Data linkage

What is data linkage?

Data linkage is the connection of records within and between data sources that are thought to relate to the same individual.

Data linkage is managed by data linkage units across Australia. If you want to access linked data in your research, you'll need to work closely with the relevant data linkage unit. No matter which one you work with, there are several phases in a data linkage journey.

Data sources

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 go to school, visit a hospital, when you get married or when you have a baby. This data is collected by different organisations.

These records typically contain information about our identity (such as name, address and date of birth), diagnosis and treatment. These records are captured in separate systems and data linkage is required to connect the information for research and service planning.

It's a complex process

Highly specialised computer software handles most of the matches, but data linkage units are often required to assess whether the data matches are accurate. With millions of data points being handled at any one time, creating accurate and secure linkage keys is a complex and involved process.

What happens when the data is linked?

Linkage keys are attached to data required for analysis and supplied to researchers for approved linked data projects. The researchers can then connect linkage keys across datasets to compare and analyse the data without knowing a person's identity.

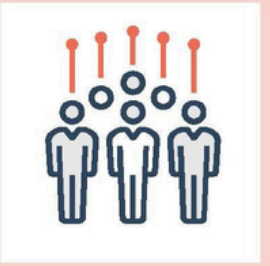
DATA CONNECT AUSTRALIA

NCRIS National Research Infrastructure for Australia An Australian Government Initiative

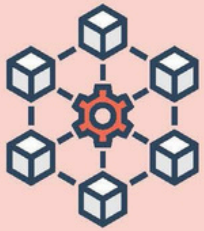
Linking data

Data is linked via a four-step process.

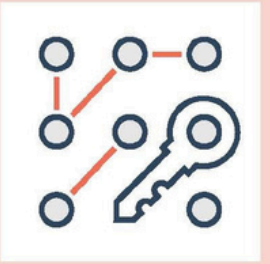
1 Identifiable data from a dataset is received, cleaned and standardised, so it can be read consistently by a computer program.



2 Identifiable details from the data such as name, address and date of birth are loaded into a table in a database that allows for broad comparisons to be made.



3 Data is run through a computer program that compares details from multiple datasets. This process can be customised depending on the data involved.



4 When a match is found across records, a 'link' is made in the form of a unique identification code. This unique code is called a 'linkage key'. Linkage keys are extracted and encrypted for approved research projects.