service:
- Email: researchdatamanagement@lshtm.ac.uk
- Web: http://www.lshtm.ac.uk/research/researchdataman/rdm_service.html

---

[6] http://www.mrc.ac.uk/documents/doc/data-management-plan-template/

1. Description of the Data

1.1 Type of study

Write a brief description of the study's purpose in 1-3 lines of text. This may describe the topic that will be investigated, the objective that you wish to achieve, or the hypothesis that will be tested. Your description should be sufficiently accurate for a layperson to understand, without having a detailed knowledge of the underlying science.

> **Example:**
>
> A mixed method study that will investigate mobility practices of UK-based men and women during later life in relation to their immediate social environment.

1.2 Types of data

Describe each of the data outputs with which you will be working, covering primary data to be created, and secondary data obtained from elsewhere. Your description should focus upon the information content, rather than the technical details of how it is stored. Key information to provide include:

- *Type of information:* Interviews, fieldwork diaries, 3D models, lab experiments
- *Form in which it will be held:* audio recordings, still images, videos, software etc.
- *Creation method:* observation, experimentation, or simulation

It is advisable to write between 3-6 lines of text, depending upon the number and complexity of data outputs. If you are working with several outputs, you may find it useful to assign a number to each, in order to track their development in later sections of the Data Management Plan.

> **Example:**
>
> The following qualitative data will be created while performing ethnographic research at three sites - Bristol, York, and London:
>
> 1. Audio recordings and transcripts of formal/informal interviews and informal discussions. Approximately 175-200 individual interviews and 10-15 focus groups will be performed;
> 2. Hand-written fieldwork diaries and memos produced while 'shadowing' individuals and participating in group activities;
> 3. Photographs of locations and other documentary data;
>
> The following quantitative data will be obtained and analysed:
>
> 4. The six available waves of the English Longitudinal Study of Ageing (ELSA) will be downloaded from the UK Data Service and integrated using the MPlus software. All derived variables from our analysis (dimension of social capital, trajectories of physical activity) will be deposited with the UKDS following project completion.

Further Information
- LSHTM: Introduction to Research Data
  < http://www.lshtm.ac.uk/research/researchdataman/introduction/researchdata_intro.html >
- LSHTM: Find Data to Reuse
  < http://www.lshtm.ac.uk/research/researchdataman/create-organise/find_data.html >

1.3   Format and scale of the data

Describe the technical composition of each of the data outputs listed in 1.2. Questions that you should consider for this section include:

- What software tools will be used at each research stage? Do these tools encourage/require you to use a specific file format?
- Will you require specific hardware, such as a data capture device to record interviews? If so, what formats are supported by this device?
- What formats and standards are considered best practice within your research field?

The MRC do not expect you to use the same format throughout; data can be converted into different formats as the need arises, on the proviso that no information is lost during the export process.

You are encouraged to apply community standards and open formats where feasible. If you must use proprietary or little-known technologies, the reason that these have been chosen should be established.

Example:

Audio recordings of each interview and focus group (approximately 215 in total) will be captured as MP3 using a digital Dictaphone. Each will be transcribed using Transcriber Pro and stored as ASCII plain text (output 1). Fieldwork diaries and memos sites will be selectively transcribed by the fieldworker into ASCII plain text (output 2). Photographs will be held in JPEG format - the only format supported by the camera (output 3).

Further Information
- LSHTM: Choosing File Formats
  < http://www.lshtm.ac.uk/research/researchdataman/create-organise/fileformats.html >
- UK Data Service: Recommended File Formats
  < http://ukdataservice.ac.uk/manage-data/format/recommended-formats.aspx >

2. Data collection / generation

The second section focuses upon the process of acquiring data for use within the study. The MRC state the approach should build upon evidence provided in your Case for Support. Therefore, you should ensure that the following information is provided:

- *If you are using an existing dataset:* the CfS should identify the datasets to be used and how access will be obtained.
- *If you are producing new data:* the CfS should explain the reason that new data creation is necessary, e.g. your research covers a topic that has received little attention.

2.1 Methodologies for data collection

Describe the method(s) that will be applied to create capture, or obtain each data output in your study. Key questions to consider include:

- Who are your study participants and where are they located?
- Who will provide the data and how will access by obtained?
- How will you collect data and what methods will be applied?

