DATA LINKAGE TECHNICAL MANUAL

Introduction

This technical manual aims to support researchers and stakeholders who are considering projects involving data linkage. It is an instructional document which describes: the quality and completeness of ACT Health data available for linkage; the requirements and processes involved in obtaining approvals to conduct data linkage, project feasibility assessment and data governance; critical timeframes, system requirements and costs; as well as the stakeholders involved in data linkage.

For an overview and introduction to data linkage, see the *ACT Population Health Data Linkage in the ACT* document created by the Data Linkage Team, ACT Health, which this technical manual has been developed to complement. It is recommended and assumed that researchers and stakeholders who are reading this manual will have already read through the *ACT Population Health Data Linkage in the ACT* document, will have a data linkage research concept in mind, and will have a rough idea of their project's methodology. Readers of this document will likely know if their project involves connecting an external data set or study cohort to ACT Health data or whether they just want to link two or more ACT Health datasets together.

> **Key fact 1: Working with linked ACT population health data involves working with <u>more</u> than just the Data Linkage Team, ACT Health**
>
> To access linked population health data in the ACT, researchers need to work with the organisation which is responsible for holding the data (in this instance, the Data Linkage Team, ACT Health), the organisation responsible for linking the data (such as, for example, the Centre for Health Record Linkage, or "CHeReL") as well as the Human Research Ethics Committee(s).

Relevant Data Linkage Organisations

One of the principles in protecting the privacy of the personal information used in data linkage is maintaining the separation principle. This means that no one person ever has access to both the identifying information needed to link datasets and the content data for analysis. To achieve this there are several authorised, independent organisations that are responsible for linking data and maintaining linked data sets. These include the Centre for Health Record Linkage (CHeReL – NSW/ACT), dedicated data linkage units for other jurisdictions, the Australian Institute of Health and Welfare (AIHW) Data Integration Services Centre and the Australian Bureau of Statistics (ABS) Data Integration Service. Often these centres have a specific focus in the data they keep and link, for example CHeReL is primarily involved in linking ACT and NSW health data while the AIHW predominantly links to Commonwealth-owned data sets.

1

The CHeReL and AIHW Data Integration Services Centre are part of the Population Health Research Network (PHRN), which is an Australian Government Initiative to build a nationwide data linkage infrastructure capable of securely and safely managing health information from around Australia. It is a requirement that all data linkage activities utilising ACT Health data must be conducted through recognised data linkage facilities such as the CHeReL or the AIHW and is conducted according to best-practice, privacy-protecting principles.

The Centre for Health Record Linkage – CHeReL (www.cherel.org.au)

The Centre for Health Record Linkage (CHeReL) makes linked data accessible to the research, government and not-for-profit sectors. The CHeReL operates as a regional data linkage centre for NSW and ACT as part of Australia's first national data linkage network, the PHRN. The CHeReL enables release of billions of records each year in accordance with regulatory and governance frameworks. Records are linked using sophisticated machine learning techniques and include records from a core linkage system (the Master Linkage Key or MLK) that contains data on over 15 million people. The system is regularly linked to other datasets and across sectors, allowing health researchers to access linkage to justice, education, community services and transport data.

The system is continuously updated and current information is available from the CHeReL website (http://www.cherel.org.au/master-linkage-key). The CHeReL team can provide detailed advice on designing and planning data linkage projects and seeking necessary approvals to access linked data. A charge for services is determined on a partial cost recovery basis. To work with ACT Health datasets, it is necessary to work with both the ACT Health Data Linkage Team and the CHeReL.

Australian Institute of Health and Welfare – AIHW (www.aihw.gov.au)

The AIHW provides a service for linkage with Commonwealth (national) datasets, including the Medicare Benefits Scheme (MBS), Pharmaceutical Benefits Scheme (PBS), and the National Death Index (NDI). Where a project involves linkage of both ACT and Commonwealth-owned datasets the CHeReL conducts the linkage of ACT datasets and collaborates with the AIHW for the second stage linkage to the Commonwealth-owned dataset(s).

However, if researchers are connecting information about their cohort to the Commonwealth-owned datasets held by the AIHW, but without connecting them to ACT Health data, there is no need to work with the CHeReL and the ACT Health Data Linkage Team; researchers can go directly to the AIHW.

Providing data linkage assistance to researchers and governments is a core service of the AIHW Data Integration Service Centre: https://www.aihw.gov.au/our-services/data-linkage/researcher-resources. They can help set up data linkage projects, negotiate and arrange access to suitable datasets, and carry out all the necessary technical processes to link datasets safely and securely for researchers to use to answer their research questions. A charge for services is determined on a cost recovery basis in accordance with Australian Government Cost Recovery Guidelines and depends on the complexity of the project.

