RDSOP17
Research Data Management and Security

Greater Manchester Mental Health NHS Foundation Trust
Title of Standard Operating Procedure: RDSOP17 Research data management and security

Document Summary: The purpose of this SOP is to describe the procedures for managing research data and ensuring that research data is stored in a safe and secure manner.

Document Author: Sue Dobson, Research Grants Co-ordinator

Target Audience: Trust-wide, Research Community, Internal and External Researchers

Consultation: R & I Office, research community and R & I Committee members

Approval Committee: R & I Committee

Cross Reference Document(s): Research Approval Policy All Trust R & I SOPs

Contact details for further information: Sarah Leo, Head of Research & Innovation Office 0161 271 0076 researchoffice@gmmh.nhs.uk

<table>
<thead>
<tr>
<th>Minimum Monitoring Requirement</th>
<th>Frequency</th>
<th>Process for monitoring</th>
<th>Evidence</th>
<th>Responsible Individual(s)</th>
<th>Response Committee(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Review of SOP content</td>
<td>Annually</td>
<td>Review by Author, redraft, submission to R &amp; I Operational Group</td>
<td>Minutes of R &amp; I Operational Group</td>
<td>Head of R &amp; I Office</td>
<td>Research &amp; Innovation Committee</td>
</tr>
</tbody>
</table>
Contents

1. Introduction ........................................................................................................................................ 3
2. Purpose ................................................................................................................................................ 4
3. Roles and Responsibilities .................................................................................................................. 4
  3.1 Duties within the Organisation ........................................................................................................ 4
4. Procedure ............................................................................................................................................ 5
  4.1 Data Management and Security Procedure – Basic Terms .............................................................. 5
  4.2 Data Management .............................................................................................................................. 5
  4.3 Data Security ..................................................................................................................................... 6
  4.4 Data Transfer ................................................................................................................................... 7
  4.5 General Data Protection Regulation – Data Transparency ............................................................... 7
5. Appendices .......................................................................................................................................... 8
  5.1 Data Management Plan ..................................................................................................................... 8
6. References and Bibliography .............................................................................................................. 12
1. Introduction

All data collected for research purposes must comply with the Data Protection Act 2018 (DPA). This includes data collected from patients, families and staff. The DPA outlines the conditions that must be met by organisations and individuals who hold or process personal identifiable information about living individuals:

- Personal data should be fairly and lawfully processed
- Data should be adequate for the purpose and not excessive
- Data should be obtained for specified purposes
- Data should be accurate and up to date
- Appropriate measures should be taken to ensure confidentiality and security of the data
- Data should be processed in accordance with the rights of the data subject
- Data should not be kept longer than is necessary
- Data should not be transferred outside of the European Economic Area without an adequate level of protection

UK Policy Framework for Health and Social Care Research v3.3 07/11/17 supports DPA and states that: “The appropriate use and protection of patient data is paramount. All those involved in research must be aware of their legal and ethical duties. Particular attention must be given to systems for ensuring confidentiality of personal information and to the security of those systems”.

Around the same time as the DPA was published, a review was commissioned by the Chief Medical Office of England “owing to increasing concern about the ways in which patient information is being used in the NHS in England and Wales and the need to ensure that confidentiality is not undermined. Such concern was largely due to the development of information technology in the service, and its capacity to disseminate information about patients rapidly and extensively”. A committee was established under the chairmanship of Dame Fiona Caldicott, Principal of Somerville College, Oxford, and previously President of the Royal College of Psychiatrists. Its findings were published in December 1997.

The ‘Caldicott’ principles and recommendations apply specifically to patient-identifiable information, and emphasise the need for controls over the availability of such information and access to it. In particular, a Caldicott Guardian, appointed in each NHS organisation, has specific responsibilities to oversee an ongoing process of audit, improvement and control.

The six Caldicott principles, applying to the handling of patient-identifiable information, are:

- Justify the purpose(s) of every proposed use or transfer
- Don’t use it unless it is absolutely necessary, and
- Use the minimum necessary
- Access to it should be on a strict need-to-know basis
Everyone with access to it should be aware of their responsibilities, and understand and comply with the law.

The 2018 DPA is the key legislation covering all aspects of information processing. This includes security and confidentiality of personal information. The Caldicott requirements provide the framework to put the DPA into operation.

