



An Australian Government Initiative



Our Children
Our Communities
Our Future

Australian Early Development Census **Data Linkage Policy**

June 2014



Contents

1	About this Policy	3
2	Introduction to Data Linkage for research purposes	3
2.1	Why link data?.....	4
3	Privacy.....	4
4	Introduction to the AEDC.....	5
4.1	AEDC Domain Scores	5
4.2	Demographic Data	6
5	Governance	6
6	Costs	7
7	Access	7
7.1	Application for access.....	7
7.2	Upon receiving AEDC microdata	8
8	Acceptable Linkage Models	8
8.1	Incorporation of AEDC linkage variables into master linkage files.....	8
8.2	Creating de-identified, linked datasets without using a master linkage file	8
8.3	Using a third party linkage facility	8
8.4	Source institution doing the linkage	9
8.5	Linkage with consent.....	9
8.6	Authorised Third Party Linkage Facilities	9
9	Storage and Security	9
9.1	Security during data transfer.....	10
9.2	Breaches.....	10
10	Restrictions on Use and Publication.....	11
10.1	Inappropriate research topics	11
10.2	Publication.....	11
11	Retention or Destruction	11
11.1	Identified and identifiable data	11
11.2	De-identified data	11
12	Transparency and Openness – public reporting	12
13	Complaints.....	12
14	Glossary	13
15	Appendix 1: Assessing Data Linkage Applications	15

1 About this Policy

The Australian Early Development Census (AEDC) aims to help improve early childhood development outcomes of all Australian children by providing communities, governments and policy makers with the information they need to plan and evaluate their efforts in achieving outcomes for children. It is therefore desirable that a large quantity of AEDC data be made available that is suitable for a range of purposes.

In all AEDC data releases the challenge is to ensure that maximum use can be made of the data without compromising the integrity and privacy of information about individual children, and community confidence in the AEDC program. The principles and protocols governing AEDC data release are set out in the *AEDC National Implementation Data Protocol* ('the *AEDC Data Protocol*').

Accordingly, AEDC data will be made available for linkage subject to approved projects, specifically for statistical and research purposes. Access will be subject to appropriate approvals, exemptions and the availability of suitable linkage infrastructure. This policy makes clear the governance arrangements and required protocols for making AEDC data available for data linkage.

2 Introduction to Data Linkage for research purposes

Data linkage is the bringing together or linkage of records of an individual, household, business unit or other entity from either within or across a number of datasets. There are two main methods by which this is done:

1. 'Deterministic' linkage – involving the exact, one-to-one character matching of linkage variable(s) across two or more data collections, e.g. an identification number common to each collection.
2. 'Probabilistic' linkage – records are compared and if a number of variables from each record are deemed to be the same, then those records are considered to belong to the same person or entity and be included in the combined data file.

Data linkage is undertaken for three main reasons:

1. Individual client management/clinical purposes. The records of an individual are linked to examine the details relating to an individual, for example linkage of records between hospitals to enable more effective care provision.
2. Audit and fraud control purposes, i.e. linkage between Centrelink and the Australian Tax Office.
3. For statistical, research and policy purposes. The records of individuals are linked to gain a better understanding of patterns of service usage, the effect of care, the effectiveness of services and interventions or health issues affecting groups, not individuals.

Other terms used to describe the process of data linkage include record linkage, data matching, record matching, data integration and record integration.

When data are linked for statistical purposes, individuals are identified only to enable the link to be made. When the linkage is complete, the identity of the individual is no longer of any statistical interest. The linked dataset is used to report statistical findings about the population or sub-populations.

In contrast, when data are linked for client management/clinical and audit/fraud control purposes, individuals are identified not only to enable the link to be made, but also for administrative use or treatment subsequent to the linkage.

AEDC data can only be linked for statistical, research and policy purposes. Linked data are not to be used for individual client management/clinical purposes, or for audit and fraud control purposes.