Example:

The project team will organise individual interviews, focus group meetings, and informal 'shadowing' of identified participants in each location. We will cover Bristol in months 4-5, York in months 6-7, and London in months 8-9. An audio recording will be captured for each interview/meeting (output 1), which will be subsequently transferred to a professional transcriber in month 10.

Fieldwork diaries and memos (output 2) will be completed by the fieldworkers during the process of 'shadowing' individuals and participating in group activities. These will be reviewed and selectively transcribed by the fieldworker into ASCII plain text. Similarly, digital photographs (output 3) captured during the process will be appraised and kept if they are considered relevant to the study objectives.

The ELSA datasets, which covers 11,000+ individuals, will be downloaded from the UK Data Service, merged and analysed using MPlus (https://www.statmodel.com/). The derived dataset will be exported into formats suitable for re-deposit with the UK Data Service, most likely CSV or tab-delimited text (output 4).

Further Information
- LSHTM: Create and Organise Data
  < http://www.lshtm.ac.uk/research/researchdataman/create-organise/index.html >

2.2 Data quality and standards

Consider how you will ensure your data will be suitable for your research needs. As a starting point, you should consider the following questions:

1. How will you ensure the correct information is captured and stored in a form that is easy to analyse?
2. What problems may occur during your research and how could you avoid them?
3. Are there any standards or practices within your research field that should be applied?

It's advisable to perform regular validation checks, to ensure your data is fit for purpose and that no unexpected problems have arisen. Recommended activities are outlined in the following table.

Prior to capture	During capture	Following capture
Review methods & protocols used by similar projects and adopt common standards	Define checkpoints to verify that you are capturing all required information	Double-entry of information (if transcribing from a paper source)
Test instruments to confirm they are working	Record multiple measures	Review data to confirm that all information has been acquired
Check instrument calibration to confirm it is setup correctly	Document research workflow	Check validity – are values in range? Are they accurate and consistent

Example:

A data collection protocol will be developed prior to fieldwork taking place. This will outline questions to be raised in interviews and focus groups, as well as guidelines on the recording of fieldwork diaries/memos and type of photographs that should be taken. Fieldworkers will receive training on the data collection protocol from the Principal and Co-Investigator prior to entering the field. Data outputs recorded during fieldwork will be reviewed by the fieldworker at the end of each day and by the Principal Investigator at the end of each week. Any errors will be addressed, e.g. by contacting interviewees with additional questions.

The ELSA dataset possesses extensive documentation. The Co-Investigator will contact the data creators in the event that aspects of the data are unclear. The Co-I will also document methods applied to generate new variables, describe and annotating each in accordance with the same standards applied to the source data.

Further Information

- LSHTM: Create and Organise Data
  < http://www.lshtm.ac.uk/research/researchdataman/create-organise >
- LSHTM: Standard Operating Procedures
  < https://intra.lshtm.ac.uk/trials/sops/index.html >
- UK Data Service: Quality Assurance
  < http://ukdataservice.ac.uk/manage-data/format/quality.aspx >

3. Data management, documentation and curation

The third section examines the approach that will be taken to store and maintain data during the project lifetime and following its completion.

3.1 Managing, storing and curating data.

Describe the approach you will take to store research data during the project lifetime. Data should be held in a managed environment operated by LSHTM or a partner institution, where feasible. If you are working in the field and unable to access institutional systems, you will need to apply measures to manage your own data. Key questions that should be considered are:

1. *Where will you store data during the project lifetime?* LSHTM researchers working within the university have several storage systems available. An overview of these can be found in the LSHTM Storage Guide at http://www.lshtm.ac.uk/research/researchdataman/store/lshtm_storage_systems.pdf

2. *How many copies of the data will be kept?* If you are managing your own storage, you should store data on at least 3 different storage media, one of which is in a different location. LSHTM storage systems are automatically replicated onto multiple storage media.

3. *Who will be responsible for performing backups*? If you are working with project partners or managing data in the field, you should establish who is responsible for performing backups. LSHTM storage systems are maintained by IT Services.

4. *How often will data backup be performed?* Backups are only useful if they contain a recent version of your data. LSHTM storage systems are automatically backed up each night.

5. *How will you keep data secure?* Personal, confidential and sensitive data must be kept securely to prevent accidental data release. Consult LSHTM guidance on keeping data secure at http://www.lshtm.ac.uk/research/researchdataman/store/.

