Tasmanian Data Linkage Unit

Governance Framework

Title:	**TDLU Governance Framework**
Version No:	**V1.0**
Approved and Commenced:	**July 2022**
Review Frequency:	**Annually**
Revision Date:	**July 2023**
Responsible Officer	**Manager, TDLU**
Approved by:	**Director, TDLU**
Supporting Information:	
Change Log:	

Key Partners

TDLU
Tasmanian Data Linkage Unit

CONTENTS

1. Introduction

The Tasmanian Data Linkage Unit (TDLU) was established in 2010 and is operated by the Menzies Institute for Medical Research under a contractual arrangement between the Department of Health Tasmania (DoH) and the University of Tasmania (UTAS). The TDLU's role is to facilitate data linkage for ethically approved research, planning, and policy-making projects in the interests of the public good. Data linkage refers to methods that bring together information from different sources that relate to the same individual or event, which is achieved using methods that minimise privacy risks (Population Health Research Network; PHRN, 2021).

2. Purpose

Broadly, governance refers to a range of regulations, rules, processes, and principles that ensures high quality practice and accountability while reducing the risk of harm. With respect to research, *The Australian Code for the Responsible Conduct of Research* (2018) states that an appropriate governance framework should be in place that considers quality, safety, privacy, risk management, financial management, and ethical acceptability.

The purpose of this document is to provide a Governance Framework to:

- Clarify the roles and responsibilities of key stakeholders
- Guide decision making concerning data linkage
- Ensure high quality practices with respect to data linkage
- Outline rules and processes relating to the Master Linkage Map (defined in section 6.2) and research-ready datasets (defined in section 6.4) associated with the TDLU

3. Scope

This framework is relevant to:

- TDLU staff
- Key stakeholders (i.e. data users and data custodians)
- All data linkage undertaken at the TDLU
- Research-ready datasets that are stored and managed by the TDLU

4. Overview of the TDLU

4.1. Funding

The TDLU is funded by the Australian Government as part of the PHRN. The PHRN is an initiative of the Australian Government and is conducted as part of the National Collaborative Research Infrastructure Strategy. The PHRN was established to build a

nationwide data linkage infrastructure capable of securely and safely managing health and other information from around Australia. This network of data linkage units now operates in every Australian state and territory. The network has greatly improved the way linkable health and related data are made available for approved research purposes.

The TDLU receives additional funding from the Tasmanian DoH and the TDLU charges fees for the provision of linked data services to users. The TDLU also receives in-kind contributions from the Menzies Institute for Medical Research (Menzies), UTAS, and the Tasmanian DoH.

4.2. Structure

There are three distinct teams that operate under the TDLU: the Client Services, Data Linkage, and Data Integration teams (see Figures 1 and 2). These teams are physically separated and individually staffed to maintain high levels of data security and to uphold the separation principle (see section 6.1 for more details).

