NSLHD-RESEARCH DATA MANAGEMENT GUIDE

NSLHD My Research Hub is an innovative support centre designed to help researchers navigate the research process. The Hub contains the centralised support services for research at NSLHD, including research ethics and governance, clinical trials support, research strategy, grants and funding, finance and workforce support, REDCap, data management, biostatistics, research training, communications, and research promotion.

The responsible conduct of any research encompasses the appropriate generation, collection, access, use, analysis, disclosure, storage, retention, disposal, sharing and re-use of data and information. The importance of appropriate data management and handling is being increasingly recognised in the health and clinical research community. Research institutions are expected to develop a clear strategy outlining the critical aspects and steps to govern the appropriate management of data in accordance with national policies and ethical guidelines. This document provides a comprehensive guide to data management for researchers.

In research involving human subjects, sharing and protecting data is of utmost importance, especially when the data contains personal and sensitive information. In research communities, a tension on the topic of data sharing often exists. However, in most cases, data can still be shared while upholding the spirit of data protection and the principles of ethical sharing of data.

The key challenges expected to be overcome by appropriate management and sharing of research data include:

1. Engagement of clinical research team with data management and sharing processes
2. Ensure data management and auditing are a routine part of clinical research
3. Ensuring all types of dealings with personal health data, including electronic and paper-based records, are accessed in accordance with NSW Health Code of Conduct
4. Access to data is legitimate and for HREC approved purposes only
5. Maintaining security of confidential and/or sensitive information, including that stored on communication and data collection devices, and servers

Objectives of this Guide

This guide has been written to assist researchers in deciding how to access and manage research data in accordance with NSW Health policies. The guide recommends that researchers apply appropriate safeguards and security measures when accessing and managing data and acknowledge relevant research standards for data access. Furthermore, this guide provides operational guidance in accordance with the obligations of the NSW Health Code of Conduct (PD2015_049) for research and the NSW Health Data Governance Framework (GL2019_002).

Specific purposes include:

- providing guidance to support the appropriate storage, access and management of health data used for research, in accordance with relevant policies
- providing research staff with assistance and practical tips for designing and complying with relevant data management policies, procedures and guidelines
- ensuring the protection of patient health information at all stages of data handling

Alignment with relevant NSW Health policies and guidelines

This guide supports the Australian Code of Conduct for the Responsible Conduct of Research and the NSW Health Data Governance Framework, among others. Both the Code and the Framework describe the broad principles and responsibilities relevant to correctly managing the data and information in research and refer to the organisational roles, decision rights, and accountabilities of people and systems as they perform health data related activities.

Critical enablers of this guide

Project: Data is accessed and used for the purpose approved by the HREC only

People: The requesting person has the appropriate authority to access the data

Settings: The platform or tool used for sharing minimises the risk of unauthorised access

Data: Appropriate and necessary security and protections are applied to the data

Output: The output from the data access and management process is safeguarded and shared only among authorised personnel

Australian Guidelines and Legislation relevant to Data Management

- Australian Code for the Responsible Conduct of Research (2018)
- Guide to Managing and Investigating Potential Breaches of the Australian Code for the Responsible Conduct of Research (2018)
- National Statement on Ethical Conduct in Human Research (2007) updated 2018
- Management of Data and Information in Research (2019)
- ICH Good Clinical Practice (GCP) - Integrated Addendum to ICH E6 (R1) Guideline for Good Clinical Practice E6 (R2)
- International Conference on Harmonisation / Good Clinical Practice (ICH/GCP) Guidelines
- NSW Ministry of Health Privacy Manual for Health Information 2015
- NSW Health Code of Conduct (PD2012_018)
- NSW Health Data Governance Framework 2019
- Privacy and Personal Information Protection Act 1998 NSW
- Health Records and Information Privacy Act 2002

Use of data in accordance with ethical and governance approvals

Information provided in the research ethics and governance applications should clearly outline details on research data management, including the consent, confidentiality, identification, collection, use or disclosure of data.

