



Australian Government
Australian Institute of Health and Welfare



Data linkage
services for
clients



What are 'data linkage' and 'data integration'?

Data linkage, data integration and 'record matching' are all terms referring to the same process.

Governments and other organisations keep various separate databases for their own purposes. These databases could contain information on, for example, people admitted to a particular hospital, participation in education or employment programs, what medicines have been dispensed to whom, who has received a family payment and so on.

Data linkage is a powerful process for identifying multiple appearances of people within one database and for combining information about people across different databases of the kinds listed above while still preserving privacy (and also subject to ethical approval and the permission of the data owners).

In this way it is possible to tell a bigger story than would be possible from one database alone. For example, we recently linked data on women who received the human papillomavirus (HPV) vaccine with separate cervical screening data. This provided new information on the effectiveness of the vaccine in reducing cervical abnormalities without the need for a new and costly information collection to provide this data.

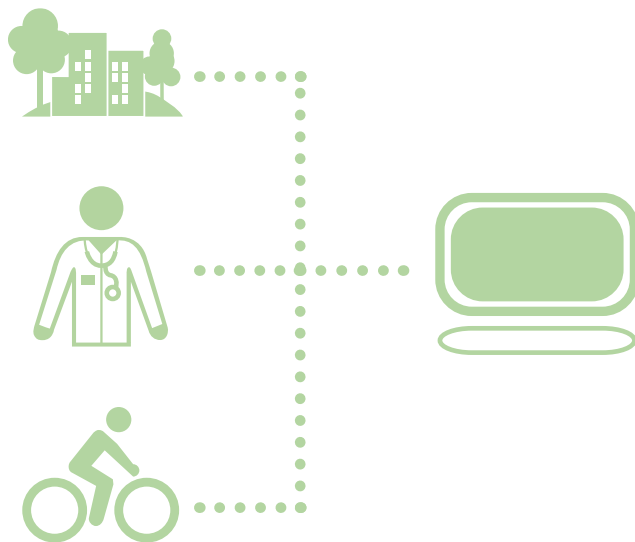
Data linkage techniques can be used in other interesting ways—for example in understanding recording of Indigenous Australians in various data collections, especially if they have differing Indigenous status in different datasets.

Advantages of data linkage to clients

Modern data linkage has enormous potential for providing new insights into people's health and wellbeing that would otherwise be difficult or expensive to obtain.

It can increase the value and use of existing data holdings at low cost by providing new information without the expense or intrusion of new data collections.

The new insights and dimensions revealed through data linkage can in turn drive the development of new, relevant and insightful policies and practices that make a real difference to the lives of Australians.



AIHW linkage services

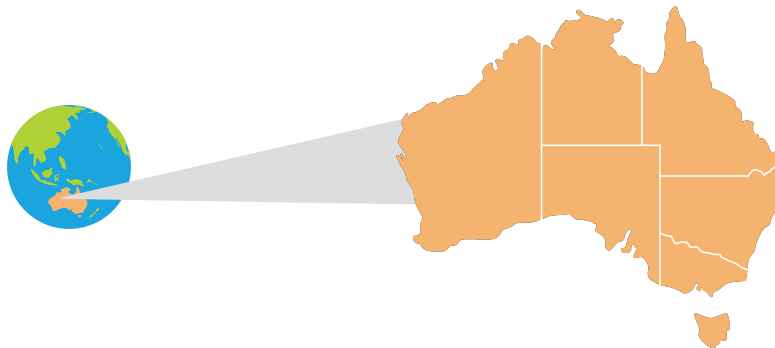
Providing data linkage assistance to researchers and governments is a core service of the Australian Institute of Health and Welfare.

We 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.

We have well-established expertise across a range of linkage techniques and are regarded as international leaders in this field.

We are also currently one of only two accredited Commonwealth Integrating Authorities in Australia. Clients wanting to undertake high-risk research projects involving the use of Commonwealth datasets **MUST** use an accredited Authority.

As part of our service we can potentially link to the several national databases we maintain at the Institute. Two of the largest are the **National Death Index**, which contains information on all deaths and their causes in Australia since 1980, and the **Australian Cancer Database**, which contains information on all cancer notifications since 1982.



Why hire the AIHW?

Security, privacy, best practice

To secure accreditation as one of only two Commonwealth Integrating Authorities we had to meet extremely stringent criteria covering project governance, capability, data management, security, and the protection of privacy and confidentiality.

A few of the many measures involved include: use of the 'separation principle' to ensure privacy—where personal information is stored in separate files from transactional information; a physically secure area with access available only to data integration staff; a separate secure network; all access logged; integrated data confidentialised; and data encrypted as part of the archiving process.

In carrying out its data linkage activities the AIHW abides by the National Statistical Service (NSS) *High level principles for data integration involving Commonwealth data for statistical and research purposes* and *Best practice guidelines*.

The AIHW Ethics Committee

The AIHW Ethics Committee plays a central role in data governance at the Institute, which includes approval and oversight of the ethical acceptability of any data linkage projects, and the release of any ensuing data in compliance with the terms of the Privacy Act and the strict confidentiality provisions of the AIHW Act.



