

Factsheet 4 – Information about Record Linkage

What is Record Linkage?

Record linkage allows researchers and policy makers to study trends and patterns in whole population groups, by bringing together routinely collected information from different sources. The method links together an individual's records from different agencies in a way that protects anonymity (that is, without personally identifying any individual). It provides a safe and secure way to bring together relevant information that enables researchers to examine relationships between events at a population (not an individual) level.

Record linkage is often used to answer research questions that cannot be studied accurately in smaller samples (where outcomes of interest may be rare, and therefore require large numbers of individuals for them to occur at all), or when use of self-reports in interviews or questionnaires presents problems of feasibility or accuracy. For example, people can't accurately remember events that happened a long time ago, but records that were made at the time of the event (e.g., at a child's birth) are likely to be more accurate than trying to remember detailed information years later. Using record linkage, health services information for an entire population can be combined with information from other departments or agencies, such as education, to study questions that could not be answered accurately using any other method. For example, researchers could see whether birth weight has an influence on children's readiness to learn at school, or whether better social and emotional functioning at school entry is related to later scholastic achievement.

Record Linkage can also bring together a child's records with those of his/her parents to provide information about important influences in a child's life. For example, to find out how significant events in the parents' lives, such as hospitalisation for serious illness or a court appearance, might impact on children's health and wellbeing. Collection of information about these possible influences on the lives of children could not be gathered accurately for a population cohort in any other way. We are fortunate that the NSW government has provided the infrastructure to enable this research to be undertaken in a way that protects the privacy of the people involved.

To see an example of the process of possible linkage of the Middle Childhood Survey (MCS) data to other datasets visit [An example of Record Linkage](#).

Who gives permission for records to be linked?

Before a record linkage can be undertaken, multiple approvals are required to ensure that all Commonwealth and State privacy and security regulations and laws are upheld, that the research is of significant scientific merit (that is, it asks important and relevant questions), and that there is no risk that an individual may be personally identified.

The following approvals must be obtained before a record linkage project can be undertaken:

- An authorised Human Research Ethics Committee (e.g., the NSW Population and Health Services Research Ethics Committee) MUST approve the linkage project, and will set an expiry date by which the record linkage project must be completed;
- Data custodians within the organisation responsible for each dataset MUST approve the use of their data in the linkage project;
- The Record Linkage Authority (e.g., in NSW this is the Centre for Health Record Linkage) MUST approve the feasibility of the linkage project;
- All government-owned data MUST be approved for linkage by the relevant Department's own ethics committee.

How is Record Linkage done?

In Australia, there are several third-party record linkage providers for data owned by State/Territory or Commonwealth Agencies. One of these linkage providers is the [Centre for Health Record Linkage \(CHeReL\)](#). This is the agency in NSW that provides linkage services for the NSW Child Development Study (NSW-CDS). It is important to note that the anonymous linkage of data requires cooperation from the CHeReL in direct liaison with Data Custodians within the organisations that are providing data for linkage, as follows.

For each research linkage study, the CHeReL receives a set of personal identifiers (information such as Name, Date of Birth, Postcode) from each Data Custodian for all the individual records to be linked. Using these sets of personal identifiers from different datasets, the CHeReL computer finds personal identifiers that 'match' (that is, belong to the same person), and generates a new, Project-specific Linkage ID code ('Linkage IDs') for these 'matches'. The CHeReL pairs Linkage IDs with the unique 'Record ID' in each Data Custodian's original file, and then sends these paired Linkage IDs and Record IDs (for all matched cases) back to the Data Custodians of each organisation. This is the end of the process for the CHeReL.

Then, each Data Custodian uses the new Linkage IDs to extract the actual data (the information needed by the researchers) from their database, and removes the Record IDs from the research data files, leaving only the research data and the Linkage IDs. Each Data Custodian then sends the researcher the Linkage IDs and their corresponding record content (data) **without any personal identifying information**. This is the end of the process for the Data Custodian.

Next, the Research Team receives from each Data Custodian the original data and Project-specific Linkage IDs, and links together data from different sources (e.g., Education, Health, etc.) that relates to the same person, but without being able to identify anyone personally. For further information about how the CHeReL operates record linkage, [click here](#).

How is privacy protected?

The research databases for the NSW Child Development Study will not contain any information that could identify a child, parent/caregiver, or school. Only **non-identifiable data** will be provided to the NSW-CDS Research Team.

The Middle Childhood Survey (MCS) questionnaire responses (data) and personal identifying information of participants in the MCS will be stored separately (held by separate organisations), so that no individuals in the MCS can be identified. The MCS data will be coded by a unique MCS-ID number. Only the researchers conducting the study will have access to the data collected during the MCS, and the University of New South Wales will be the Data Custodian of MCS data. On behalf of the University of New South Wales, the CHeReL will hold only the MCS-ID and personal identifiers (without having any access to the MCS data).

The Record Linkage processes ensure that researchers cannot re-identify study participants, because the researchers do not hold the personal identifiers for the research participants – they access only the research data that is coded by the unique MCS-ID number. Staff at the CHeReL will only access the personal identifiers associated with each MCS-ID number, but have no access to research data.

For more information, contact the NSW Child Development Study free helpline on **1800 901 922**.