- All relevant ethical and governance approvals must be obtained before starting any research study
- Data may only be used per the ethics and governance approvals.

Refer to the NSLHD Research Governance and Human Research pages for details on the submission process for ethics and governance applications.

Research data accountability

Research data accountability refers to the transparent use of data collected during the implementation of a research project. This will help determine whether the proposed data usage is appropriate and in accordance with privacy principles of health information and the ethical collection, sharing and reporting of research data in compliance with relevant legislations and policies to ensure appropriate data stewardship. Consequently, this would enable researchers and research offices to have a communication trail to identify individuals and entities to be held responsible for any misuse of information. Data access and usage accountability must be clearly articulated to ensure the data integrity of electronically captured research data.

Data Standards

To achieve and maintain adequate standards of research data, all original data must be retained in the accepted data storage platforms. However, only the approved data parameters can be accessed following adequate and appropriate de-identification of patient personal information. Personal identifying data can only be accessed if the research protocol requires it for the purpose of communication with patients and if the patient has provided appropriate consent to access their personal health information and/or data linkages with authorized entities. In any case, it is important to ensure rigorous adherence to data standards which may also often be guided by the design of the research study.

Research Data Management Plan

It is advisable to devise a research data management plan (RDMP) at the time of ethics submission to the HREC. This RDMP can be concise but clearly state the data access management and reporting related roles to be undertaken for the monitoring and reporting of the research study and identify the responsibilities and accountability of researchers or position holders who will be accountable for all data handling. Researchers and investigators listed in a research grant must ensure that a data management plan is in place.

NSLHD internal researchers and clinicians, external collaborators, or research students listed in the grant and stated in the RDMP will only be considered authorised personnel for ethical access to data. External collaborators and research students (for example, USYD researchers and HDR students) may access and store data on the NSLHD servers. Any reproduction or external sharing must be in accordance with the protocol. It is advised that such plans may be included within the RDMP when seeking HREC approval.

Principles of responsible management of research data

The responsible management of research data must be in accordance with the Australian Code for the Responsible Conduct of Research. This encompasses:

- The need to develop a comprehensive research data management plan (RDMP) [Appendix A- NSLHD RDMP Template] for the handling, storing, and reporting research data in consensus with all researchers involved in a research project.
- State the indicators and parameters required at each stage of monitoring and reporting the research processes and outcomes.
- Transparency in declaring specific interests in all associated components of the research data being accessed.
- Share and communicate all relevant methodology and approaches to responsibly access, store and manage research data.
- Set up a strategy for accountability of accessing and reporting findings from research data in compliance with all relevant legislations.

Components of a Research Data Management Plan

Backup

A sound and credible data backup strategy must be implemented and continued throughout the project and post-project period. This is one of the most critical components to address. Consider including an encrypted cloud backup so that data is not lost even in the face of local catastrophe. In the case of online or network based electronic database, it is important to ensure the automated backup process is active and functioning. If using REDCap, the collected data are automatically backed up on the NSLHD SharePoint secure server. External storage devices like USB, hard drives are not recommended to store patient/research data as risk of loss or security breach may be more common. If for any reason external storage devices have to be used, recommend storing the devices in secured lockers or drawers with defined personnel access only.

Survey of existing data

If using existing or pre-collected data must include a detailed description of data sources, data types and justifications. Explain which components of the data will be captured from which sources, and make sure to use a relevant citation, link or DOI.

Data to be collected/generated

Describe the data that the project will generate including data types (e.g. anthropometry, EEG, MRI), formats (e.g. text, numeric, images, open ended) and the estimated amount of research data to be generated and/or used in the research while ensuring the contextual validity of the data. Describe data in general terms that address the type and amount/size of data expected to be collected and used in the project (e.g., fMRI images from ~50 research participants, maternal depression scores for 350 women in 6 weeks postpartum). Descriptions may indicate the data modality (e.g., imaging, genomic, mobile, survey), level of aggregation (e.g., individual, aggregated, summarised), and/or the degree of data processing that has occurred (i.e., how raw or processed the data will be).