Expertise, experience, networks, trust and value-for-money

Other key advantages to using the AIHW are:

- **Expertise**—we are the leading health and welfare statistics agency in Australia, with an enviable track record of over 2,500 health, welfare and housing reports over the last 25 years.
- **Data linkage experience**—we carry out 50–80 linkage projects a year of varying size, complexity and scope. We were the first organisation, for example, to undertake data linkage between hospitals and residential aged care, and between child protection, juvenile justice and homelessness services.
- **Extensive networks**—the AIHW has a strongly collaborative approach to its work, with strong relationships with the Commonwealth, and state and territory governments, universities and other expert groups, national committees, and with the private sector.
- **Access to data**—because of our reputation and our networks we have an enviable record in being able to arrange access to a wide variety of useful government and non-government datasets for data linkage purposes. As both a national ‘node’—or linkage centre—within the Population Health Research Network, and as an Integrating Authority, we are uniquely placed to enable safe and secure linkage to data from state and territory databases as well as to Commonwealth data. And we also hold some of the most useful large health, welfare and housing databases ourselves. This can still take time but we work with data custodians to enable safe use of their data for important research.



- **Trust**—our extensive and effective data governance framework helps ensure that we deliver on our mission of providing authoritative information and statistics to Australians while also complying with our legal, regulatory and governance obligations in acquiring, handling and releasing data.
- **Ethics**—we work with an experienced, effective and independent AIHW Ethics Committee with efficient online application and approval processes (all data linkage projects need ethical approval).
- **Value-for-money**—our rates are comparable to our private and public sector competitors, and we have been carrying out projects under contract since we were established in 1987.

The data linkage process

There are several different variations of data linking, but broadly the process is as follows.

With sophisticated data linkage tools, the requisite computing power, statistical and data management skills, and a totally secure environment, people's names and other personal details from one database can be matched with their names and details in one or more other databases, even if the two databases differ in the way they record people's details (for example surname followed by initials, or full name with surname last). These processes can be used to link databases containing millions of records.

Sometimes this work may involve the use of a 'statistical linkage key', which is developed for each person or entity. This might, for example, consist of the second, third and fifth letters of the surname plus the second and third letters of the person's first name, sex and date of birth. This supports data linkage without the need to release personal identifiers.

The services we offer

We offer a range of services, depending on the needs of your project.

In broad terms we undertake all the linkage work to produce a linked dataset.

Services we offer include:

- Help and advice with project conception and design.
- Investigation of data linkage and analytical methods.
- Negotiating access to suitable datasets, while maintaining strong confidentiality and privacy protections.
- Assistance with approvals (data use, ethical acceptability).
- Routine data linkage to our own datasets (especially the National Death Index and Australian Cancer Database).
- More complex data integration, including those using other datasets not necessarily held by the AIHW.



- Undertaking data integration involving Commonwealth data, for high-risk research projects—as one of only two agencies permitted to do so.
- Data cleaning and standardisation services to enhance data linkage and quality.
- Ongoing secure data storage and management.
- Statistical analysis, and report writing and production services using linked data, as well as other datasets.

Different data access solutions can be arranged to suit the requirements of the data owner and the analytical needs of the client.

Client researchers could, for example, manipulate the resulting de-identified linked data themselves (in an approved secure environment) in order to produce their own analyses and reports—or the AIHW could also do that work on behalf of the client.

For some approved projects it may be possible to release linked data directly to researchers for analysis and other approved purposes.



Project examples

At the AIHW, data linkage projects come in all sizes, grades of complexity and scope, on a very wide variety of topics related to the health and wellbeing of Australians.

A 'routine' health data linkage project might involve matching people who received treatment 'X' with subsequent hospital attendances or how long they lived after treatment.

At the other end of the spectrum it could be possible to link various health, welfare, education and income and support payments datasets to show how to get better outcomes and value from health and welfare programs and expenditure.

Examples of projects under way or completed at AIHW include:

- ***Pathways in Aged Care***—this linked dataset covers aged care assessments and use of 7 different Commonwealth aged care service programs from 2002 to 2011, as well as deaths. The study showed that even where people are eligible for residential aged care they prefer to remain in their homes for as long as they can.
- ***Cancer risk in people exposed to computed tomography scans***—Medical Benefits Schedule data on 680,000 CT scans were linked to the Australian Cancer Database to examine if there was increased incidence of cancer after CT scan exposure. The study showed exposure to CT scans in childhood increased the incidence of cancer.



- ***Impact of a population-based HPV vaccination program on cervical abnormalities***—this collaborative study between the Australian Institute of Health and Welfare (AIHW) and the Victorian Cytology Service (VCS) linked the National HPV Vaccination Program Register with Victoria’s Pap test Register. The study was the first in the world to show a population-based HPV vaccination program resulting in a fall in cervical abnormalities within 5 years of implementation.
- ***Dementia care in hospitals***—this study, commissioned by Alzheimer’s Australia, used linked data to investigate the experience of people with dementia in the New South Wales hospital system. It showed that people with dementia stay in hospital longer than other patients and have higher costs of care, and that the condition is often not recorded in patient records.
- ***Diabetes Care Project***—this 3-year pilot study was conducted by McKinsey and Company, with AIHW being contracted to coordinate the data linkage components across a range of health data sources. The study evaluated whether new models of care deliver better-quality outcomes for people with diabetes.
- ***Homelessness, income support and employment pathways***—we are undertaking a project funded by the Department of Social Services to link Specialist Homelessness Services client data with income support and employment services data from the Commonwealth Departments of Human Services and Employment. The study will to identify the service delivery patterns and pathways used by clients.
- ***Child Protection data and educational achievement***—this project involves linking Child Protection data with educational achievement data as measured by NAPLAN (National Assessment Program—Literacy and Numeracy) testing.
- ***Child immunisation and health***—two projects we are undertaking in 2014–15 involve linking child immunisation and health outcomes data.

Further information



Visit the data linkage pages on the
AIHW website at



www.aihw.gov.au/data-linking/

Contact the AIHW Data
Integration Services Centre at



linkage@aihw.gov.au

Printed 2014.

Photos of building interior by Ben Wrigley.