The EU’s General Data Protection Regulation (GDPR) came into effect on the 25 May 2018. For an organisation such as the GMMH, our role and level of involvement changes with each particular study. However in each case we are required to be transparent in how personal data is collected, processed, accessed and stored as well as informing service users of the safeguards involved to retain compliance with appropriate legislations, and their rights around this.

2. Purpose

The purpose of this SOP is to describe the procedures for managing research data and ensuring that research data is stored in a safe and secure manner.

3. Roles and Responsibilities

3.1 Duties within the Organisation

It is the responsibility of the Research & Innovation Office to make Trust R & I SOPs available to all research active staff working on Trust-approved research studies.

It is the responsibility of the study Chief Investigator (CI) or local Principal Investigator (PI) to ensure that up-to-date copies of Trust R & I SOPs are available to research staff.

It is the responsibility of the study Chief Investigator or local Principal Investigator to ensure a link to up-to-date SOPs is filed in the Investigator Site File and research teams are aware how to access SOP on the intranet and internet.

It is the responsibility of the study Chief Investigator or Principal Investigator to designate if the SOPs of another organisation are to be followed for a study. For example those of a commercial sponsor. If there is significant conflict between the external SOP and the Trust R & I SOP it is the responsibility of the CI or PI to resolve these with the Research Office prior to starting the study.

It is the personal responsibility of all staff to follow Trust (or the designated alternative organisations) procedural documents.

The Research & Innovation Office is responsible for managing Trust R & I SOPs including their approval, dissemination and archiving. All Trust R & I SOPs must be made available and published on the internet and intranet.
4. Procedure

4.1 Data Management and Security Procedure – Basic Terms

Personal data. Personal data is any information that may lead to the identification of a living person that if released would put them at significant risk or harm or distress.

Anonymised data. Anonymised data is where all patient or participant identifiers (which can include name or initials, address, date of birth, hospital or NHS number) have been permanently removed. Anonymised data are not covered by the DPA.

Pseudo-anonymised data. Pseudo-anonymised data is where all personal identifiers (which can include name or initials, address, date of birth, hospital or NHS number) are replaced with a unique identifier (e.g. patient study number). A code should be held separately from patient identifiers, and allow for study un-blinding if required by the protocol.

Source documents. Source documents are original documents, data and records (e.g. hospital records, clinical and office charts, laboratory notes, diaries or evaluation checklists, laboratory results, x-ray or other scan results, pharmacy dispensing records). These are the ‘essential documents’ as described by ICH GCP that allow the evaluation and verification of the research study and data collected.

Source data. Source data are all information in original records and certified copies of original records of clinical findings, observations, or other activities in a research study necessary for the reconstruction and evaluation of the study. Source data are contained in source documents. PARIS and DATIX data are used as source data in a research study.

For practical application of above definitions please refer to the IG Handbook: https://newintranet/GMMH-Policies/Pages/Information-Governance.aspx

4.2 Data Management

All research data containing personal information about patients or participants should be anonymised. For certain studies, such as clinical trials of an investigational medicinal product, this may not be possible and the data should be pseudo-anonymised. In these instances, un-blinding procedures for safety reasons should be clearly defined in the protocol.

To ensure data governance compliance a data management plan should be composed for each research study. This can either be incorporated in the protocol/IRAS form or a separate plan can be composed (see appendix 1 for a suitable template)

Clinical trial data should be collected onto a Case Report Form (CRF). This can be a printed, optical or electronic document and designed to record all the information detailed in the protocol. All staff completing CRF’s should be trained to do so.
Corrections to the printed CRF should be initialled and dated by the PI or persons delegated to do so. Correction fluid should never be used.

Clinical trial data from the CRF may be entered onto a database or other data management system. For clinical trials, data entry should be completed by the Data Manager, Data Custodian or other person designated to do so by the Principal Investigator. The data entry process (e.g. single data entry or double data entry methods) should be decided based on the size and complexity of the study. For small single site studies, single data entry (one person) with source verification is usually appropriate. This will involve a visual check between what is recorded on the paper CRF and what is entered onto the data management system.

Missing values, values outside of normal ranges or inconsistencies entered onto CRF’s should be reported to the PI.

Source data verification can be conducted by members of the research team or independent monitors. This will involve cross-checking the data entered onto the CRF against source documents for accuracy.

All research staff using a data management system should be trained in its proper operation. A training record should be kept and updated following any software changes or upgrades.

The data management system should be appropriate for the analysis planned, ensure data integrity and be auditable. In the case of certain data management systems (e.g. Microsoft Excel), it may be necessary to save and date older versions of the database or print copies.