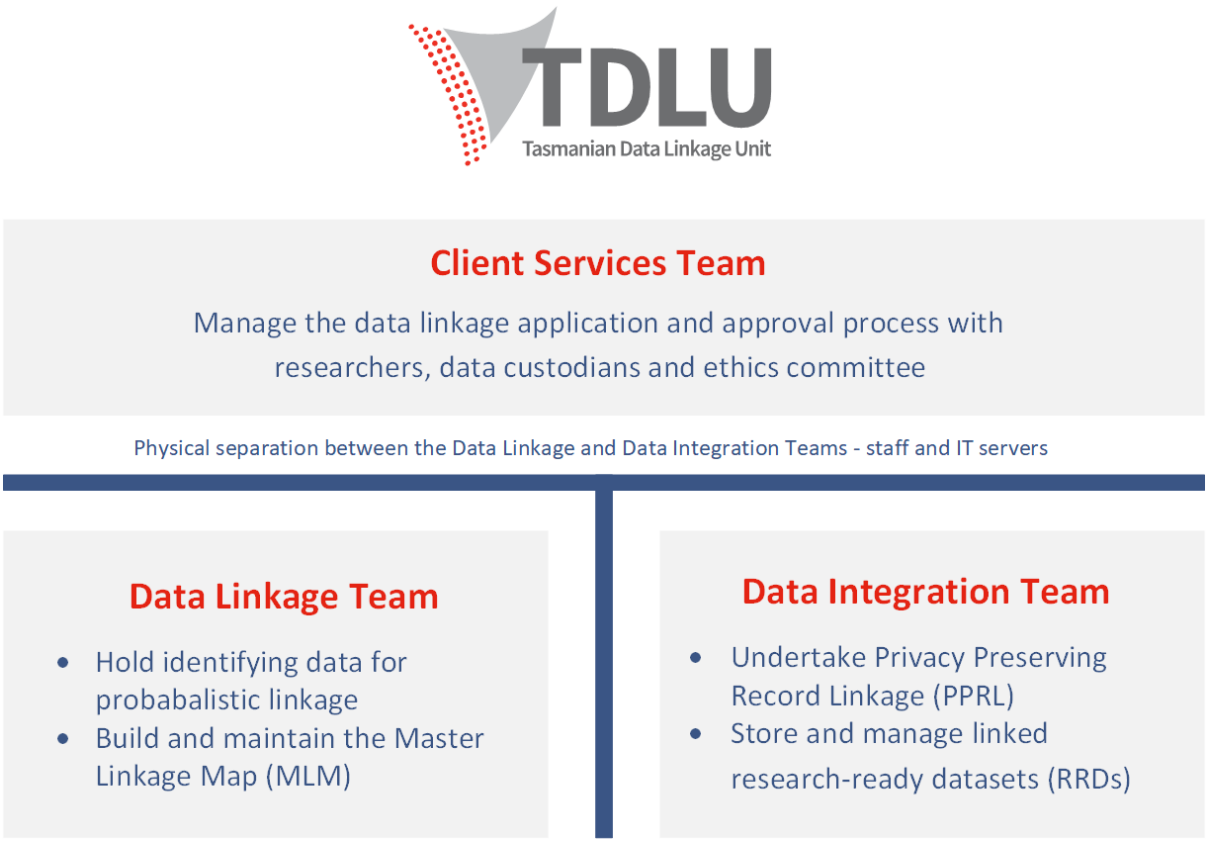

Figure 1. Summary of roles for the three, physically separate TDLU teams.

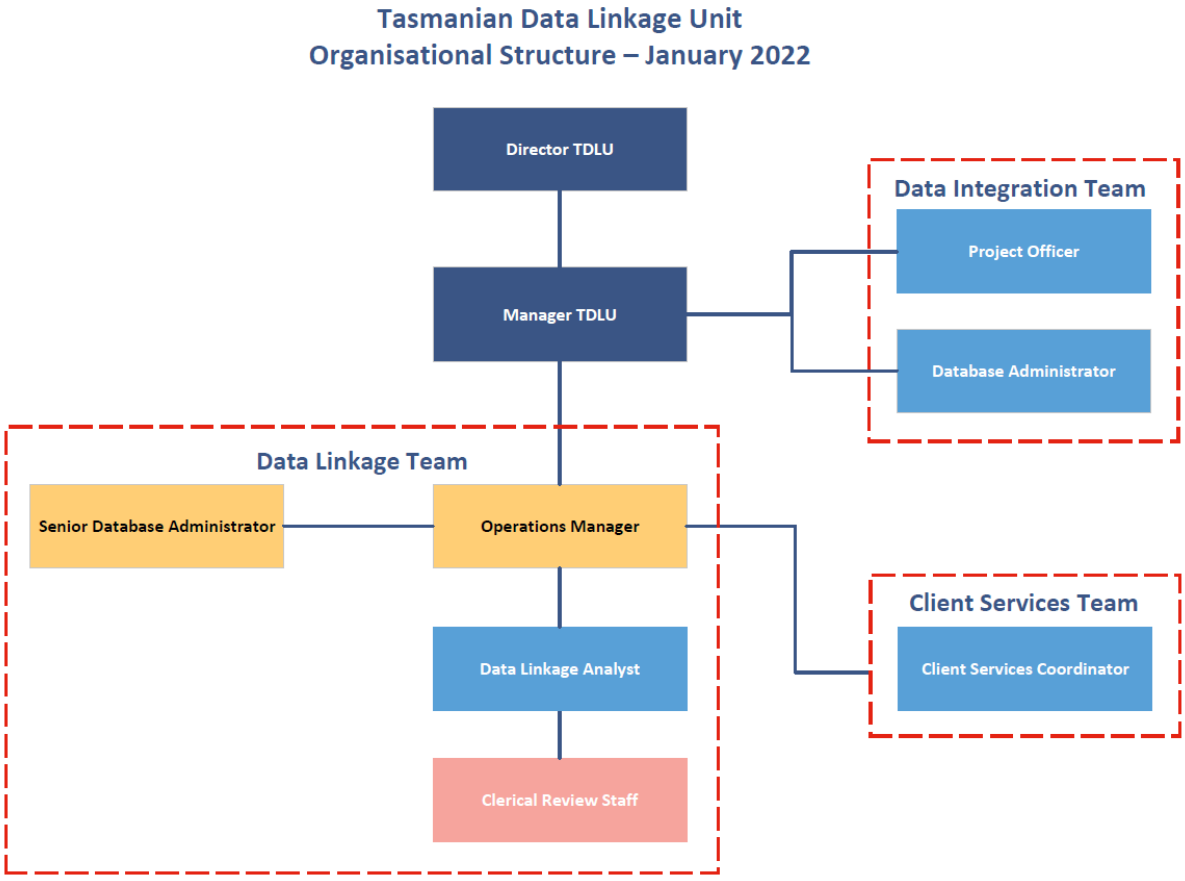

Figure 2. Organisational structure of the TDLU.