The Secure Unified Research Environment – SURE (www.saxinstitute.org.au/our-work/sure/)

Where a project requires Medicare Benefits Schedule, Pharmaceutical Benefits Scheme, Centrelink or certain other data, it is a data custodian requirement that the linked dataset must be stored and analysed within an approved secure research access environment.

Approved secure research access environments include:

- The Secure Unified Research Environment (SURE)

- The AIHW's onsite secure data laboratory.

SURE is a high-powered computing environment aiming to bring researchers together from across Australia and the world to collaborate on large-scale projects tackling major health and social issues such as population ageing, diabetes and mental health. The SURE provides a remote-access data research laboratory for analysing routinely collected data, allowing researchers to log in remotely and securely analyse data from sources such as hospitals, general practice, cancer registries and linked datasets. The SURE was established with funding from the Australian Government National Collaborative Research Infrastructure Strategy (NCRIS).

When researchers are collaborating across jurisdictions or organisations it can be helpful to use the SURE – it can assist researchers at different sites to work together using the same data set and research analysis tools, ensuring consistency and security is maintained. If a project only uses NSW and ACT data collections, it is not a requirement of NSW Ministry of Health or ACT Health to use SURE. More information about the SURE, including the registration process, access charges, training and system requirements, can be found here: https://www.saxinstitute.org.au/our-work/sure/using-sure/.

What ACT Datasets are Currently Available for Linkage?

ACT Health makes a range of datasets available for linkage through the CHeReL. The availability of datasets can change over time, so it is important to check the CHeReL website for the most recent information – see: www.cherel.org.au. Other important information on the CHeReL website includes the details of datasets including data dictionaries.

- ACT Admitted Patient Care (APC) – records inpatient separations (discharges, transfers and deaths) from ACT public hospitals (The Canberra Hospital and Calvary Public Hospital).

- ACT Cancer Registry (ACTCR) –registers all new cases of cancer diagnosed in ACT residents (excluding non-melanoma skin cancers) and all deaths of people with cancer.

- ACT Maternal Perinatal Data Collection (MPDC) (also known as the Perinatal Data Collection, PDC, and formerly known as the Midwives Data Collection) – is a population-based surveillance system covering all births in ACT hospitals, as well as homebirths. Currently only public hospital births are available via CHeReL.

- ACT Notifiable Diseases Management System (ACT NDMS) (including foodborne and communicable diseases) – is a register of the diagnosis of notifiable infectious diseases and conditions in the ACT.

- Births / Deaths – the ACT Registrar of Births, Deaths and Marriages (ACT BDM) keeps a permanent record of all registered births in the ACT. Mortality information for deaths occurring in the ACT is contained in two datasets: The Registry of Births, Deaths and Marriages (ACT BDM) death registrations and Australian Coordinating Registry (ACR) Cause of Death Unit Record File (COD URF – available from 2006 onwards).

- ACT Emergency Department Information System (EDIS) – contains records of patient presentations to public hospital emergency departments in the ACT (Canberra Hospital and Calvary Public Hospital), including diagnoses, triage categories and sources of referral.

- ACT Kindergarten Health Check (KHC) is a screening program that provides health checks for children attending kindergarten in the ACT. The health check includes vision, hearing, height and weight and development.

- ACT Ambulance Services (ACTAS) data is available for the purpose of data linkage for all electronic records from 2012, including dispatch and clinical data.

- The CHeReL is an Australian Early Development Census (AEDC) Approved Data Linkage Unit and links to the AEDC – a population-based measure of children's development in their first year of full-time school. The AEDC datasets provide widespread coverage of this group - national participation has been in excess of 96 per cent in each Census undertaken, in 2009, 2012, 2015 and 2018. Information on how to apply for linked AEDC data and other resources, including the AEDC Data Dictionary and application forms, can be found through the AEDC website: www.aedc.gov.au.

- The Australian and New Zealand Dialysis and Transplant Registry (ANZDATA) records the incidence, prevalence and outcome of dialysis and transplant treatment for patients with end stage renal failure.

What national datasets are available for linkage to the ACT datasets in a linkage project?

A range of national datasets are available from the AIHW. Again, the availability of datasets can change over time, so it is advisable to check the AIHW website (https://www.aihw.gov.au/our-services/data-linkage).