2.0 Why link data?

Linking existing datasets leverages more information from the combination of individual datasets than is available from the individual datasets taken separately. They have the potential to:

- provide new insights for policy, highlighting key issues for government response
- provide information on the short and long term impacts of policy across Australian businesses, population groups and geographic areas
- assist in costing new government based programs
- assist in simulating the impact of changes in policy parameters in government programs
- provide insight into the transitions of population groups between government support programs over time and identify factors that influence these transitions.

In addition, linked datasets provide public benefits in terms of an expanded range of official statistics, improved research, and a better informed community.

3 Privacy

As stated in the *AEDC Data Protocol*, the AEDC data is owned by the Australian Government (currently represented by the Department of Education). Variables collected as part of the AEDC dataset make the identities of individuals in the dataset apparent, or are such that the identities of these individuals could be reasonably ascertained. This means that the data is personal information and is subject to the principles of the *Privacy Act 1988*.

In addition the *AEDC Data Protocol* specifies that the privacy of teachers and children are to be maintained in accordance with national, state and territory privacy legislation and policies. Any entities seeking to undertake research on a linked dataset containing AEDC information will need to show they are operating in accordance with the relevant legislation, including any requisite exemptions to use the AEDC data for a purpose secondary to the original purpose of collection (i.e. to undertake the linkage).

To further maximise the privacy of the children represented in the AEDC dataset, any linkage projects will need to use protocols separating the variables required for linkage from the final dataset (Kelman, Bass and Holman 2002)¹. The key features of these protocols are:

- Only variables required for linkage (mainly demographic variables) are provided to the individuals undertaking the linkage.
- Once linkage is undertaken, these demographic variables can be deleted by the

¹ Kelman, C.,W. Bass, A.J. Holman, C.D.J. (2002) 'Research use of linked health data – a best practice protocol', *Australian and New Zealand Journal of Public Health*, vol 26 (3), pp 251-5

individuals undertaking the linkage.

- Project specific IDs are generated for each linked record, for each research project.
- The variables required for research are de-identified and provided directly from the data custodians to the researcher.
- The individuals undertaking the research using the linked dataset must not have access to the identified data, or the linkage variables used in the linkage process, to protect the privacy of children in the AEDC dataset.
- The de-identified linked dataset is destroyed once the objectives of the research are met (or a retention plan, with scheduled reviews, is agreed).
- The output from the research is made public.

4 Introduction to the AEDC

The AEDC is a population measure of children's development as they enter school. It is an adapted version of the Canadian Early Development Instrument (EDI), developed in response to communities' increasing interest in knowing how their children were developing. A population measure places the focus on all children in the community and therefore the AEDC reports on early childhood development across the whole community. It is recognised that moving the focus of effort from the individual child to all children in the community can make a bigger difference in supporting early childhood development.

The Council of Australian Governments (COAG) has recognised the need for all communities to have information about early childhood development, and has endorsed the Australian Early Development Census (AEDC) as a national progress measure of early childhood development. The data will be collected every three years.

The AEDC measures five areas of early childhood development from information collected through a teacher-completed checklist covering:

- physical health and wellbeing
- social competence
- emotional maturity
- language and cognitive skills (school-based)
- communication skills and general knowledge.

4.0 AEDC Domain Scores

AEDC domain scores are calculated based on teacher responses to 104 developmental questions for each child. There are five developmental domains. The processes involved in creating AEDC domain scores have been licensed for use in the AEDC by the Offord Centre for Child Studies in Canada. Domain scores are represented by a number between 0 and 10. A higher domain score indicates a higher level of development within a particular domain.

The macrodata includes responses to each AEDC checklist question for each child and a score for each domain. However, domain scores have been developed for use at a group level and have not been psychometrically tested for application in relation to individual

children.

4.1 Demographic Data

The AEDC microdata also include a range of demographic data allowing for more detailed analysis of particular social and cultural groups. These data were collected using the standard school enrolment process and cover:

- age
- sex
- geographic location in which the child lives
- special needs status
- Aboriginal/Torres Strait Islander status
- English as a second language status
- country of birth.

AEDC demographic variables such as geography, language spoken at home and country of birth have been coded using classifications, such as the Australian Bureau of Statistics Socio-economic Indices for Areas (SEIFA) and Australian Standard Geographical Classification (ASGC).