4.3. Monitoring, Reviews, and Reporting

The TDLU has a range of monitoring, review, and reporting arrangements. Reviewing and monitoring is completed by several parties from different standpoints:

PHRN - The PHRN monitors the TDLU's compliance with funding agreements, reviews performance and finances, conducts audits, and provides guidance on best practice data linkage. The TDLU is required to submit regular milestone and progress reports to the PHRN for review.

Tasmanian DoH – The DoH contracts UTAS to operate the TDLU. All monitoring reports, budgets, and financial statements provided to the PHRN are also provided to the DoH. Additionally, the TDLU meets with the DoH on a quarterly basis to monitor and review the operational aspects of the TDLU.

Tasmanian HREC - The TDLU operates under the approval (reference #H0026587) of the UTAS Human Research Ethics Committee (HREC). The HREC requires yearly progress reports from the TDLU regarding the functioning of the unit from an ethical standpoint.

The ethics committee are notified of staffing or operational changes to the TDLU and if there are any data breaches or misconduct relating to the TDLU. The HREC are also notified if the TDLU becomes aware of any relevant data breaches, researcher misconduct, or any other relevant data linkage incidents occurring outside of the TDLU.

TDLU Management Committee - The TDLU's Management Committee provide high level oversight of the unit and is chaired by the DoH. More information on the Management Committee can be found in the Management Committee's *Terms of Reference*.

Menzies – The TDLU is staffed by Menzies/UTAS employees and operates within the Menzies organisation and management structure. The TDLU is required to comply with all UTAS governance and reporting requirements.

Data custodians – The TDLU would not operate without the support and good will of data custodians. Formal Data Exchange Agreements are in place with key State Government Departments for the ongoing supply of data for linkage. The TDLU provides data custodians with information about each data linkage project and a project cannot proceed until data custodians have provided written approval for the supply of data to the project.

TDLU Data Linkage Team – The Data Linkage Team prepares reports and data linkage statements regarding each data linkage project. These documents cover linkage methodology, quality assurance, and linkage rates. Each report is provided to researchers on completion of their data linkage project.

TDLU Data Integration Team – The Data Integration Team, who manage research-ready datasets, prepare documents that report the scope of the dataset, dataset quality and limitations, the application and approval process, and conditions of the project.

TDLU Client Services Team – The Client Services Team monitors and reviews data linkage applications and approvals across all data linkage projects.

5. Roles and Responsibilities

5.1. TDLU

The roles and responsibilities of TDLU staff are outlined below. There are several shared responsibilities across staff of the Client Services, Data Linkage, and Data Integration teams. All staff are required to sign documentation that outlines conditions regarding data management and adherence to UTAS policies and procedures. Furthermore, all staff are responsible for maintaining acceptable levels of data security and privacy over data held by the TDLU.

A common role of all TDLU staff is to assist in the development, review, and maintenance of policies and procedures that support the operation of the TDLU. Staff also ensure that TDLU operations comply with ethical and privacy requirements, policies, and frameworks, as well as report any conflicts of interest or data breaches that are relevant to the TDLU.

TDLU Director

The TDLU Director holds academic (UTAS) responsibility for the TDLU. The Director provides high level strategic advice and guidance to the TDLU Manager.

TDLU Manager

The TDLU Manager has a range of duties that include (but are not limited to) implementing the TDLU's strategic plan, developing and delivering project output and milestones, collaborating with stakeholders, developing working relationships and collaborative linkages, securing funding, managing budgets, promoting the TDLU, managing and supervising reports, overseeing work performance and professional development, and representing the TDLU at stakeholder meetings.

TDLU Data Linkage Team

The Data Linkage Team maintains the Master Linkage Map, collaborates with data custodians to facilitate the supply of data for linkage, conducts data linkage, provides information on quality assurance, and generates linkage keys and associated linkage reports. The Data Linkage Team also assist data users with data linkage applications.

TDLU Data Integration Team

The Data Integration Team primarily conduct Privacy Preserving Record Linkage, generate linkage keys, and manage the governance, storage, and access of linked research-ready datasets within the TDLU. The Data Integration Team also assist data users with applications to access data from research-ready datasets.