- National Death Index (NDI) – a national database which contains records of all deaths occurring in Australia since 1980. The data are obtained from the Registrars of Births, Deaths and Marriages in each State and Territory. For more information see: https://www.aihw.gov.au/about-our-data/our-data-collections/national-death-index

- Medicare Benefits Scheme (MBS) – Medicare Australia collects primary health care service events for which Medicare payment claims are made using the Medicare Benefits Schedule (a list of codes). Some of these codes can be disease specific but most are about service utilisation. For a full list of the codes see: www.mbsonline.gov.au/internet/mbsonline/publishing.nsf/Content/Home.

- Pharmaceutical Benefits Scheme (PBS) – provides a database of usage of the medicinal and therapeutic products available under the Pharmaceutical Benefits Scheme (PBS). For more information see: www.pbs.gov.au/pbs/home.

Key fact 2: There are several stages involved in accessing linked data
There are differences in the processes of obtaining ACT linked datasets to obtaining NSW datasets. For example, NSW datasets require custodian approval before ethics, while ACT datasets require ethics before custodian approval. For more information see the CHeReL website: https://www.cherel.org.au/apply-for-linked-data.

Requesting data: How do I access/request/use linked data?

The following flowchart and staged description show the process of accessing linked data from the Data Linkage Team at ACT Health.

"I want to conduct a data linkage project about …"
1. **Researcher: Plan your study.**
2. **Researcher: Contact the CHeReL (to discuss your project).**
3. **Researcher: Complete the CHeReL *Combined Protocol and Application for Data***

form.
4. **Researcher: Complete the required ethics approval forms.**
5. **Researcher: Provide documents to the CHeReL for a technical feasibility review.**
6. **Researcher: Obtain ethics approvals from the ACT Health Human Research Ethics Committee (HREC) and from the Calvary Public Hospital Bruce HREC (if seeking Calvary Public Hospital data).** **Note: NSW datasets require custodian approval <u>before</u> ethics.**
7. **Researcher: Approach ACT Health Data Linkage Team for in-principle Data Custodian sign off.**
8. **ACT Health Data Linkage Team: Coordinate approval from data custodians and notify researcher and the CHeReL of approval.**
9. **CHeReL: Process the data linkage.**
10. **ACT Health Data Linkage Team: Seek Data Custodian approval to release linked data.**
11. **ACT Health Data Linkage Team: Release linked data to the researcher.**
12. **Researcher: Conduct data analysis and provide findings to ACT Health Data Linkage Team before release or publication of results.**

Step 1: Researcher – Plan your study

Planning your study will involve carefully considering:

- Your research question – which must be feasible and have clear merit.

- The population health data that will help answer your question – ensuring that the data you are considering is clearly justified by the research question, noting that broad data requests, e.g. asking for all episodes of care for all ages may not be an appropriately targeted request for a specific question.

- Inclusion and exclusion criteria – for example, clearly stating if you require data for specific age groups, locations or a diagnosis subset.

- If the data will be able to accommodate your inclusion and exclusion criteria.

- For example:

  o Is data available for the time period you are considering?

- o Is data available in single age years or age-ranges?
- o How is the geographical data coded? By suburb? Postcode?
- o Is data about both public and private hospitals requested and available?

It is necessary to consider if the data holdings match your requirements and are available by checking, for example, the CHeReL website. There are different data dictionaries across different jurisdictions and datasets.

Step 2: Researcher – Contact the CHeReL

It is vital to commence discussing your research proposal with the CHeReL before you make your application for data, such as what is and is not feasible, and what is the best data to match your research question and plan. The CHeReL Research Project Manager is happy to discuss your project with you. They can be reached by phone: (02) 9391 9924 or email: MOH-CHeReL@health.nsw.gov.au. Based on this discussion, you may need to return to Step 1 and further consider the design of your research project before discussing your proposal again with the CHeReL.

Step 3: Researcher – Complete the CHeReL *Combined Protocol and Application for Data form*

The CHeReL *Combined Protocol and Application for Data form* and instructions on how to submit it can be found here: http://www.cherel.org.au/apply-for-linked-data-act.

It is advisable to provide a range of supporting documentation when submitting your *Combined Protocol and Application for Data form* to the CHeReL. This information will include a list of the proposed variables required for analysis and a draft of your Human Research Ethics Application (HREA) generated in the Research Ethics Governance Information System (REGIS). Information on how to set up and use REGIS can be found at the following website: https://regis.health.nsw.gov.au/.

This application will allow the CHeReL to review the form, provide feedback and a quote.