Example:

Audio recordings and photographs captured in the field will be moved to the Data Collector's laptop and uploaded to their personal area on the LSHTM network each night using Filr (http://www.lshtm.ac.uk/its/staffservices/filr/). This system is maintained in accordance with LSHTM's Information Management & Security Policy, backed-up and virus scanned on a daily basis. Audio recordings will be kept until transcription has been completed and verified, after which they will be deleted. The laptop's hard disk will be encrypted, to protect against accidental release in the event that the disk is stolen.

Further Information
- LSHTM: Storage Guide
  < http://www.lshtm.ac.uk/research/researchdataman/store/lshtm_storage_systems.pdf >
- LSHTM: Keep Data Secure
  < http://www.lshtm.ac.uk/research/researchdataman/store/ >
- UK Data Service: Store your Data
  < http://ukdataservice.ac.uk/manage-data/store.aspx >

3.2 Metadata standards and data documentation

Documentation provides contextual information that is essential to understand and reproduce the research. Consider the documentation that should be created to describe your research. Key information that may need to be documented include the:

- *Research Context:* A description of the aims, objectives, hypotheses, no. of subjects and other pertinent information on the research project.

- *Methods:* A description of the methods applied to collect, analyse and use data. This may include an explanation of collection protocols, sample design, instruments used, etc.

- *Data provenance*: An outline of where you obtained data and how.

- *Quality controls:* An explanation of validation techniques applied to calibrate equipment and check results.

- *Content:* Define variables, acronyms, units of measurement, and other pertinent information on the dataset.

- Consent forms: Information on the permissions assigned by study participants.

It is helpful to consider how someone with only limited understanding of the research – a third party or yourself in 10 years – will attempt to interpret your data. What type of questions would they ask? If there is insufficient documentation, they may misinterpret the information, or locate an alternative dataset that is easier to analyse.

Metadata as a type of structured documentation

Metadata is a type of structured documentation that is routinely created at different stages of the data management process. Many software products will automatically capture metadata, or provide you with the option to specify metadata attributes, such as the creator's name, software tools used, creation date, and machine configuration options. It may be stored within the dataset, or within a separate file such as a spreadsheet.

To aid data interoperability, you are strongly encouraged to use community standards to describe and structure data. Examples include:

- *International Statistical Classification of Diseases and Related Health[7] (ICD-1012)*: A medical classification from the World Health Organisation that "codes for diseases, signs and symptoms, abnormal findings, complaints, social circumstances, and external causes of injury or diseases".

- *Data Documentation Initiative[8] (DDI):* an international standard for describing data from the social, behavioural and economic sciences.

- *Standard Operating Procedures:* Clinical researchers often apply institution and domain specific SOPs that outline specific processes for data management. LSHTM's Clinical Trial SOPs can be found on the School intranet (https://intra.lshtm.ac.uk/trials/sops/).

The Digital Curation Centre maintains a catalogue of disciplinary metadata standards (see 'Further Information'), which may be useful.

---

[7] http://www.who.int/classifications/icd/en/
[8] http://www.ddialliance.org/

> **Example:**
>
> The project will follow UK Data Service guidelines for describing quantitative and qualitative data (http://ukdataservice.ac.uk/manage-data/document/data-level). In addition, a user guide/data paper will be created, containing information suitable for use by a lay person.
>
> Documentation will be developed carefully to ensure it provides sufficient information to understand the content, without providing information that may be used to re-identify individuals. Each output will be assigned a unique identifier to enable data linking across multiple outputs.
>
> Names, labels and descriptions for all variables, fields, records and their values contained within tabular datasets will be recorded in a data dictionary. Processing activities performed in STATA will be captured as a set of do files, which will be commented throughout in order to provide a clear record of its purpose.
>
> Interview transcripts will be labelled with the interviewee ID and other relevant information, such as age range, gender, and interview date. Each file will be labelled with the filename and version number to minimize confusion.

Further Information
- LSHTM: Documenting your Data
  < http://www.lshtm.ac.uk/research/researchdataman/describe/ >
- LSHTM: Standard Operating Procedures
  < https://intra.lshtm.ac.uk/trials/sops/ >
- UK Data Service: Document your Data
  < http://ukdataservice.ac.uk/manage-data/document.aspx >
- Digital Curation Centre: Disciplinary Metadata
  < http://www.dcc.ac.uk/resources/metadata-standards >

3.3 Data preservation strategy and standards

Research data often has value beyond the project lifetime. Describe the approach that will be taken to maintain data in the long-term, beyond the project lifetime. As a starting point, you should consider the following questions:

1. *What data is necessary to verify and reproduce your research findings? What could be used in future research?*
   A research study is likely to produce a large amount of data, only some of which may be important. It's advisable to apprise your data in the final months of your study and decide what should be kept.