TDLU Client Services Team

The Client Services Team provides information and advice regarding data linkage to researchers and other stakeholders, coordinates and assists with data linkage applications, manages project fees, and liaises with data custodians regarding approvals for data to be used in projects.

5.2. DATA CUSTODIANS

Data custodians are organisations or parties who hold specific data collections. Data custodians store, transport, and ensure safe custody of data. The roles and responsibilities of data custodians involve:

- Reviewing and approving data exchange agreements with the TDLU
- Completing duties listed in respective data exchange agreements
- General management and oversight of relevant data
- Upholding data quality and integrity
- Accountability for the governance and security of respective datasets
- Reviewing and approving, or rejecting, applications to access datasets
- Extracting cohorts and securely transferring identified data, when required, to the Data Linkage Team
- Sending de-identified clinical/activity data to researchers and/or the TDLU Data Integration Team

5.3. Data Users

Data users include researchers, policy makers, and other individuals or groups who have approval to access linked data. The roles and responsibilities of data users are as follows:

- Apply for, and manage, ethics applications and data linkage applications for respective projects
- Provide the TDLU with relevant ethics documents and ethics approvals
- Comply with all data management and data handling conditions specified by the approving ethics committee
- Bear responsibility for protecting de-identified, linked data according to the rules and conditions under which approval was granted for the project
- Conduct analyses and research, or research-related activities, on linked de-identified data according to approved documents
- Effectively manage any data quality or ethical issues, including privacy breaches
- Comply with all terms and conditions listed in relevant TDLU documents
- Merge de-identified data, when necessary, that has been received from data custodians and/or the TDLU
- Ensuring that respective data linkage projects are scientifically, ethically, and methodologically (including the design, conduct, and analysis) sound

6. Data Governance

Data governance refers to exercising decision making and authority on matters concerning data and includes rules and guidelines around the use of data (Data Governance Institute, 2022). The following sections introduce data linkage concepts to provide a context for the TDLU's data governance arrangements while latter sections outline the governing principles regarding data linkage associated with the TDLU.

6.1. THE SEPARATION PRINCIPLE

The TDLU protects the privacy of personal information by establishing and maintaining strict security of all data that is provided by data custodians. The TDLU achieves this with the separation principle, which refers to a best-practice method that protects the identity of individuals in a population. The separation principle involves the separation of identifying data (identifiers) from clinical/activity data, in addition to the separation of relevant roles (see Figures 3 and 4). This principle ensures that no single party involved in linkage (i.e. data linkage unit and user of a linked dataset) can access demographic and clinical/activity data other than the data custodian. This governance principle helps protect privacy and enables data linkage to be ethically justifiable.