Researchers need to consider timelines and scheduling to ensure that the timing for accessing linked data matches project timelines. While the Data Linkage Team at ACT Health and the CHeReL are committed to providing the most up-to-date data for linkage projects, it takes time for data to be processed and made available to researchers. Timelines can be affected by the type of data requested, the number of different organisations involved, the complexity of the project and the number of approvals that are required.

Step 4: Researcher – Complete the required ethics approval forms

All studies using linked data must have appropriate ethical and legal basis. In most cases researchers will need to complete the following documents:

- *Combined Protocol and Application for Data form*
- Variables lists (available on the data dictionaries page of the CHeReL website)

All studies linking data provided by the ACT Health Data Linkage Team must gain approval from the ACT Health Human Research Ethics Committee (HREC). The submission of a *Waiver of Consent* application will be required for all data linkage projects. For advice and guidance regarding ethical issues, please contact the ethics secretariat at: ethics@act.gov.au.

For studies that require cross-institutional data (e.g. all public hospital Emergency Department and Admitted Patient Care), ACT Health HREC will facilitate streamlined ethical approval with Calvary Public Hospital Bruce, the University of Canberra and Australian National University HREC. Applications submitted to the ACT Health HREC will be reviewed and approved first by ACT Health and then submitted for expedited review through Calvary or the relevant university HREC. Further information about the Calvary Public Hospital Bruce HREC is available from the website: https://www.calvarycare.org.au/public-hospital-bruce/research-and-ethics/.

Please note that evidence of custodian approval must be provided with the ethics application for NSW datasets.

Step 5: Researcher – Provide documents to the CHeReL for a technical feasibility review

You will need to email a completed copy of the following documents to the CHeReL, to enable a technical feasibility review:

1. *Combined Protocol and Application for Data form* (as mentioned in Steps 3 and 4).
2. *NSW Privacy Form* (for NSW data only).
3. Data variables list for each dataset.
4. Data flow chart.
5. Draft of your HREA generated in REGIS.

CHeReL will assess your forms and provide a technical feasibility letter that must accompany your ethics submission. For NSW datasets, CHeReL will seek data custodian approval on your behalf for datasets in the MLK. For ACT datasets, the process for data custodian approval is slightly different (see Step 7).

Step 6: Researcher – Obtain Human Research Ethics Committee (HREC) approval

The ACT Health HREC aims to ensure the highest ethical standards are maintained in research projects carried out in the ACT and Australia. Human research to be undertaken in

ACT Health or using ACT Health data cannot proceed without approval from the ACT Health HREC or a National Mutual Acceptance (NMA) exemption where ACT Health or Canberra Health Services sites are named as participating sites on the approval notification. Research with a NMA exemption must still meet the site governance requirements and must not commence without appropriate authorisation. For more information about the NMA, please visit: https://www.health.act.gov.au/research/research-ethics-and-governance/nma-scheme.

To progress to the next stage in the process of accessing linked ACT Health data, the researcher must prepare and submit an ethics application to the ACT Health HREC. The CHeReL technical feasibility letter will form part of this ethics application. More information about the ACT Health HREC, including a checklist of requirements for the submission, fees and the committee meeting schedule can be found here: https://health.act.gov.au/research/research-ethics-and-governance.

Please note:

- You may also need to complete supplementary documentation if your research includes Aboriginal and Torres Strait Islanders.

- You will also need to apply to the Calvary Public Hospital Bruce HREC if you require ACT Public hospital data. This committee meets every two months. The ACT Health HREC can only provide approval to use hospital data from The Canberra Hospital.

- If your research links to the MBS, PBS or NDI you will also need to apply for ethics approval from the AIHW Ethics Committee. This committee meets five times a year.

- Submissions for research seeking approval through the ACT Health HREC must be made via the REGIS platform, using the HREA form.

If the ACT Health HREC determines that the researchers and the project comply with the ethical and legal responsibilities outlined in the National Health and Medical Research Council National Statement on Ethical Conduct in Human Research, then it grants ethics approval. At this stage the researcher may also be in the process of applying for ethics approval from other organisations and jurisdictions depending on the project.

Key fact 3: Projects require separate ethics and custodian approval

It is important to note that Data Custodian approval is not automatically granted if a project has ethics approval.