2. *How long must / should data be kept?*
   The LSHTM Records Retention and Disposal Schedule[9] and MRC guidelines state that data must be kept for a minimum of 10 years following project completion. Data that has been produced in a MRC-funded population and patient based study will need to be retained for at least 20 years.

3. *How will access to the information content be maintained across changing technology?*
   Data should be stored in an open, well-documented file format that is supported by a wide range of software tools, wherever possible. The use of text-based formats, such as CSV and tab-delimited, is preferable to binary formats, such as Microsoft Access. Consult the URLs in 'Further Information' for recommendations.

4. *Who will look after your data?*
   Finally, you should consider how data will be managed in the long-term. Services such as the UK Data Service (http://ukdataservice.ac.uk/) and UK BioBank (http://www.ukbiobank.ac.uk/) may be willing to

---

[9] https://intra.lshtm.ac.uk/informationmanagement/records/retention.html

accept and maintain your data. If there are no designated data curators in your research field, data may be deposited with the LSHTM RDM Service for safe-keeping.

Example

Primary data will be deposited with the UK Data Service in a format considered appropriate for long-term preservation, following recommendations outlined at http://ukdataservice.ac.uk/manage-data/format/recommended-formats.aspx. Tabular data will be deposited in STATA format and tab-delimited text, and qualitative data (such as interview transcripts) will be deposited in RTF format. Any enhancements made to the ELSA datasets, originally obtained from the UK Data Service, will also be provided at the same time.

In accordance with guidelines outlined in the LSHTM Records Retention and Disposal Schedule and MRC requirements, anonymized data will be kept for a minimum of 10 years following project completion. However, there are no barriers that prevent it being retained in perpetuity.

Further Information
- LSHTM: Determine retention requirements for research data
  < http://www.lshtm.ac.uk/research/researchdataman/curate-preserve/research_retention.html >
- LSHTM: Convert data to an appropriate file format
  < http://www.lshtm.ac.uk/research/researchdataman/depositdata/fileformat.html >
- Digital Curation Centre: How to Appraise and Select Research Data for Curation
  < http://www.dcc.ac.uk/resources/how-guides/appraise-select-data >
- UK Data Service: Recommended Formats
  < http://ukdataservice.ac.uk/manage-data/format/recommended-formats.aspx >

4. Data security and confidentiality of potentially disclosive information

The 4th section explores the approach that will be taken to safeguard data from unauthorised access. Any information outlined in this section must correspond with that produced as part of your ethics review.

4.1 Formal information/data security standards

First, identify any formal information standards that will be followed by your study. As a minimum, LSHTM researchers are expected to follow the LSHTM 'Information Management & Security Policy', which sets out requirements for the storage of personal, confidential and sensitive data.

For collaborative research, you will need to identify the security standards and policies implemented by the lead institution and that which will be storing data. Many LSHTM research projects use UCL's ISO 27001 storage system to hold their data, for example. If you are using an ISO compliant system, you should state the institutional registration number in your Data Management Plan.

If you intend to use a 3rd party storage provider (e.g. a cloud service), you should ensure that they conform to relevant standards in the health sector, such as storage within the European Economic Area (EEA). IT Services can provide support and guidance on this topic.

Example:

Study team members will be trained on the principles of data security covered in Good Clinical Practice and will follow the LSHTM Information Management and Security Policy at all times.

Further Information
- LSHTM: Information Management & Security Policy
  < http://www.lshtm.ac.uk/its/informationsecurity/policy/ >
- LSHTM: Keep Data Secure
  < http://www.lshtm.ac.uk/research/researchdataman/store/ >
- Digital Curation Centre: Information Security Management- The ISO 27000 (ISO 27K) Series
  < http://www.dcc.ac.uk/resources/briefing-papers/standards-watch-papers/information-security-management-iso-27000-iso-27k-s >
- UK Data Service
  < http://ukdataservice.ac.uk/manage-data/store/security.aspx >

4.2 Main risks to data security

Health research is a highly regulated field that requires researchers to implement a range of practices to ensure data is stored securely, particularly if it has been collected from human participants. This question encourages you to take a 'worst-case scenario' approach, considering potential problems that could occur and how you will avoid these, or at least minimize the likelihood that they will happen.