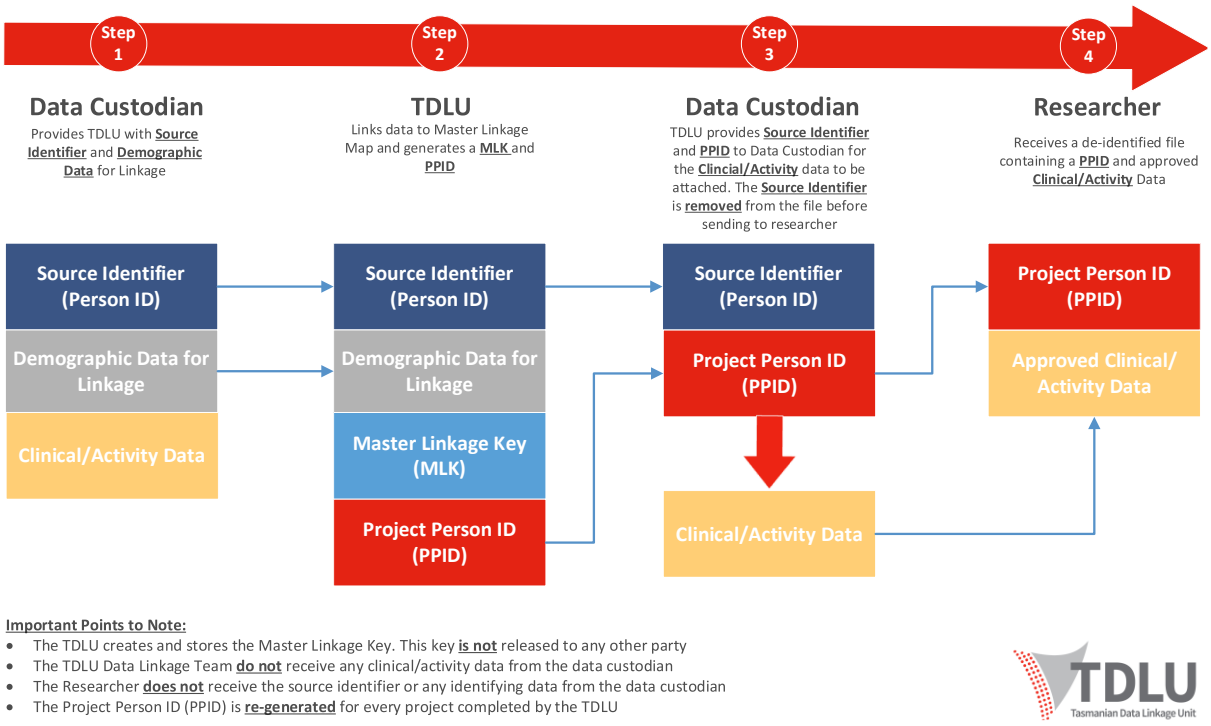

Figure 3. The separation of demographic and clinical/activity data during data linkage is facilitated by the creation of project person identifiers (PPIDs), which enhance the privacy of data.

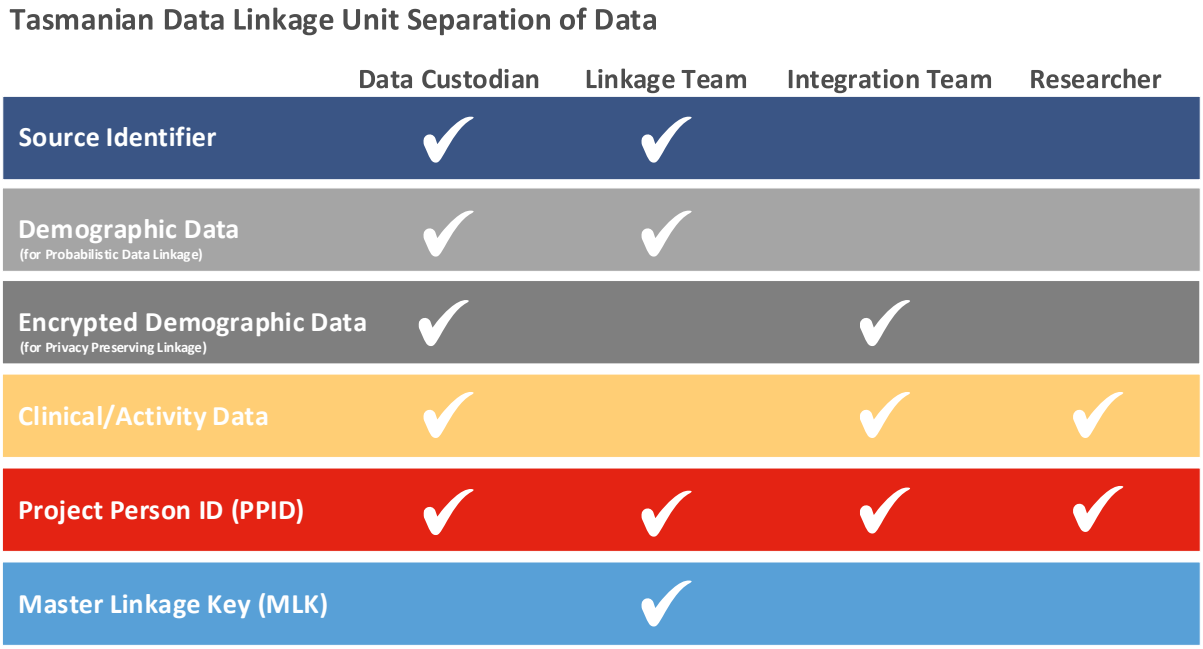

Figure 4. This diagram shows the level of data access between roles. The TDLU's Data Integration Team specifically deals with research-ready datasets (see section 6.4 for further details).

6.2. DATA LINKAGE

Data linkage at the TDLU typically involves data custodians supplying the TDLU Data Linkage Team with basic demographic data from various datasets in accordance with data exchange agreements or one-off agreements for specific linkage projects. The Data Linkage Team links and stores this data within the Master Linkage Map. The Master Linkage Map groups together records for individuals and only contains demographic data necessary for linkage purposes. Unique Master Linkage Keys (otherwise referred to as linkage keys) are stored in the Master Linkage Map; each record with the same Master Linkage Key belongs to the same individual in the population. Datasets covered by data exchange agreements are updated at regular time points as specified in respective agreements. In this arrangement, clinical/activity data is retained by data custodians. The TDLU generates Project Person IDs (PPIDs) that are then sent to data custodians along with their source identifier to enable extraction of the approved data variables. The data custodian sends the approved data variables and PPIDs directly to the data user.