5 Governance

The *AEDC Data Protocol* states that the Australian Government is the owner of the data (currently represented by the Department of Education). The Social Research Centre (SRC) acts as manager and host of the AEDC data.

Governance is actualised by the AEDC National Committee. This National Committee is chaired by the Department of Education and membership is documented in the *AEDC Data Protocol*. Among other functions, the Committee approves and oversees the administrative processes to facilitate the linkage of AEDC data to other relevant data.

As the owner of the data, authorisation is required from the Department of Education in order to approve the release of data for data linkage.

Research projects using linked data are required to nominate a body, referred to in this Policy as the Integrating Authority, which is ultimately accountable for the data linkage aspect of the project, leading it through its approval and implementation. This body will undertake the following functions:

- Ensure appropriate governance is in place, an open approval process is followed, including documentation of the proposal, the impact on privacy, the expected costs and benefits and the outputs.
- Be responsible for the ongoing management of the de-identified linked dataset, ensuring it is kept secure, confidential and fit for the purposes for which it was approved.
- Be responsible for initiating and managing the regular review of the project (where it has been agreed that the project will be ongoing).

6 Costs

There are costs involved in linking data. They include the costs involved in creating the linkage keys and Data Custodians may choose to charge for data extraction (for providing data for the creation of linkage keys and de-identified data to researchers). The AEDC Data Manager, SRC recovers the costs associated with producing the customised microdata file after the linkage process.

The Department of Education will not pay any costs charged by other stakeholders in research projects linking AEDC data. The applicant will need to negotiate the payment of the Integrating Authority and any relevant Data Managers or Custodians.

7 Access

7.0 Application for access

Chief Researcher(s)/Investigator(s) may apply to undertake a single research project or multiple research projects (a set or family of research projects) using an AEDC linked dataset.

Chief Researcher(s)/Investigator(s) may specify one or all research projects in their initial AEDC Data Linkage Application Form when they seek approval for the creation of the AEDC linked dataset. However, if a researcher wants to do research using an AEDC linked data set, already created for their use, for an approved AEDC data linkage project they may request approval to use that data set rather than incur the cost of having a new AEDC linked data approved and created.

Initially, the integrating authority and researcher should contact the AEDC Data Management Team to discuss the proposed research project/s with a view to ensuring the project/s is within the scope of acceptable research uses of AEDC and that AEDC linked data is suitable to address the aim of the project/s and any associated research question/s. The researcher must obtain ethics approval from a Human Research Ethics Committee (HREC) for the linkage of AEDC data, unless the provision of data for the purpose of the project is authorised or required by legislation. In addition the researcher is responsible for undertaking any exemptions from the relevant legislation to use the AEDC data for a purpose secondary to the original purpose. Once these approvals and/or exemptions have been obtained, the researcher in conjunction with the Integrating Authority needs to complete an *AEDC Data Linkage Application Form*.

The *Data Linkage Application Form* in conjunction with HREC approval is submitted to AEDC Data Management Team for assessment ensuring that the data being sought and the way it will be used will meet the desired outcomes of the applicant and the information being sought will not breach any privacy provisions. The application will then be considered by AEDC National Committee.

The researcher will also need to complete an Australian Early Development Census Individual Deed of Licence for Australian Researchers or Organisational Deed of Licence available from AEDC Data Management website (AEDCdata.com.au).

The AEDC National Committee, through its Secretariat, further assesses the application to ensure that it meets the principles of the *AEDC Data Protocol* and approves or rejects the application. Criteria to be used to assess the application are in Appendix 1.

7.1 Upon receiving AEDC microdata

Any researchers that are provided access to AEDC data (Individual Users) are required to sign an Individual Deed of Licence or if the researcher works for an organisation, an Organisational Deed of Licence may be signed by the Organisation's Delegated and Data Manager.

8 Acceptable Linkage Models

There are three acceptable linkage models.

1. AEDC linkage variables supplied to data linkage facilities to create links with master linkage files.
2. The creation of de-identified, linked research datasets – either using a third party linkage facility, or within an institution that is supplying data for linkage but the actual linkage occurs behind a firewall to maintain the separation of linkage variables and content variables.
3. Linkage with consent.