1. *What potential problems may arise?*

   Consider the events that will have a negative impact upon your research, such as accidental data loss and unauthorized release of personal information. The latter is particularly serious, since it would breach Data Protection legislation and potentially result in reputation damage, financial repercussions, and legal action.

2. *What is the likelihood that the problems will occur?*

   Risk assessment is inherently speculative. Many methodologies exist that can help you to determine the likelihood that a problem will occur. The RDM Service has developed a risk assessment template that may be used as a starting point (see 'Determine data storage needs and risks').

3. *How will you avoid each problem or reduce the likelihood that it will take place?*

Your approach to risk management should take into account relevant policy decisions and technical barriers that will be implemented. A strategy for reducing the likelihood of accidental disclosure may include the following activities:

- Collect anonymous information only, or remove identifiable information at the earliest opportunity
- Store data within the European Economic Area (EEA);
- Apply access controls (username and passwords) to limit data to authorized users only
- Encrypt data using an appropriate algorithm
- Adopt a Data Transfer Agreement that prohibits re-identification of study participants

---

Example:

There is a risk that data containing personal information may be accessed by unauthorized individuals and study participants are identified, either directly or through linkage to other data. To minimize risk, identifiable information (names, addresses, etc.) will be replaced with internal identifiers at the earliest opportunity. Personal information that must be kept will be held on a secure, isolated machine, separate from the research data.

Devices used for data collection in the field, such as laptops, will be encrypted to protect data in the event that the device is lost or stolen. In addition, data capture procedures will state that photographs should only be taken of the environment in which investigators are working, or study participants who are willing to be identified.

All paper consent forms that contain personal information will be stored in a secure, locked cabinet within LSHTM.

Data will only be made available in an anonymized form. If there are concerns that it cannot be fully anonymized, it will be made available on condition that end users apply for access and sign a Data Transfer Agreement indicating that they will not share the data or attempt to re-identify individuals.

Further Information
- LSHTM: Information Management & Security Policy
  < http://www.lshtm.ac.uk/its/informationsecurity/policy/ >
- LSHTM: Keep Data Secure
  < http://www.lshtm.ac.uk/research/researchdataman/store/ >
- LSHTM: Determine data storage needs and risks
  < http://www.lshtm.ac.uk/research/researchdataman/store/storage_intro.html >
- MRC: Personal Information in Medical Research
  < http://www.mrc.ac.uk/documents/pdf/personal-information-in-medical-research/ >
- MRC: Human Tissue and Biological Samples for Use in Research: Operational and Ethical Guidelines
  < http://www.mrc.ac.uk/news-events/publications/human-tissue-and-biological-samples-for-use-in-research/ >
- MRC: Information Security Policy
  < http://www.mrc.ac.uk/documents/pdf/mrc-information-security-policy/ >
- UK Data Service
  < http://ukdataservice.ac.uk/manage-data/store/security.aspx >

5. Data sharing and access

The 5th section focuses upon the process of making data available for access and re-use.

LSHTM policy on data sharing

The LSHTM Research Data Management Policy states that research data which substantiate research findings should be made available for access and use in a timely manner, within the boundaries of conditions established by contractual, legislative, ethical, or other requirements.

MRC policy on data sharing

The MRC encourage applicants to make research data available for access and use in a "timely and responsible manner", but recognize that not all data is suitable for sharing. Population and patient studies must comply with additional guidelines for sharing research data[10].

5.1 Suitability for sharing

Review the data outputs listed in 1.2 and consider which, if any, can be made available for access and use. As a starting point consider the following questions:

1. What data produced and/or used in your project is necessary to validate your research findings?
2. Do any barriers exist that prevent or limit opportunities for data sharing?
3. If so, what action can be taken to resolve these issues / what conditions need to be in place to enable data to be shared?

The MRC places considerable emphasis upon the sharing of data generated by a project. However, they recognize that there are issues surrounding the sharing of data on human participants (even it has been anonymized) that cannot be easily addressed. If it is not possible to make data available under any circumstances, you should clearly describe the reason for this. For example, data may contain commercially sensitive information, be impossible to anonymise without removing its context, or be used as part of a patent application.