Researchers can request to access specific datasets associated with the Master Linkage Map for a set timeframe as well as request new datasets to be linked to the Master Linkage Map. Linkage projects that are conducted by the Data Linkage Team are therefore customised to suit the researchers' specific requirements.

6.3. Data Linkage Methodology

The Data Linkage Team uses probabilistic data linkage techniques to facilitate the construction and maintenance of its Master Linkage Map. Probabilistic record linkage assumes that there is no common identifier within the source databases being linked that can identify an individual in the population with complete accuracy. Probabilistic record linkage calculates the probability that two records belong to the same individual using multiple pieces of identifying information including first name, other names, surname, gender, date of birth, date of death, and address.

A record identifier is provided by the data custodian that enables the return of linkage keys so that clinical, service, or activity data can be attached to the final, de-identified, dataset.

6.4. Research-Ready Datasets

The TDLU is approved to store research-ready datasets. Research-ready datasets are de-identified, pre-linked datasets that contain data from multiple data custodians. A single research-ready dataset can be used to answer many different questions without the need to continually re-link data and regenerate what is essentially the same, or a very similar, end-product. Research-ready datasets improve the speed of data access which, in turn, increases the use of linked data.

The process for research-ready datasets is, in brief, that an individual or party commissions (the project commissioner/commissioner) the TDLU to develop a research-ready dataset. The research-ready dataset may consist of datasets already associated with the Master Linkage Map as well as new datasets that are not linked to the Master Linkage Map. Once relevant approvals and agreements are finalised, data custodians transfer identifiable data to the Data Linkage Team who generate PPIDs that data custodians attach clinical/activity data to. Data custodians transfer this de-identified clinical/activity data, along with PPIDs, to the Data Integration Team for storage. At times, data custodians may not wish to transfer identifiable data to the TDLU, in which case privacy preserving data linkage methods are employed to encrypt identifiable data. The TDLU is considered the curator of the research-ready dataset given that the dataset is stored within the TDLU's Data Integration Team.

Once a research-ready dataset has been established, the project commissioner can apply to access subsets of data from their research-ready dataset on numerous occasions for separate projects. Other, independent parties/individuals can also apply to access subsets of data from the research-ready dataset. All parties or individuals wishing to use subsets of data from a research-ready dataset must obtain relevant approvals (see section 8). Following relevant approvals, the Data Integration Team extract data variables for the requested period and/or cohort and project specific person identifiers are generated by

the Data Integration Team to provide to data users along with clinical/activity data. New PPIDs are generated for each project to prevent the linkage of data across multiple projects, which enhances privacy.

6.5. Governance - All Data Linkage Projects

The following section outlines the general governance of all data linkage projects and datasets that are associated with the TDLU:

Security and Privacy

- TDLU staff and other parties involved in data linkage must use and store data and information in a secure, ethical, and lawful manner
- Data must only be used for the purposes outlined in formal documents, agreements, and approved applications
- Any approved party/individual that receives data is responsible for the security and privacy of that data. This includes not passing data onto other, unauthorised parties/individuals
- Any approved party/individual that holds data is responsible for the destruction of that data
- The TDLU Data Linkage Team do not receive, or have access to, clinical/activity data
- New PPIDs will be generated for each project to maintain privacy

Transparency

- The TDLU facilitates data linkage, it does not monitor or review projects with respect to scientific or ethical integrity. However, the TDLU may request to review ethics applications in relation to specific data linkage matters
- The TDLU conducts data linkage in accordance with approved ethics documents
- The TDLU will notify data users of any conditions stipulated by data custodians
- Data users will be provided with policies, procedures, and other documents relevant to TDLU data linkage
- Data users will receive a data linkage statement from the TDLU upon completion of each data linkage project
- The TDLU should be acknowledged in reports, publications, and other work
- Data users are requested to email manuscripts for publication to Menzies.tdlu@utas.edu.au at least two weeks prior to submission. The TDLU will review the wording relating to data linkage methodology and check for appropriate acknowledgement of data custodians and the TDLU. In accordance with relevant data exchange agreements, the TDLU may forward manuscripts to

data custodians to provide them with the opportunity to review and provide feedback

Permissions

- Any Australian-based entity can apply to the TDLU to use linked data
- All parties wishing to use linked data will require approval from data custodians, the Tasmanian HREC, and the TDLU
- Data custodians reserve the right to approve or deny access to their datasets

Quality and Training

- The TDLU will strive to ensure all staff are appropriately skilled and knowledgeable in data management and privacy matters
- TDLU staff will aim to conduct all work in a timely manner
- Errors may naturally be present in large administrative datasets, for which the TDLU does not take responsibility
- While the TDLU conducts quality checks on data, data users should conduct their own quality checks and report any issues to the TDLU to support ongoing quality improvement