8.0 Incorporation of AEDC linkage variables into master linkage files

There are several established linkage facilities in Australia which have created, or in the process of creating, master linkage files. These files are mappings or concordances between the local identifiers used by each data custodian represented in the master linkage file and the purpose created master linkage keys allocated by the linkage facility and which are consistent for individuals across all of the datasets involved. These linkage facilities do not hold service or event data (content variables); this infrastructure provides the capability to create de-identified linked datasets subject to appropriate approvals being obtained. Current examples include the Western Australia Data Linkage Unit, SA-NT DataLink and CHeReL (NSW).

To create the linkages between the AEDC dataset and a master linkage file, identifying variables need to be released from the AEDC dataset to the linkage unit. This is subject to the *Privacy Act 1988*.

The identifying variables only need to be provided to the linkage facility to create the linkages. Once the linkages are created, any research projects that propose to use de-identified, linked datasets created using the master linkage file still need to obtain approval from the AEDC National Committee, but do not need Public Interest Determinations.

8.1 Creating de-identified, linked datasets without using a master linkage file

An Integrating Authority may represent one or more institutions proposing to create a de-identified, linked dataset using AEDC data and data held by the institutions. In these circumstances, it is critical to maintain the separation of linkage variables and content variables, ensuring that people with access to the identifying variables (undertaking the linkage) do not provide information that may identify an individual to the people undertaking the research on the resulting de-identified linked dataset. This can be effected by using a third party linkage facility to undertake the linkage. Alternatively one of the institutions contributing data can undertake the linkage but does the linkage in an electronic restricted access environment where analysts do not have access to identifiable variables (i.e. name,

address, date of birth) used for linkage.

8.2 Using a third party linkage facility

This is a similar process to producing a master linkage file and then creating a de-identified linked dataset for research from that master file. The difference is that the linkages are created only once, specifically for the purpose of then creating the de-identified linked dataset.

Once the custom linkage file is created, the links can then be used to create the de-identified linked file for the research project.

8.3 Source institution doing the linkage

If a third party linking facility is not used to create the custom linkage file, an alternative is for one of the institutions contributing data to also undertake the linkage. In this scenario it is critical that the linkage be undertaken in a separate secure server with restricted access to maintain the separation of linkage variables from content variables.

Once the custom linkage file is created in the restricted separate server and project linkage keys (PLKs) assigned, the AEDC content data can be requested to create the de-identified, linked dataset in the restricted access server. Once the de-identified, linked dataset is created it can be moved outside of the restricted server-and utilised by researchers for the approved project.

8.4 Linkage with consent

The third acceptable linkage model is linkage with consent. Researchers who have datasets where parents/guardians have consented to the linkage of their children's' records with other data held about them can apply to link to the AEDC records that relate to the children whose parents/guardians have provided consent.

In this scenario, the identifying information is provided by the researcher to SRC to identify the AEDC records that relate to the children whose parents/guardians have provided consent. These AEDC records can then be provided to the researcher.

8.5 Authorised Third Party Linkage Facilities

A number of linkage facilities have already been deemed to meet the minimum requirements for having capacity to undertake linkage and the procedures, protocols and infrastructure to maintain the privacy of AEDC data. Projects that use these linkage facilities to undertake the linkage do not have to separately demonstrate proof of capacity in these areas. The authorised linkage facilities are listed in the *AEDC Data Linkage Application Form*.

9 Storage and Security

AEDC microdata must be stored in a manner excluding access by anyone except authorised Individual Users within the organisation.

The following are guidelines for storing the AEDC Unit Record Data and results of analyses:

- If identified AEDC unit record data is stored on a stand alone computer, i.e. one not connected to any other computer, it must be kept in a locked room and/or have password protection to prevent access by others.