The MRC encourages projects that intend to use 3rd party data to negotiate a licence that will allow it to be openly shared. However, as these decisions are often beyond the control of the project team, only limited importance is placed upon this aspect. In most cases it is sufficient to state how access to 3rd party data may be obtained, i.e. application to XYZ.

The LSHTM RDM Service has published a decision tree for use in selecting an appropriate data sharing approach. This can be found at
http://www.lshtm.ac.uk/research/researchdataman/depositdata/datasharingmethod.html

> **Example:**
> The project will make available the anonymised interview transcripts (output 1), transcribed fieldwork diaries (output 2) and selected photographs (output 3) through LSHTM Data Compass.
>
> The English Longitudinal Study of Ageing is available for access via the UK Data Service. Any enhancements made by the project will be submitted to the UKDS for consideration.

Further Information
- LSHTM: Identify Data to Share
  < http://www.lshtm.ac.uk/research/researchdataman/share/datatoshare.html >
- UK Data Service: Planning for Sharing
  < http://www.data-archive.ac.uk/create-manage/planning-for-sharing>
- DCC: How to Appraise and Select Research Data for Curation
  < http://www.dcc.ac.uk/resources/how-guides/appraise-select-data >

---

[10] http://www.mrc.ac.uk/news-events/publications/mrc-policy-and-guidance-on-sharing-of-research-data-from-population-and-patient-studies/

5.2 Discovery by potential users of the research data

Describe the approach you will take to promote and encourage take-up of your data outputs.

LSHTM researchers are expected to offer research data that they generate to an appropriate data repository or data enclave. The Registry of Research Data Repositories (http://service.re3data.org) may be used to locate an appropriate subject / domain specific repository. Alternatively, you can upload it to LSHTM Data Compass (http://datacompass.lshtm.ac.uk/).

If you intend to make data publicly available through a third party system, you must register its location with the School, in compliance with the LSHTM RDM Policy[11]. This may be achieved by emailing the URL to the LSHTM RDM Service.

> **Example:**
>
> Data outputs 1, 2 & 3 will be deposited with LSHTM Data Compass (http://datacompass.lshtm.ac.uk/) for long-term curation and preservation, where it will be assigned a Digital Object Identifier (DOI). The DOI will be cited in project reports and journal publications through a Data Access Statement or citation list. Descriptive metadata on each data collection held in LSHTM Data Compass will be made available through OAI-PMH, RSS and ATOM in various formats (Dublin Core, MODS, METS) for use by third party services, such as research data catalogues.
>
> Information on each data collection will be provided to the MRC for inclusion in their Research Data Gateway for Health Sciences catalogue (https://www.datagateway.mrc.ac.uk/).

Further Information
- LSHTM: Choose a Data Sharing Approach
  < http://www.lshtm.ac.uk/research/researchdataman/depositdata/datasharingmethod.html >
- LSHTM: Create Discovery Metadata to improve research visibility
  < http://www.lshtm.ac.uk/research/researchdataman/describe/discovery_metadata.html >
- LSHTM: Deposit Data to LSHTM Data Compass
  < http://www.lshtm.ac.uk/research/researchdataman/depositdata/ >

5.3 Governance of access

The next step is to consider how data access will be made available. Funders and journals encourage researchers to make anonymised data freely available under a permissive licence, such as Creative Commons Attribution (CC-BY). However, in cases where this is not possible, some form of controlled access may be appropriate.

The MRC state population health and patient-based research must follow three additional requirements for access governance[12]:

1. Study governance of access, the criteria and processes, must be appropriate and proportionate to the nature and scale of the study, the level of risk and the likely demand for access.
2. Access processes must include independent advice and / or oversight.
3. Criteria and processes governing access must be transparent and readily discoverable.

A growing number of journals specify conditions on how requests for controlled access data are submitted and evaluated. For example, PLOS state "it is not acceptable for the authors to be the sole named individuals responsible for ensuring data access". Please contact the LSHTM RDM Service for advice on this topic.