Data owners and stakeholders

Include who owns the data and who will be responsible for it. This can include a description of the data roles and responsibilities for all personnel involved in handling the data management and ownership of the data. The data owner from a research project may be the chief/principal investigator, data sponsors, funding organisations. A general outline of the responsibilities of a Data custodian (senior managers), Data stewards (Data manager, data integrity officer, system administrator, REDCap Administrator), and Data users are provided below:

- Data Sponsor can be a senior role responsible for the overall management of the data collection and are responsible for the oversight, direction, guidance, of data to be collected and ensuring appropriate resources for the data custodians.
- Data custodian manages the data collection system and implements the data delivery process (data collection, storage, security, disposal, administration, quality assurance, maintaining data standards, data access, governance approvals).
- Data stewards would report to the data custodian to develop, update, and operationalise the established data policies and enact data management in line with those policies. Often this role overlaps with either the data custodian or the data integrity officer.
- Data Integrity Officer/ Data manager will assist the data custodians in defining and controlling the data with high level expertise in the content of the data they manage. They will implement the data management functions including identifying and improving data quality issues, resolving data management challenges, monitoring data management activities, training and assisting research staff on data collection requirements, and participating in data quality improvement programs.
- Data users are other staff with any form of access to the data. This can also be external stakeholders contracted to assist with data collection or monitoring. All study personnel

are responsible for collecting and maintaining the privacy, integrity, and security of the data held by their respective projects.

'Ownership' of research data can be difficult to determine, especially when the research project involves investigators from multiple research institutions. For research conducted within the purview of the NSLHD My Research Hub, most data are likely to be generated from within the LHD facilities. Often data extracted from external sources may be required. It is advised that researchers consider the source of data and claim ownership according to protocol. Agreements covering the ownership, stewardship and control of research data must be clearly stated in the CTRA. In instances when the research staffs themselves are moving between institutions, it is advised for the data ownership and stewardship to be held and maintained by a specific role holder and not the person themselves.

File formats

Be explicit about the format of data files that will be generated from the research project. Justify the use of the specific format (for example- .txt, .pdf, .csv, .xls, .dta). have plans for the feasibility of data file conversion or migration from one format to another. Be aware of the potential risk of loss of data or corruption of data if the migration is not pre-planned or is not feasible.

Metadata

Metadata is the data about the data, which is usually a structured description of the content, quality, condition, or other characteristics of data. May include description of the forms of metadata that will be generated and explain the standards that will be followed. Include all the information required for the data to be read and interpreted. A structured description of the characteristics of the data, including its contents, quality, and formats, provides a shared meaning and allows comparisons across similar datasets. Creating metadata makes it easier for future researchers to retrieve, use, and manage data.

Access and security

The Australian Code for Responsible Conduct of Research clearly outlines the responsibilities of researchers and institutions in ensuring secured access to research data. Any form of data generated from an NSLHD investigator initiated or external research project that has been approved by the HREC for sharing must be made available to authorised parties upon request and upon presenting appropriate documentation. This includes all forms of data, including medical records, audio-visual files, health information, and any output tables or summary information generated from a research study. A data custodian will remain in charge of deciding options for sharing data via open or mediated access as outlined by the HREC approval.

Actions that can be taken to ensure secured access:

- Safeguarding the regulated access to data will restrict the sharing of sensitive information and data that research participants have not consented to. Such measures ensure that research data meant for public sharing and dissemination is made accessible to those authorised to handle the data from the specific research project. Such access control is extensively relevant and significant for research data available electronically or online. The platform in which these electronic or online research data are collected, stored and made available is important. The larger the research project or a research project involving multiple institutions, including NSLHD, the higher the need to set up a tighter network with higher levels of access control. The NSLHD SharePoint has significant safety parameters in place to ensure that research data access from external platforms are regulated and mediated by authorised personnel only.
- Data access must always be provided on a ‘needs only’ basis, as outlined in the HREC approved research protocol. Consideration needs to be given as to whether de-identified data is sufficient to fulfil the request requirement.
- All those given access to any platform or system (including eMR or REDCap) can have a secure individual login using a relevant NSLHD email address. Passwords must not be shared with any unauthorised personnel.
- To minimise any form of disclosure, have robust governance measures in place to approve and monitor access to research data. Have a plan to investigate if any research data has been misused.