- If identified AEDC unit record data is stored on a computer network, it must be kept in a directory with suitable restricted access.
- Identified AEDC unit record data, whether stored on a stand alone computer or a secure network, must be managed in a way which is consistent with the separation principle² i.e. no-one can see the identifying or demographic information in conjunction with the analytical data.
- Any stand alone computer on which de-identified AEDC Unit Record Data is stored must be kept in a locked room and/or have password protection to prevent access by others.
- Any information or results stored on a computer network must be de-identified and kept in a directory with suitably restricted access.
- Any printouts or any physical medium containing AEDC Unit Record Data (such as a CD-ROM provided by the AEDC) must be de-identified and kept in a locked room or secured in a locked cabinet when an individual User is not present.
- Datasets and printouts that contain or reveal AEDC Unit Record Data must be de-identified and tracked so that they can be destroyed when they are no longer needed.

Printouts can be destroyed securely by shredding. Organisations must also provide for a secure way to delete computer files so that they cannot be recovered by an unauthorised person.

These restrictions apply to AEDC Unit Record Data and to outputs from analysis that potentially reveal information on individual children.

9.0 Security during data transfer

When AEDC unit record data is in transit, appropriate security measures must be used to ensure the only people who have access to the data are authorised. This may include (but is not limited to) use of encryption software.

When AEDC data containing identifying variables is in transit, a record of information exchange, release authorisations and recipient acknowledgements must be maintained to ensure the passage of the data can be tracked through each step, e.g. AEDC National Committee authorising the data to leave SRC, record of the data leaving SRC, receipt of the data by the linkage unit (or Integrating Authority). This record is maintained by SRC.

9.1 Breaches

Breaches of an Individual Deed of Licence or Organisation Deed of Licence will result in that individual's access to the data being withdrawn and their organisation's access being withdrawn. In addition, **all** breaches should be referred to the Department of Education's legal area and to the Department of Education Privacy Contact Officer.

² The separation principle is one mechanism to protect the identities of individuals and organisations in datasets. The separation principle means that no-one can see the identifying or demographic information, used to identify which records relate to the same person or organisation (e.g. name, address, date of birth), in conjunction with the content data (e.g. clinical information, benefit information, company profits). Instead, staff can see only the information they need to do the linking or analysis. So, rather than someone being able to see that John Smith has a rare medical condition, or the profits earned by Company X, the person doing the linking sees only the information needed to do the linking (e.g. John Smith's name and address) and the analyst just sees a record, with no identifying information, showing that a person has a rare medical condition together with any other variables needed for analysis (e.g. broad age group, sex).

10 Restrictions on Use and Publication

10.0 Inappropriate research topics

The AEDC National Committee will not approve any research that involves:

- the creation of school league tables
- the identification of individuals, including:
 - those with specific learning disabilities or developmental delay
 - making recommendations, on an individual basis, of students to be provided with special assistance, held back a year etc.
- usage as an individual diagnostic tool, including scoring individual children as developing well or being developmentally vulnerable
- reflecting the performance of a school or the quality of teaching

These topics are inappropriate uses of the AEDC dataset and may lead to misinformed analysis and decisions.

10.1 Publication

It is a condition of accessing AEDC data for data linkage for statistical or research purposes that the results of the project be published after clearance by the AEDC National Committee.

Information derived from AEDC Unit Record Data may only be published in such a way as to ensure that data relating to individual children is not disclosed and that the integrity of the AEDC program is not compromised by any inappropriate use.

A copy of all publications derived from analysis of AEDC data must be provided to the AEDC National Committee at least one month prior to the proposed publication date as per the *AEDC Data Protocol*.

There is scope for exemptions to this process to be granted on a case-by-case basis, with the agreement of the Data Custodian (the Commonwealth Department of Education) and affected jurisdictions, after consultation with the AEDC National Committee.

11 Retention or Destruction

11.0 Identified and identifiable data

Once the identified data (linkage variables) have been used to facilitate data linkage, then the identified data may be destroyed. Confirmation of this destruction must be sent to SRC as per the Microdata Application form and detailed in the disposal of AEDC data section.

The exception to required destruction is the creation of enduring master linkage keys by incorporating AEDC identifying variables into a Master Linkage File. In these situations, a retention plan is required with review periods set at a minimum of every three years. Review is undertaken by application to the AEDC National Committee.