6.6. Governance - Research-Ready Datasets

In addition to the above governance, the principles that underlie the governance of research-ready datasets are as follows:

Security and Privacy

- The Data Integration Team does not have access to the Master Linkage Map or identifiable data
- The TDLU will use established data linkage methods, including privacy preserving record linkage, to reduce the risk of breaching an individual's privacy
- The TDLU, in liaison with the project commissioner, reserves the right to cap the number of discrete datasets released for one project to, for instance, reduce the risk of re-identification
- The TDLU will only release subsets of data and will not release an entire research-ready dataset for a single project, including to commissioners of research-ready datasets

Permissions

- Research-ready datasets are intended to be shared with multiple, independent parties due to their potential research value

- Commissioners have the right to set conditions regarding their respective research-ready datasets. For example, a commissioner may choose to withhold the release of certain data variables from a given dataset
- Commissioners of research-ready datasets can make an unlimited number of requests to access subsets of data from respective research-ready datasets, subject to relevant approvals
- The TDLU, in liaison with the commissioner, reserves the right to set a *minimum* number of datasets to be released from a research-ready dataset to a given project
- The TDLU will maintain a registry of research questions that have been approved to use linked data to avoid duplication. This registry will be attached to data custodian approvals

Transparency

- Data users may be required to acknowledge project commissioners in outputs, which will be specified by the respective project commissioner

Quality

- The Data Integration Team may clean data, when necessary, in research-ready datasets prior to data being transferred to data users
- Any data cleaning conducted by the TDLU will occur once for each research-ready dataset
- For each research-ready dataset, the TDLU will create (when necessary) a document in collaboration with the project commissioner that specifies the rules, limitations, and quality of the dataset
- The TDLU cannot identify the error rate in projects where only de-identified data (i.e. in datasets involving privacy preserving record linkage) has been provided to the TDLU for linkage. However, approximate error rates can be calculated by the TDLU if a sample of identifiable data is released to the Data Linkage Team for a given dataset

6.7. Data Destruction Obligations

Long term and/or permanent retention of research data is permissible if the data is of interest to the public and if there are no agreements or rules specifying that the data must be destroyed. Data that is held by the TDLU relates to the public and therefore the data will be stored by the TDLU until the unit ceases to exist or relevant data exchange agreements expire or are terminated, or upon agreement between the TDLU and data custodian to destroy (or return) the data.

Data custodians reserve the right to stipulate the timeframes for the destruction of data within their data exchange agreement or approval documentation with the TDLU. The TDLU will retain data provided by data custodians for a minimum of six months from the time the data user receives all requested linked data. Data will then be stored or destroyed in accordance with agreements made with the respective data custodian. The TDLU maintains a register of data receipt and data destruction and advises data custodians when data has been destroyed.

Researchers must adhere to *The Australian Code for the Responsible Conduct of Research (2018)* regarding data destruction obligations.

Refer to UTAS Research Data Management Procedure for further details on data management.

7. Pricing of Data Linkage

The cost of data linkage projects is determined according to the TDLU's pricing models and pricing policy. Projects are assessed on a case-by-case basis.

8. Applying For Linked Data

All individuals or parties wishing to use linked data (including subsets of data from research-ready datasets) must apply to, and receive approvals from, the following:

1. The TDLU
2. A NHMRC-registered HREC, and
3. The data custodian responsible for each dataset

It is the applicant's responsibility to submit applications to the TDLU and HREC. The TDLU will manage the approval process with data custodians. The TDLU's website provides further information on the data linkage application and approval process.

Throughout the application process the TDLU will endeavour to:

- Provide an acknowledgement of receipt of data linkage applications within five business days
- Review expressions of interest that are lodged via the Online Application System (OAS) within 10 business days, or inform the applicant if there is an anticipated delay
- Approve applications in a timely manner. However, the timeframe to approve applications may vary due to a range of factors, such as the applicant's timeliness in completing tasks related to the application process

Data custodian approvals may take several weeks. A courtesy reminder email will be sent by the TDLU to the data custodian if required. The TDLU has no control over the timelines of ethics committees and data custodians.

9. Management of Collaborative Projects

- All parties involved in data linkage projects are required to adhere to *The Australian Code for the Responsible Conduct of Research* (2018)
- All stakeholders involved in data linkage must follow the TDLU's *Governance Framework*
- Conflicts relating to the TDLU will be managed by the TDLU's Director and/or Manager and will do so according to the UTAS Conflicts of Interest Procedure

10. Risk and Risk Management

The TDLU adheres to UTAS procedures regarding risk (see Appendix 1). Additional details regarding risk management are provided in the TDLU's *Security Framework*.