Data organisation

Include details on how the data files will be named, organised into folders, transferred between users or devices, and synchronised across multiple machines. Explain how the collaborative work using the research data will be managed and how data users will keep track of multiple versions of the data files and documents. Establish rules for data organisation and clearly describe the file naming and folder structures. Research data files and folders need to be labelled and organised in a systematic way agreed upon by the entire research team, so they are identifiable and accessible for current and future users. Ensure team consensus/agreement to use standard file naming and versioning plans. Keep in mind that it is important for someone new to the project to be able to follow the workflow easily. Some examples can be:

- File name: Be consistent and use descriptive names
- File structure: Create folders according to the file naming conventions, document their purpose and ensure the files are stored in the correct folders.
- Versioning: Establish naming and storage rules for different versions of the same data.
- Inventory: Track data files using a simple spreadsheet/database/collaborative tool

Storage

Data storage is the practice of storing electronic and paper-based data. This can be done online on the NSLHD SharePoint or on REDCap (also on the NSLHD SharePoint). All online stored data can be hosted and accessed through the internet. Off-line storage on USBs or CDs are not allowed due to the high risk of data loss. While storing data, it is important to remember that it must remain authentic, reliable, discoverable, accessible, usable, protected, and preserved for as long as required or authorised. NSLHD researchers responsible for data storing and sharing must check the accuracy of data, especially when data is auto populated from online data management platforms or SharePoint. All relevant staff who are handling project specific research data must be trained in the appropriate management of research data. Researchers must have a clear plan for the role/person responsible for the data storage and its management.

Preferred digital platforms for collection, storage and maintenance of data:

eMR
All patient specific personal health information is collected, used or disclosed in the course of routine care. The primary purpose of collecting personal health information in the hospital setting is the provision of care. As such, the use of health information, including data from electronic and paper medical records, and departmental databases, for research is of a secondary purpose. The Health Records and Information Privacy Act 2002 (HRIPA) limits the use of health information for purposes other than the primary purpose. Researchers can only use electronic medical records upon appropriate approval from the HREC and in accordance with HRIPA.

REDCap
REDCap (Research Electronic Data Capture) is a secure web application for building and managing online surveys and databases. The system was developed by a multi-institutional consortium based at Vanderbilt University but is hosted on NSLHD servers and managed by the REDCap Administrator. My Research Hub encourages NSLHD investigators and researchers to use REDCap for operational and clinical research purposes. The data is then held at the NSLHD server, and all forms of system and access related security are managed according to NSW Health data sharing policies.
Secondary use of medical records that are manually or systematically extracted onto the relevant study specific REDCap projects are acceptable only if the NSLHD REDCap platform is used for data collection upload, and storage. This is particularly important if the research project is an NSLHD investigator-initiated project.

Excel spreadsheet

While Excel is a great tool for collecting and storing data simply and quickly, it is not recommended for the purpose of data collection. Excel spreadsheets can only be used to store research data tables once relevant data outputs are extracted from the eMR or REDCap data capturing systems. However, it is pertinent that any data outputs and data tables stored in an Excel spreadsheet must be encrypted and protected with at least 2 layers of password protection which are accessible to authorised study personnel only.

Bibliography management

Specify the bibliography management tool (for example- EndNote, Zotero, Mendeley, Revman) that will be used and how the references will be shared across other research group members.