11.1 De-identified data

Once the approved purpose of the project, or set (family) of research projects, is met, the de-identified linked dataset (and any related datasets) will be destroyed, unless a retention plan is agreed to by the AEDC National Committee. Only retention plans with specified review

points by the AEDC National Committee will be considered. Reviews will be no more than three years apart.

Confirmation of destruction is sent to SRC.

Archiving of linked data is restricted to confidentialised datasets (subject to an approved retention plan).

12 Transparency and Openness – public reporting

Stakeholders and the community are kept informed of linkage to AEDC data by publication on the AEDC Data Management website. The Department of Education will publish on the website appropriate details of the project.

13 Complaints

Any complaints received in relation to data linkage, and any resolutions to these complaints, will go through a process of being communicated to the researcher's HREC, the AEDC National Committee, the Royal Children's Hospital (RCH) HREC, Melbourne and Princess Margaret Hospital for Children Ethics Committee, Perth.

A complaint about AEDC data linkage may come from a:

- data subject, that is a person who is concerned at the actions of a researcher in relation to their or their child's information
- researcher who has sought approval to link AEDC data.

If a complaint is made by a data subject regarding the actions of a researcher and data linkage, the researcher is to notify their organisation's HREC, as well as the AEDC National Committee (who will notify the RCH HREC and Princess Margaret Hospital for Children Ethics Committee of the complaint). The researcher is required to have their organisation's HREC approval for linkage to the AEDC data and this HREC will have its own procedures to handle complaints.

If a researcher makes a complaint regarding data linkage, (for example being refused an application to link AEDC data with other datasets) then the complaint will be communicated to the AEDC National Committee, which may review the decision. The National Committee will inform the RCH HREC and Princess Margaret Hospital for Children Ethics Committee of any complaints received and plans for resolving the complaint.

If a data subject or researcher does not believe that their complaint has been resolved by this process, they may make their complaint to the Department of Education or to the Office of the Australian Information Commissioner.

14 Glossary

Authorised Third Party Linkage Facility

A facility that undertakes data linkage on behalf of one or more stakeholders and has been authorised by the AEDC National Committee as having an acceptable minimum infrastructure and capacity.

Content Variables

Data items or variables that contain information about services provided to individuals or events that happened to individuals. It is highly unlikely that these variables alone could be used to identify an individual.

Data Custodian

The entity that has the legal authority to grant access to data. Throughout this policy the term has been used to more generally refer to institutions or agencies that hold data.

Data Linkage

The bringing together or linkage of records of an individual, household, business unit or other entity from either within or across a number of datasets.

De-identified dataset

A dataset of unit records that has had identifying variables removed from each record.

Identified dataset

A dataset of unit records with identifying variables included in each record.

Identifiers

See Identifying variables.

Identifying variables

Variables that can be used to identify individuals. In most situations this refers to variables like name, address and date of birth. In some situations, other information which reveals rare situations can act as identifying variables. For example, a dataset that allows cross-tabulation of a rare type of cancer by very small geographic areas may reveal the only person in an area with that type of cancer and this could lead to the identification of that person.

Another slightly less intuitive example is when a characteristic is shared by everyone in a category, e.g. all five year olds in an AEDC community are developmentally vulnerable in a particular domain. That immediately identifies **any** five year old in that AEDC community as developmentally vulnerable in that domain.

Individual User

An Individual User is an individual in an organisation who is authorised to access AEDC data and who has signed an Individual Deed of Licence or who is an Approved User under an Organisational Deed of Licence.

Integrating Authority

A body which is ultimately accountable for the data linkage aspects of a project, leading it through its approval and implementation). This body will undertake the following functions:

- Ensure appropriate governance is in place, an open approval process is followed, including documentation of the proposal, the impact on privacy, the expected costs and benefits and the

outputs.

- Be responsible for the ongoing management of the de-identified linked dataset, ensuring it is kept secure, confidential and fit for the purposes for which it was approved.
- Be responsible for initiating and managing the regular review of the project (where it has been agreed that the project will be ongoing).

Linkage File

A file created from subjecting two (or more) datasets to data linkage. The file will contain (but is not limited to) linkage keys connecting the idiosyncratic identification numbers used by the source datasets.