All computer systems used in the conduct of clinical trials of investigational medicinal products (databases, Electronic Data Capture etc), must have their current release number and date of release, last release number and date of release and validation status recorded.

For information on storage and archiving of files, please refer to RDSOP21.

### 4.3 Data Security

All non-electronic data such as paper CRFs and other documents, audio and video recordings, should be kept in locked filing cabinets in lockable rooms only accessible by authorised personnel. Transfer of non electronic copies should be minimal and only when necessary. Secure lockable bags/locks have to be considered. Please contact R&I Office for further guidance.

Electronic data should only be stored on devices that are backed up regularly such as NHS Trust or other servers (e.g. University). This should be confirmed with IT support or by consulting Trust or local policy documents.
Backup electronic copies should be kept in a separate secure location to the master copy. These should be updated daily.

Data should be password protected and access limited to authorised personnel. Authorised users should login with their own account details and should never login to provide access to another user.

Laptops and memory sticks used to store data should be encrypted by Trust, local IT support or the study coordinating centre.

### 4.4 Data Transfer

Research data transferred within the European Economic Area (EEA) must be fully or pseudo anonymised. Any codes held to unblind or unlock data identity should not be sent.

Research data transferred outside the EEA must be fully or pseudo anonymised. An agreement should be in place documenting that the data will be held or processed according to the principles outlined in the DPA (2018). Any codes held to unblind or unlock data identity should not be sent.

Participants should be consented for their data to be transferred to a third party.

Approvals from the relevant ethics committee/HRA and/or Caldicott Guardian should be obtained and relevant contracts signed before data is transferred to a third party.

### 4.5 General Data Protection Regulation – Data Transparency

As a research Sponsor, GMMH acts as the “Data Controller”. The Data Controller determines the purpose and means of processing personal data. It is the Sponsor who determines what data is collected for the research study through the protocol, case report form and/or structured data fields in a database.

As a research Host (i.e. where the Sponsor is another organisation, such as a university or commercial company), GMMH acts as the “Processor”. The Processor collects personal data on behalf of the Data Controller.

It is our priority that research participants are aware of how their data will be accessed and processed for research purposes. The Trust’s research transparency statement can be found at:-

[https://newintranet/GMMH-Policies/Pages/Information-Governance.aspx](https://newintranet/GMMH-Policies/Pages/Information-Governance.aspx)

Researchers should liaise with the research Sponsor to ensure that the appropriate transparency statements are inserted into Protocols and Participant Information.
Sheets. The HRA has provided some useful guidance template at :-
https://newintranet/GMMH-Policies/Pages/Information-Governance.aspx

As a Trust we have a responsibility to produce regular Information Governance updates on data flows used within the department. It is the researcher and the CI/PI’s responsibility to report any changes to data flows to ensure full R&I oversight and accurate reporting.

5. Appendices

5.1 Data Management Plan

(Document based on mrc.ukri.org template)
The following template should be used to develop a Data Management Plan (DMP) to accompany a research proposal.
If you opt NOT to use the template the topics listed in the template should be addressed in the research study protocol.
0. Study Title

Exactly as in the study protocol

1. Description of the data

1.1 Type of study

Up to three lines of text that summarise the type of study (or studies) for which the data are being collected.

1.2 Types of data

Types of research data to be managed in the following terms: quantitative, qualitative; generated from surveys, clinical measurements, interviews, medical records, electronic health records, administrative records, genotypic data, images, tissue samples,...

1.3 Format and scale of the data

File formats, software used, number of records, databases, sweeps, repetitions,… (in terms that are meaningful in your field of research). Do formats and software enable sharing and long-term validity of data?

2. Data collection / generation

Make sure you justify why new data collection or long term management is needed in your Case for Support. Focus in this template on the good practice and standards for ensuring new data are of high quality and processing is well documented.

2.1 Methodologies for data collection / generation

How the data will be collected/generated and which community data standards (if any) will be used at this stage.

2.2 Data quality and standards

How consistency and quality of data collection / generation will be controlled and documented, through processes of calibration, repeat samples or measurements, standardised data capture or recording, data entry validation, peer review of data or representation with controlled vocabularies.

3. Data management, documentation and curation

Keep this section concise and accessible to readers who are not data-management experts. Focus on principles, systems and major standards. Focus on the main kind(s) of study data. Give brief examples and avoid long lists.