Data sharing

It is important to have a data sharing access plan during the study period and at the end of the study. Data sharing cannot be permitted unless all appropriate requirements have been met by both parties and all ethics and governance approvals have been obtained. This includes participant consent and agreements. Important issues to consider when sharing data

- Ethics and governance approval for using the data for current and future studies
- Participant consent regarding sharing of the data
- Disclosure and use of data as per participant information sheet and consent form
- Any new use of the collected data has been approved by the HREC
- Appropriate licensing of the data for researchers to share research data with others and to govern subsequent use of data
- The release of data from NSLHD may require a Material & Data Transfer Agreement (MDTA). Please contact the Research Office to discuss this. Please find the MDTA template here

Participant informed consent in relation to data collection, use, retention and sharing

- Include details of what the participant has consented to in terms of the data usage. For example, collection of data/tissues, data labelling (identified, coded etc.), sharing of data with other researchers/organisations, participation in sub-studies, consent to future contact, or consent to future research. Consent documentation must inform participants how research data will be stored, preserved, and used in the long term, how privacy will be maintained, specific conditions under which data access may be granted to external organisations/researchers and a description of information describing risks related to how data might be used. The database management platform should be able to identify a participant’s terms of consent so that the data can be appropriately used/stored/shared during all phases of the data life cycle.
- When a participant consents to data sharing, researchers should abide by the conditions of consent when sharing data and provide the conditions/restrictions of consent to a third party or collaborators with who were not stated as conditions of the consent. If the participant consents to future use of their data, researchers should

abide by the consent conditions when using the data in future studies.

- Where participants are provided with a choice for the use of data via the participant informed consent form, researchers should implement a means of tracking the choices (using a consent database) so that data is used in accordance with the individual participant’s permission. A consent database could include the date, time and conditions of the consent and specify whether the participant agreed to internal/external sharing of data, accessing personal health records from myhealth or use MRN to access clinical history. It is important to be specific on what aspects of data access and sharing the participant has agreed to and record it on the consent database. In addition, it is important to record which participants did not consent or gave partial consent to data access. This database must be adhered to at the time of data access to ensure only consenting participants’ information are accessed and shared.

The concept of data de-identification

- Study data and documents should be maintained to ensure participant privacy during all phases of the data/documentation life cycle. All data collected within NSLHD may be classified as fully identified data, re-identifiable data, and de-identified data.
- Fully identified data: A combination of factors (e.g., date of birth; sex; postcode) may make information potentially identifiable, particularly in the case of an uncommon disease or condition and in a small community.
- Re-identifiable data: Providing a unique identifying number (e.g., Unique Reference Number [URN], Medical Reference Number [MRN], etc.) or using a technique that allows a researcher to go back to identify a particular patient. It must be noted that any data containing potentially identifiable information falls within the HRIP Privacy Act and policies.
- De-identified data: Does not comprise any identifying factors or features in a data set. A completely de-identified data does not trigger compliance with the Act as these no longer contain ‘personal information’.

A plan for data retention and archiving must also be in place. Data retention refers to identifying how long the data should be retained and stored appropriately and applying all safeguard measures of data security. This allows others to discover, use and learn from the documents and ensures continued access to the data.

A data dissemination and publication plan may be incorporated, outlining the strategies to disseminate the research findings. It is crucial to ensure that at the time of publication and dissemination, the data sharing plans and participant consent regarding data usage continue to be adhered. All confidential and de-identified data must remain as such.

Destruction

Destruction of data is physically damaging a data medium so that it is not usable or read on any device or platform. Authorized disposal of research data must be done in a way as to render them unreadable and leave them in a format from which they cannot be reconstructed in whole or in part. A register of destroyed records must be maintained for future reference and accountability. Information and biospecimens used in research should be disposed of in a safe and secure manner, consistent with the consent obtained and any, legal, and ethical requirements appropriate to the study. A plan must be included for the circumstances in which data destruction will be warranted, and the responsible person/role must be clearly stated. The data destruction steps for paper-based or electronic data must be described.

Responsibilities

At each stage of the plan, it is essential to note the study personnel or the role responsible for each step. It is better to identify the role responsible for each action, not the person, as the person may change over time.