---

[11] http://researchonline.lshtm.ac.uk/612422/

[12] http://www.mrc.ac.uk/news-events/publications/mrc-policy-and-guidance-on-sharing-of-research-data-from-population-and-patient-studies/ (p10)

> **Example:**
>
> An aggregated dataset will be made available as open access using a permissive licence, such as Creative Commons Attribution (CC-BY). No restrictions will be placed upon access.
>
> Interview transcripts that cannot be fully anonymized will be made available through a controlled access method. Access requests will be reviewed by a Data Access Committee (consisting of project members & external experts), which will make decisions in accordance with the project's access procedures.
>
> The English Longitudinal Study of Ageing (ELSA) is available through the UK Data Service.

Further Information
- LSHTM: Decide how access should be provided
  < http://www.lshtm.ac.uk/research/researchdataman/depositdata/access_permissions.html >
- MRC: Policy and Guidance on Sharing of Research Data from Population and Patient Studies
  < http://www.mrc.ac.uk/news-events/publications/mrc-policy-and-guidance-on-sharing-of-research-data-from-population-and-patient-studies/ >

5.4   The study team's exclusive use of the data

The MRC encourage "timely data sharing", which balances the project's need to publish research on the data that they have collected, with the broader need to provide appropriate access to publicly funded research. They do not establish specific timescales, however, beyond an expectation that the stated period must be clearly defined and justified.

State the time frame in which your data will be made available, outlining any policies, dependencies, or other factors that will influence your approach.

Decisions on data sharing timescales are often influenced by four factors:

1. *Subject domain:* Data sharing practices vary between research disciplines. Genetics and physics researchers, for example, have established practices for data sharing that emphasise early publication of research results[13], whereas other domains publish at a later date.

2. *Study Type:* For longitudinal studies, it is advisable to adopt a granular approach to data sharing, releasing subsets of data at distinct periods, rather than wait until the end of the project.

3. *Resource Type:* Community resources, such as those defined in the Fort Lauderdale Principles[14] and Toronto statement[15] that will have broad utility should be made available during the project lifetime.

4. *Publication plans:* Many journals expect data that underpin research findings to be made available at the point of publication, so that it can be validated by others.

---

[13] http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/Public-health-and-epidemiology/WTDV030690.htm

[14] http://www.wellcome.ac.uk/About-us/Publications/Reports/Biomedical-science/WTD003208.htm

[15] http://www.nature.com/nature/journal/v461/n7261/full/461168a.html

> **Example**
>
> Journal papers will be written and submitted towards the end of the project. At this point, data that underpin these papers will be made available for access
>
> The study team will make data outputs 1, 2, and 3 available in their entirety within 6 months of project completion, in line with conditions outlined in 5.3. This embargo period is requested to allow time for additional analysis and further publication of research findings.

Further Information
- MRC: Policy and Guidance on Sharing of Research Data from Population and Patient Studies
  < http://www.mrc.ac.uk/news-events/publications/mrc-policy-and-guidance-on-sharing-of-research-data-from-population-and-patient-studies/ >
- LSHTM: Timescales for Data Sharing
  < http://www.lshtm.ac.uk/research/researchdataman/share/when_to_share.html >

5.5   Restrictions or delays to sharing, with planned actions to limit such restrictions

Consider the factors that may limit opportunities for data sharing and the steps you will take to address these requirements. Relevant factors include the need to:

- *Protect participant confidentiality:* The identity of study participants should be protected, unless they have given permission to be identified.

- *Comply with informed consent agreement:* The MRC expect consent procedures to include provision for data sharing. However, it recognizes that limitations upon permitted use may be necessary for health data. For example, participant may allow data to be shared, on condition it is only used to investigate specific medical conditions or research questions.

- *Protect Intellectual Property Rights:* Health research often requires use of data that incorporates third party rights, e.g. government, funding bodies, research labs, and others. Steps should be taken to clarify these rights and address any requirements.

- *Submit patent applications:* A study that wishes to patent its research may request that specific restrictions are applied to data sharing. E.g. an embargo period.

- *Protect commercial confidentiality:* Research that involves use of commercial data, or which includes a commercial partner may require additional protection. Limitations may be established that restrict who is allowed access to the data and the permitted use.

Possible measures that may be applied to address these issues include: the implementation of access controls, redaction of personal / sensitive information, and development of a Data Transfer Agreement. These may be expressed as shown in the following table.