Step 7: Researcher – Approach ACT Health Data Linkage Team for Data Custodian sign off

Following ethics approval, the researcher must provide a comprehensive application to the ACT Health Data Linkage Team for forwarding to obtain Data Custodian approval on the

researchers' behalf. It is important to be aware that Data Custodian Approval will not automatically be granted, even if ethics has been approved. In considering the project, the ACT Health Data Custodians consider aspects such as how the data will be reported, and its merit.

While open, hypothesis-generating research questions are welcome, the researcher/requestor must demonstrate an outcomes-driven, research-oriented, theory-based proposal. In considering projects for approval, ACT Health Data Custodians must adhere to legal and policy responsibilities of their office. As such, the analytical capacity and credibility of the research team will also be taken into account to ensure that risks are well managed and the team is well placed to make appropriate interpretations of the data.

To apply for Data Custodian approval, please email the documents below to the ACT Health Data Linkage Team at: HealthInfo@act.gov.au – indicating that your email is a request for Data Custodian approval for a data linkage project:

- A copy of your completed CHeReL *Combined Protocol and Application for Data form.*

- A description of your research and full research protocols (if not using the *Combined Protocol and Application for Data form*).

- A copy of the CHeReL technical feasibility letter.

- A copy of all ACT Health CHeReL data variables checklists.

- Ethics approval from the full range of organisations/sources from which you're requesting access to data.

- Any other supporting documentation.

ACT Health will then process your application for Data Custodian approval which can take four to six weeks depending on the complexity of the request, the number of data sets involved and the process of internal review. The process may involve some discussion between the researcher and the ACT Health Data Linkage Team.

If the Data Custodian approves the release of data for your project, you will receive a letter informing you that your project is approved. The ACT Health Data Linkage Team will then send a copy of this letter electronically to the researcher, which will need to be forwarded on to the CHeReL to enable data linkage to commence.

Step 8: ACT Health Data Linkage Team – Coordinate approval from data custodians and notify researcher and the CHeReL of approval

When you are advised of ethical and Data Custodian approvals, please notify the CHeReL and forward all documents listed on your approval letter. The CHeReL will confirm all requirements before scheduling your linkage.

Step 9: CHeReL – Process data linkage

The CHeReL then commences processing the data linkage according to its schedule of work and the requirements of the project. For updates, please email: MOH-CHeReL@health.nsw.gov.au.

Step 10: ACT Health – Seek Data Custodian approval to release linked data

When the data linkage process is complete, the ACT Health Data Linkage Team seeks approval again from the Data Custodian to release the content to the researcher. The ACT Health Data Linkage Team works with the CHeReL to provide the linked results files that will be supplied to the researcher.

While waiting for the data custodian to release data, the ACT Health Data Linkage Team will send the researcher a "Conditions of Release" agreement, which includes conditions such as, for example that you agree to provide the ACT Health Data Linkage Team with draft embargoed copies of your findings for our comment before they are released. The principal investigator completes this and returns it to the ACT Health Data Linkage Team.

Step 11: ACT Health – Release linked data to researcher

The ACT Health Data Linkage Team provides data in a flat text file format (a CSV file), sent securely as an encrypted file. The researcher will be provided with a password to decrypt the file. Other formats may be provided on request (e.g. STATA or SPSS files). If the research project is using the SURE, the ACT Health Data Linkage Team will upload the data into the SURE workspace.

Step 12: Researcher – Conduct data analysis and provide findings to ACT Health before release or publication of results

The researcher works with the data to complete their research project in compliance with REC conditions and the conditions of release. There is scope for researchers to ask questions about the data during the analysis, for example, interpretation of the variables and the continuity of variables over time. The researcher is required to provide findings to ACT Health prior to release or publication of the study.

Population Health Research Network Online Application System

The Population Health Research Network (PHRN) is a national network of Data Linkage Units located in each Australian state/territory. Where a project is using cross and multijurisdictional data, the PHRN Online Application System may be a more appropriate tool for researchers. Requirements to access and use data are similar to those outlined above. For more information, please visit: www.phrn.org.au.

Still have questions?

Contact the ACT Health Data Linkage Team

The ACT Health Data Linkage Team
Epidemiology Section, Preventive and Population Health, ACT Health
GPO Box 825
Canberra ACT 2601
Phone: ((02) 5124 9458
Email: HealthInfo@act.gov.au

Thank you and happy linking.

Accessibility
If you have difficulty reading a standard printed document and would like an alternative format, please phone 13 22 81.

If English is not your first language and you need the Translating and Interpreting Service (TIS), please call 13 14 50.

For further accessibility information, visit: www.health.act.gov.au/accessibility

www.health.act.gov.au | Phone: 132281 | Publication No XXXXX