Budget

Outline any costs relevant to implementing a data management plan and overall data handling. Identify the costs required to perform each item/activity of the plan and make a clear budget for each item. An example of costs may be the cost of the time and effort for data curation processes and localised infrastructure (those not covered by a person’s salary support or institutional overhead costs). This could also include possible costs for hardware backups, research assistant time, metadata creation and maintenance, archiving etc.

Anything else

Do not restrict to the items listed above. Individual research studies may have different requirements for data handling strategies which may not have been covered in the checklist here.

Contact for Research Data Management at NSLHD

Rebeka Freckleton
Manager, Research Strategy and Partnerships
rebeka.freckleton@health.nsw.gov.au

Shahreen Raihana
Data Manager, NSLHD
shahreen.raihana@health.nsw.gov.au

Kenneth Vuong
REDCap Administrator, NSLHD
kenneth.vuong@health.nsw.gov.au

NSLHD Research Office
NSLHD-Research@health.nsw.gov.au
+61 2 9926 4590

Appendix A- NSLHD Research Data Management Plan (RDMP) Template

Data Collection

What type of data will be collected or generated during the project?

Consider including a description of the type of data- its characteristics and features.

How will the data be collected or created, or captured?

This may include the methods or processes for producing the data, expected file formats, use of existing or third-party data and any requirements associated with its use.

Data Documentation and Metadata

What supporting documentation and metadata will accompany the data to enhance data interpretation?

This may include details of any supporting information to be developed or documented; any metadata standard, controlled vocabularies or ontologies that will be used to describe the data; quality assurance processes (calibration, validation, etc) to be applied to the data; and any processes that will be followed for documenting or organising the data such as file name conventions and directory structures.

Data Storage and Security

How will the data be stored and backed up during the project?

This may include a description of data storage and security arrangements: estimated size/amount of data; the location of where the data will be stored; the location of where the data will be backed-up to; frequency of back-up procedures, and the person responsible.

How will the investigators manage or restrict access and security?

This may include a description of how access to the data will be managed; any security or restriction issues relating to access or storage; and details of any physical or non-digital outputs that need to be stored, including their location.

Ethics, Legal Compliance, Copyright, IP and Authorship

How will the investigators manage any ethical issues?

Provide information on ethics arrangements: methods used to manage sensitive, confidential, or private information; details of any restrictions due to ethical or privacy considerations on *the data; information for consent forms relating to retention of the data and protection of privacy and confidentiality and steps taken to manage these (de-identification, anonymising etc.)*

How will the investigators manage copyright and Intellectual Property Rights (IP/IPR) issues?

Provide information on copyright and IP arrangements: details of any agreements reached with partner organisations concerning ownership of the data; any copyright or licensing restrictions; or legislative regulations or requirements associated with collecting data from/sending to countries/locations outside of Australia.

Data Sharing and Access

How will you share the data? Will the data be deposited with an archive or repository or published on the web?

This may include information on access, sharing and reuse arrangements, including what data or non-digital outputs will be retained on completion of the project; where will these be stored; will some/all the data be shared or published; what supporting information will be available to assist with interpretation of the data.

Are any restrictions on data sharing required? Are there likely to be any costs associated with making the data available for sharing or reuse?

This can include the processes or steps to be taken to protect privacy and confidentiality; any restrictions that negate sharing or re-use of the data; any requirements for mediating access to the data; intent to deposit in data repository or archive; how soon after completion of the project can the data be shared; and any costs associated with making the data available for sharing or reuse.

Data Retention, Preservation and Archiving

Which data are of long-term value and should be retained, shared, and/or preserved?

Consider providing information on data retention and disposal, including how long the data should be retained (in line with Australian Code for the Responsible Conduct of Research 2018, and NSW Health Data Governance Framework 2019); the disposal date and data disposal approval process that will be followed.

What is the long-term preservation plan for the dataset?

Consider providing information describing preservation and archiving arrangements, including the sustainable file formats that will be used for long term access; descriptive information details the organisation and structure of the data and supporting information that will be made available with the data for reuse and interpretation; the person or position *responsible for managing long-term access to the data; and any expected costs associated with long term storage of the data.*