Linkage Model

The method by which data linkage will be undertaken.

Linkage Variables

Data items or variables that facilitate linkage between datasets. These are generally identifying variables that are common across the datasets (e.g. name, date of birth, address). In some cases, service or event data can serve as a linkage variable, e.g. date of discharge from hospital and date of admission to a nursing home.

Master Linkage File

A Linkage File with the aim of creating linkage keys for every member of a given population by creating links between as many relevant datasets as possible. The Master Linkage File manages the links and linking material for all of the input datasets.

Master Linkage Key

Linkage keys allocated in Master Linkage Files.

Probabilistic Linkage

Records are compared and if a number of variables from each record are deemed to be the same, then those records are considered to belong to the same person or entity.

Project Linkage Key

Linkage keys allocated for specific projects and records for a given individual will be allocated different project linkage keys for each successive research project. This prevents the disclosure of Master Linkage Keys and ensures that researchers who access the same datasets for linkage across multiple research projects will not be able to identify individuals across research projects.

Third Party Linkage Facility

A facility that undertakes data linkage on behalf of one or more stakeholders.

15 Appendix 1: Assessing Data Linkage Applications

1. Is this project an appropriate use of AEDC data?
2. Is linkage to AEDC data essential to complete this project?
3. Are all of the requested AEDC data items essential to the completion of the project?
4. Has the researcher obtained HREC approval for the linkage of AEDC data for the project?
5. Upon creation of the de-identified linked dataset, will the linkage of requested AEDC variables with other variables pose a risk to the privacy of individuals in the AEDC dataset?

The following precautionary measures are recommended:

- Only providing higher level demographic and address data e.g. age ranges rather than date of birth;
 - Not releasing certain demographic data items together which when used in combination could increase the risk of identification (e.g. date of birth, sex and postcode);
 - If the researcher is proposing to link with a dataset owned by the researcher, the level of knowledge the researcher has as a result of working with that dataset should be considered. In these cases, greater stringency needs to be applied to maintaining confidentiality of AEDC data; and
 - Considering whether the institution undertaking the research on the de-identified linked dataset has received de-identified AEDC data before (refer to previous AEDC Data Linkage Applications). If the requesting researcher has been previously supplied with de-identified AEDC data, compare the variables requested previously and the IDs previously provided to identify the level of possible commonality with the current application.
6. Will linkage be done by an authorised third party linkage facility? (If yes, go to 10)
 - a. Will linkage be done by an independent third party linkage facility?
 - b. Will linkage be done by a supplier of one of the source datasets?
 7. Does the proposed linkage maintain the separation of linkage variables from content variables?
 - Only variables required for linkage (mainly demographic variables) are provided to the individuals undertaking the linkage
 - Once linkage is undertaken, these demographic variables can be deleted by the individuals undertaking the linkage
 - Project specific IDs are generated for each linked record, for each research project
 - The variables required for research are de-identified and provided directly from the data custodians to the researcher
 - The individuals undertaking the research are different from, and do not communicate with, the individuals who undertake the linkage
 8. Does the institution undertaking the linkage have suitable capacity and infrastructure?
 9. Does the institution undertaking the linkage have suitable security to curate the linkage variables?
 10. Is the retention and destruction plan for the linkage variables satisfactory?
 11. Is the security plan for the curation of the de-identified, linked dataAEDCset satisfactory?

12. Are there plans to make the results of the project public?
13. Is the retention and destruction plan for the de-identified linked dataset satisfactory?
14. Has ongoing retention of the de-identified linked dataset been requested? Is there sufficient reason to retain the de-identified linked dataset?
 - a. Have provisions been made for planned review of the retained dataset?
15. Have the other data custodians consented to the linkage?
16. Has the public good of the project been assessed to outweigh the potential impact on privacy?
 - a. Are the committee(s) that provided this/these assessment(s) properly convened for this purpose?
17. Has there been appropriate consideration of the relevant Federal and State/Territory privacy legislation and exemptions obtained where applicable?
18. Has there been appropriate consideration of other relevant legislation and regulations. Have required approvals and/or exemptions been obtained?