

Population Health Research Net

• [Home](#) / [About Us](#) / [What is Data Linkage?](#)

What is Data Linkage?

- [What is Data Linkage?](#)
- [Benefits](#)
- [What We Have Learnt](#)
 - [Case Study 1](#)
 - [Case Study 2](#)
 - [Case Study 3](#)
 - [Case Study 4](#)
 - [Case Study 5](#)
 - [Case Study 6](#)
 - [Case Study 7](#)
 - [Case Study 8](#)
 - [Case Study 9](#)
- [Privacy and Security](#)
- [PHRN Management Council](#)
 - [Proof of Concept Reference Group](#)
 - [Access Committee](#)
 - [Ethics Privacy & Consumer Engagement Advisory Group](#)
 - [Data Transfer Working Group](#)
 - [Operations Committee](#)
- [Policies](#)
- [Current Projects](#)
- [Annual Reports](#)
- [PHRN Timeline](#)



Data linkage, in simple terms, is a method of bringing together information about people, places and events in a way that protects individual privacy. Using specially developed technology PHRN nodes will be linking existing information from different health and health related data collections to provide data to approved researchers for a range of population-based studies.

To watch a seven minute video presentation on the data linkage process (courtesy of [SA-NT DataLink](#)) please click [here](#)

Is it all about health?

Historically, research based on linked data has been used in the health sector to examine a range of health issues and provide us with a picture of the health and well-being of our community. The results from these studies have provided valuable information on the causes of and risk factors for disease, the evaluation of new approaches to prevention, early intervention and treatment, and the assessment of new and more effective ways of delivering health services.

More recently, health data has been linked to data from other areas including education, criminal justice and child protection, providing new insights into factors that can influence the health and well-being of the population.

[Read more](#)

What type of data is collected?

Throughout our lives information about us is routinely collected by government departments and other agencies. Many of our life experiences from the moment we are born until our death generate data that is collected and used for a range of purposes. For example, information is routinely collected when you visit a hospital, when you get married or divorced, when you have a child or when you are diagnosed with an infectious disease or with cancer. This information is collected via departments and agencies such as public hospitals, state and territory health departments, and Births, Deaths and Marriages registries. It includes information on hospital admissions, births and deaths, cancer and infectious diseases.



The collection of this information, or data, is generally required by law. The information is used for policy, planning, management, monitoring, evaluation and research purposes. It is stored in secure data collections within each agency and all access to the data is controlled by a Data Custodian (an authorised nominee of the agency who holds the data). Additionally, governments and other agencies may collect information from time to time for a particular purpose such as a survey of health behaviours or a specific research project.

Who is involved in data linkage?

There are three distinct groups involved in data linkage and access processes.

- **Data Custodians** - are the people who look after the data collections. Data custodians work within an organisation or agency (such as a government health department) and are responsible for the secure collection, use and disclosure of data. Data custodians collect and store personal information (eg.name, address, date of birth) and also content information (eg.health information such as diagnosis and treatment details)
- **Data Linkers** - are the people who create Linkage ID's which allow data to be linked within and between data collections. Data Linkers work in a Data Linkage Unit that is either within, or associated with, a government agency.
- **Researchers** - are the people who use the data for the purpose of analysis and research. This process is only possible after an extensive application process and approval by all relevant data custodians and a Human Research Ethics Committee (HREC).

How is data linked for research?

Creation of Linkage IDs

To allow data about the same person to be linked across different data collections, data linkers within a Data Linkage Unit (DLU) create unique Linkage IDs (a random string of numbers and letters).

To do this, the data custodians provide the personal information portion plus the local Record ID of each record in their data collections to the DLU. The data custodian requires approval from an HREC before providing the data. The other portion of the record containing the health, education or other data (known as content information) remains with the data custodians, meaning that the data linkers never have access to this data.

Upon receiving the personal information and Record IDs at the DLU, the data linkers assign a Linkage ID to each person. These Linkage IDs are stored on secure computer servers and can only be accessed by authorised DLU staff.

Data custodians provide regular updates of the personal information and Record IDs to the data linkers. The data linkers then check the new data against the existing personal information to see if they already have Linkage IDs for these records using a statistical probability method.

For each record that is determined to be for a new person in the system, the data linkers create a new Linkage ID which is then added to the DLU's Linkage ID collection.

Provision of linkable data to researchers

Researchers wishing to access the data that data custodians hold must undergo a stringent application process requiring approval from each data custodian and also from a HREC that certifies that the study is valid and in the public interest.

Once a project is approved, the data custodians and the staff at the DLU work together to determine which records are required for the study to ensure the minimum amount of information is provided to the researcher. The data linkers then use the Linkage IDs to create Project Linkage IDs that are specific for the approved study. They then send the Project Linkage IDs along with the Record IDs of the required records to the data custodians.

Using the Record IDs, the data custodians extract the required records from their collections and replace the personal information of each record with its matched Project Linkage ID. The researcher is then provided with the content data of each record and its corresponding Project Linkage ID by each data custodian.

Using the Project Linkage ID, the researcher can determine which records from different datasets belong to the same person without having access to the personal information in order to create a merged dataset for their analysis.

The access and use of research datasets is strictly controlled and managed. For more information on the PHRN's Privacy and Security measures

[click here.](#)