No.	Examples
1	No restrictions are considered necessary. However, the study team will obtain advice from LSHTM's legal department if potential issues arise.
2	To protect study participants from identification, the end user will be required to complete a Data Access Request form via LSHTM Data Compass to access interview transcriptions. The form will cover relevant topics, such as proposed use and details of the ethics approval received. Visit http://dx.doi.org/10.17037/DATA.2 to see how this approach is applied in practice. If their request complies with the participant consent agreement and they are willing to sign a DTA, they will be provided with access to the data.
3	Biological samples are limited and depletable, so access will need to be carefully controlled and coordinated. The quantity of sample that is provided will be judged against the potential benefits of the research project, with advice from appropriate experts as required.

Further Information
- UK Data Service: Legal and Ethical < http://ukdataservice.ac.uk/manage-data/legal-ethical.aspx >
- LSHTM: Data Sharing
  < http://www.lshtm.ac.uk/research/researchdataman/share/ >
- LSHTM: Decide how access should be provided
  < http://www.lshtm.ac.uk/research/researchdataman/depositdata/access_permissions.html >

5.6   Regulation of responsibilities of users

Outline any controls that must be set to limit who is allowed access, the purpose for which they may use the data, and how access may be obtained.

It may be appropriate to establish a graded access procedure, in which anonymised and aggregate data is made freely available, whereas sensitive and confidential data can only be accessed through application.

No.	Example
1	Open access data is licensed under a CC-BY-NC licence, which prohibits use for commercial purposes.
2	Controlled access data is available through application, on condition that the end user meets appropriate access conditions and signs a Data Transfer Agreement indicating they will not attempt to re-identify or contact participants.
3	Biological samples will be deposited with the UK BioBank for future use. Applicants will be required to comply with a Data Transfer Agreement prior to gaining access to data.
4	State the timescales for data sharing, outlining any policies or dependencies that will influence your approach

Further Information
- MRC: Policy and Guidance on Sharing of Research Data from Population and Patient Studies
  < http://www.mrc.ac.uk/news-events/publications/mrc-policy-and-guidance-on-sharing-of-research-data-from-population-and-patient-studies/ >
- UK Data Service: Legal and Ethical < http://ukdataservice.ac.uk/manage-data/legal-ethical.aspx >
- LSHTM: Choose a Data Licence
  < http://www.lshtm.ac.uk/research/researchdataman/share/choose_licence.html >

6. Responsibilities

The sixth section covers roles and responsibilities for data management activities within the study / consortia. You should outline three aspects:

1. The role / individual responsible
2. Task they will perform, e.g. study-wide data management, metadata creation, quality assurance
3. Institution or location (if a consortia and multi-site projects)

Individuals should be named where possible. If you intend to recruit the role at a later date, this should be stated in the Data Management Plan.

For collaborative projects you should explain how data management responsibilities will be coordinated across partners.

7. Relevant institutional, departmental or study policies on data sharing and data security

Finally, provide URLs for relevant policies and procedures applied by LSHTM, collaborating institutions, or others that influence your approach to data management and sharing.

Policy	URL or Reference
LSHTM Research Data Management Policy	http://researchonline.lshtm.ac.uk/612422/ or http://www.lshtm.ac.uk/research/researchdataman/rdm_policy.html
LSHTM Information Management and Security Policy	http://www.lshtm.ac.uk/its/informationsecurity/policy/index.html
Data Sharing Policy	See RDM Policy, principle 7 http://www.lshtm.ac.uk/research/researchdataman/rdm_policy.html#principle07
LSHTM Records Management Policy	Held on LSHTM Intranet – available on request

Further Guidance

Guidance on MRC requirements can be found at:

- MRC: Data Management Plans
  < http://www.mrc.ac.uk/research/research-policy-ethics/data-sharing/data-management-plans/ >

- MRC: Data Sharing
  < http://www.mrc.ac.uk/research/research-policy-ethics/data-sharing/ >

- MRC: Guidance for reviewers in assessing Data Management Plans
  < http://www.mrc.ac.uk/documents/pdf/data-management-plans-guidance-for-reviewers/ >

- London School of Hygiene & Tropical Medicine: MRC guidance
  < http://www.lshtm.ac.uk/research/researchdataman/plan/funder_mrc.html

- University of Bristol: Data Management Planning guide for MRC applicants
  < http://data.bris.ac.uk/research/dmp/mrc/

- University of Leicester: MRC Funding applicants: Data Management Plans
  < http://www2.le.ac.uk/services/research-data/create-data/mrc/ >