3.1 Managing, storing and curating data.

Briefly describe how data will be stored, backed-up, managed and curated in the short to medium term. Specify any community agreed or other formal data standards used (with URL references). [Enter data security standards in Section 4].

3.2 Metadata standards and data documentation

What metadata is produced about the data generated from the research? For example descriptions of data that enable research data to be used by others outside of your own team. This may include documenting the methods used to generate the data, analytical and procedural information, capturing instrument metadata alongside data, documenting provenance of data and their coding, detailed descriptions for variables, records, etc.

3.3 Data preservation strategy and standards
4. Data security and confidentiality of potentially disclosive information

This section MUST be completed if your research data includes personal data relating to human participants in research. For other research, the safeguarding and security of data should also be considered. Information provided will be in line with your ethical review. Please note this section concerns protecting the data, not the patients.

4.1 Formal information/data security standards

Identify formal information standards with which your study is or will be compliant. An example is ISO 27001. If your organisation is ISO compliant, please state the registration number.

4.2 Main risks to data security

All personal data has an element of risk. Summarise the main risks to the confidentiality and security of information related to human participants, the level of risk and how these risks will be managed. Cover the main processes or facilities for storage and processing of personal data, data access, with controls put in place and any auditing of user compliance with consent and security conditions. It is not sufficient to write not applicable under this heading.

5. Data sharing and access

Identify any data repository (-ies) that are, or will be, entrusted with storing, curating and/or sharing data from your study, where they exist for particular disciplinary domains or data types.

5.1 Suitability for sharing

Is the data you propose to collect (or existing data you propose to use) in the study suitable for sharing? If yes, briefly state why it is suitable.

If No, indicate why the data will not be suitable for sharing and then go to Section 6.

5.2 Discovery by potential users of the research data

Indicate how potential new users (outside of your organisation) can find out about your data and identify whether it could be suitable for their research purposes, e.g. through summary information (metadata) being readily available on the study website, in databases or catalogues. How widely accessible is this depository?

Indicate whether your policy or approach to data sharing is (or will be) published on your study website (or by other means).

5.3 Governance of access

Identify who makes or will make the decision on whether to supply research data to a potential new user.

Indicate whether the research data will be deposited in and available from an identified community database, repository, archive or other infrastructure established to curate and share data.

5.4 The study team’s exclusive use of the data

What are the timescale/dependencies for when data will be accessible to others outside of your team? Summarize the principles of your current/intended policy.
5.5 Restrictions or delays to sharing, with planned actions to limit such restrictions

Restriction to data sharing may be due to participant confidentiality, consent agreements. Strategies to limit restrictions may include data being anonymised or aggregated; gaining participant consent for data sharing; gaining copyright permissions. For prospective studies, consent procedures should include provision for data sharing to maximise the value of the data for wider research use, while providing adequate safeguards for participants. As part of the consent process, proposed procedures for data sharing should be set out clearly and current and potential future risks associated with this explained to research participants.

5.6 Regulation of responsibilities of users

Indicate whether external users are (will be) bound by data sharing agreements, setting out their main responsibilities.

6. Responsibilities

Apart from the PI, who is responsible at your organisation/within your consortia for:

- study-wide data management
- metadata creation,
- data security
- quality assurance of data.

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

Please complete, where such policies are (i) relevant to your study, and (ii) are in the public domain, e.g. accessible through the internet.

Add any others that are relevant

<table>
<thead>
<tr>
<th>Policy</th>
<th>URL or Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Management Policy &amp; Procedures</td>
<td></td>
</tr>
<tr>
<td>Data Security Policy</td>
<td></td>
</tr>
<tr>
<td>Data Sharing Policy</td>
<td>e.g. a study policy of sharing research data</td>
</tr>
<tr>
<td>Institutional Information Policy</td>
<td></td>
</tr>
<tr>
<td>Other:</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td></td>
</tr>
</tbody>
</table>

8. Author of this Data Management Plan (Name) and, if different to that of the Principal Investigator, their telephone & email contact details
6. References and Bibliography

Data Protection Act (2018)

Caldicott review: information governance in the health and care system (2013)
https://www.gov.uk/government/publications/the-information-governance-review

UK Policy Framework for Health and Social Care Research (v3.3 07/11/17)

ICH – Good Clinical Practice guidelines

HRA GDPR Guidance (2018)

GMMH GDPR Guidance (last accessed 19/07/2020)
https://www.gmmh.nhs.uk/gdpr-in-research

MRC Guidance on Data Management Plans (last accessed 23/07/2020)