11. Legal Matters

Data users, data custodians, and the TDLU must comply with relevant laws, regulations, codes, and policies (see Appendix 1). All stakeholders should refer to the *Privacy Act 1988* regarding Australian law and Privacy Principles. The PHRN website can be viewed for further information on legal matters relating to data linkage.

12. Misconduct

Misconduct and the management of misconduct is described in Part B of the *Australian Code for the Responsible Conduct of Research* (2018). Misconduct relates to a range of issues, such as falsification, plagiarism, and failure to adhere to approved research proposals.

The TDLU is primarily responsible for managing misconduct *within* the unit itself. The TDLU will also report, when appropriate, misconduct issues relating to data linkage projects should the TDLU become aware of such issues. However, the management of misconduct relating to individual projects firstly lies with data users involved in the project.

Proven misconduct may result in several outcomes, such as disciplinary or legal action, depending on the nature of the misconduct. Queries about misconduct can be addressed with the TDLU Director or Manager and/or a University Research Integrity Advisor or HREC.

13. Terms and Conditions

- The TDLU will require an ethics amendment for the unit should changes be needed to the unit to support data linkage and related activities
- Under no circumstances will identifying data be made available to parties who should not have access to such data
- Data provision and related services may be subject to fees

14. Useful Websites

Title	Website
PHRN	https://www.phrn.org.au/
TDLU	https://www.menzies.utas.edu.au/research/research-centres/data-linkage-unit
UTAS Ethics	https://www.utas.edu.au/research-admin/research-integrity-and-ethics-unit-rieu/human-ethics

15. Definitions

Term	Definition	Acronym
Commissioner	The project commissioner/commissioner is the individual or party who finances the TDLU to create and store a linked, research-ready dataset	
Dataset	A collection of data	
Data custodian	An entity that holds data for any purpose and is accountable for the governance of that data. Data custodians have responsibility for granting access to data	
Data governance	"Data Governance is a system of decision rights and accountabilities for information-related processes, executed according to agreed-upon models which describe who can take what actions with what information, and when, under what circumstances, using what methods." Data Governance Institute (2022)	
Data user	Any individual or party involved in accessing or investigating data. This typically includes academics from research institutions or employees of agencies	

Data variables	Variables that represent data that are measured and can be analysed, e.g., weight, height, blood pressure	
De-identified data	Data that has been through a process of de-identification whereby identifying data is removed	
Identifiable data	Data that is either identifiable data or data that is de-identified but contains sufficient detail that a person could be identified	
Linkage variables	Variables that are used to link records, such as first name, last name, and date of birth	
Master Linkage Key	Refers to an individual's unique ID, otherwise known as a 'key' or 'linkage key'. Each key is stored in the MLM	MLK
Master Linkage Map	The MLM groups together records for individuals in a population. Each individual within the MLM has their own unique 'key'. The MLM contains basic data such as name, address, date of birth and gender to enable linkage across datasets to occur	MLM
Project Person Identifier	A PPID is a project-specific, unique pseudo identifier that is supplied to researchers that refers to an individual	PPID
Primary Health Network	A network of independent organisations that aim to streamline health services and better coordinate care. There are 31 PHN's in Australia	PHN
Population Health Research Network	An Australian Government Initiative to support research through the linkage of Australia's population-based health and human services data while adhering to the separation principle	PHRN
Separation Principle	The concept of separating identifying data from clinical/activity data, as well as the separation of roles to protect privacy	

16. REFERENCES

National Health and Medical Research Council. (2007, updated 2018). *The Australian Code for the Responsible Conduct of Research*. https://www.nhmrc.gov.au/about-us/publications/australian-code-responsible-conduct-research-2018

Population Health Research Network (2021, April 21). *Data Linkage*. https://www.phrn.org.au/about-us/data-linkage/#:~:text=Data%20linkage% 2C% 20in%20simple%20terms,data%20to%20researchers%20approved%20research.

APPENDIX 1: RELEVANT LEGISLATION, CODES, PROCEDURES, AND DOCUMENTS

Legislation

- Tasmania Personal Information Protection (PIP) Act 2004
- Commonwealth Privacy Act 1988

Statements

- National Statement on Ethical Conduct in Human Research (2018)

Codes

- Australian Code for the Responsible Conduct of Research (2018)

Guidelines

- Guidelines under Section 95 of the Privacy Act 1988 (2014)
- Australian Privacy Principles Guidelines
- Australian Privacy Principles (APP)

Related University of Tasmania Documents

- Collaborative Research Procedure
- Conflicts of Interest Procedure
- Data and Information Governance Policy
- Management of Research Data Procedure
- Publications and Research Disseminations Procedure
- Research Ethics Procedure
- Risk Management Framework