

Data linkage and protecting privacy: a protocol for linking between two or more data sets held within the Australian Institute of Health and Welfare

Scope of the protocol

The protocol described in this document covers only the process of linking data sets held within the Australian Institute of Health and Welfare (AIHW) and its collaborating units (termed 'internal' AIHW linkage). It therefore applies to linkage between data sets held by the AIHW and/or its collaborating units (see Appendix 1).

There are several other circumstances when linkage is undertaken; for example, between AIHW data sets and data sets provided by external researchers, and between AIHW data sets and the National Death Index for routine ascertainment of deaths. It should be noted that the protocol described below is not necessarily appropriate for such projects.

Purpose of the protocol

Increasingly linkage between data sets is being seen as a way to make the most of available data to investigate policy and research issues. Of paramount importance in such linkage is the protection of the privacy of individuals. This policy document aims to:

1. describe the broad context within which linkage work (both actual linking and analysis of linked data) is undertaken within the AIHW;
2. outline a protocol which protects the privacy and confidentiality of individuals whose information is contained in the data sets, during both the actual linking of data sets held within AIHW and the subsequent analysis of linked data within the AIHW;
3. provide people within the AIHW practical guidelines for undertaking internal data linkage.

4. provide a description to explain to external agencies and individuals how linkage work is carried out within the AIHW.

Context: privacy and confidentiality arrangements at the AIHW

Data linkage within the AIHW occurs within the general privacy and confidentiality arrangements under which all AIHW staff work. There are three main aspects to these arrangements.

Requirements under the Australian Institute of Health and Welfare Act 1987

The security of data held by the Australian Institute of Health and Welfare has been of the highest importance since the Institute was established, and the protection of the privacy of individuals is of particular concern when linking between data sets.

The Explanatory Memorandum which accompanied the introduction of the *Australian Institute of Health and Welfare Act 1987* stated: 'An important aspect of the Bill is the provision to protect the confidentiality of personal information given to the Institute. Any publications based on the work of the Institute may not identify an individual (including a deceased person...)' (Parliament of the Commonwealth of Australia House of Representatives 1987b). The Act's confidentiality provisions are contained in Section 29 (reproduced in Appendix 2) which prohibits unlawful disclosure or communication of information held under the Act, and AIHW staff can be prosecuted in a court of law for breaching these provisions.

Furthermore, Section 29(4)(e) of the Act includes the requirement not to disclose 'the whereabouts, existence or non-existence of a document concerning a person'. Newly appointed Institute staff, including those employed on a short-term basis, and staff of collaborating units, are required to sign an Undertaking of Confidentiality as soon as they start work.

The Act establishing the Institute includes strong measures for ensuring confidentiality of data and that use of the data complies with the requirements of data providers (see Appendix 2). Established practices and procedures for handling confidential data within the Australian Institute of Health and Welfare further protect the privacy of individuals. These are discussed below.

AIHW Ethics Committee

The *Australian Institute of Health and Welfare Act 1987* requires the Institute to appoint an ethics committee. The principal responsibility of the AIHW Ethics Committee is to form an opinion of the acceptability or otherwise, on ethical grounds, of activities engaged in by the Institute, or with which it is associated. The committee considers applications:

- by AIHW units and collaborating units to commence new collections of identifiable data under the AIHW Act, or to change the scope or content of existing collections;
- for approval for AIHW units and collaborating units to link and subsequently analyse the linked data sets;
- by external researchers for access to either identifiable or non-identifiable data through linkages of research data with AIHW unit record collections.

The Institute does not permit data in its custody to be linked for client management or regulatory purposes.

The membership of the AIHW Ethics Committee is in accordance with guidelines as specified by the National Health and Medical Research Council (NHMRC), and includes the Director of the Institute (or his/her representative) and eight others appointed to provide a range of expertise and experience in health and welfare research areas, including a representative of Registrars of Births, Deaths and Marriages.

As part of the approval process, those involved with a project – both within and external to the AIHW – sign an undertaking which includes:

1. *The unit record file will not be matched, in whole or in part, with any other information for the purposes of attempting to identify individuals, nor will any other attempt to identify an individual be made.*
2. *The person/organisation will not disclose or release the information to any other person or organisation, except as statistical information that does not identify an individual.*

The full text of the undertaking that must be signed before the Ethics Committee will consider a project is contained in Appendix 3.

Limiting data access

There are a number of internal AIHW policies on data handling and reporting procedures that operationalise the confidentiality provisions of the Act. In particular, over-arching policies are set out in *AIHW Information Security and Privacy: Policy and Procedures* (AIHW 2001).

The cornerstone for handling confidential data is limiting access to data to those who need it. Consequently, staff have access to only those data sets required for their work. This is achieved through a network of data custodians. AIHW data custodians – staff at Unit Head level or Head of a collaborating unit – have responsibility for particular data sets, including ensuring compliance both with confidentiality requirements and with conditions of use set out by data providers.

Permission for access to a particular dataset is given by the relevant data custodian, and only those project officers who need access in order to fulfil their duties within the AIHW or collaborating units are given access. In addition, access to identifying data on the databases is further limited to those officers that specifically need it.

The AIHW Board has endorsed *Guidelines for Custody of AIHW Data* (AIHW 2003) which describe the responsibilities of data custodians and the processes to be

followed in carrying out those responsibilities. It also details the roles and responsibilities of other staff in monitoring data security and confidentiality. Furthermore, the Board requires regular data audits to ensure the soundness of practices regarding security of data.

The linkage protocol

The linkage protocol has three components: signed undertakings so that researchers are aware of their responsibilities, linkage principles which guide the way that linkage is undertaken, and linkage procedures which offer a practical way to implementing the principles.

The linkage undertakings

To further protect privacy, linkage between data sets within the AIHW for a particular project can be carried out only by people who have signed the Linkage Undertaking for that project. Conversely, only those staff members carrying out the actual linkage will sign the undertaking, and other project officers are not permitted to link unit record files. By signing the undertaking, those carrying out the linkage are committed to using processes which ensure that identifying data is not transferred between data sets and that linkage to data sets external to the project does not occur.

The Linkage Undertaking – reproduced in Appendix 4 – is signed as part of the Ethics Committee submission for a project, and contains confirmation by the AIHW data custodians that they are aware of this use of their data on condition of the project gaining Ethics Committee approval. A data custodian for the linked data sets is also identified in this undertaking. As with the general Ethics Committee undertaking, the Ethics Committee is informed in writing as people leave the project, and new staff involved with linking must sign the Linkage Undertaking before starting work.

Within the context of analysis carried out within the AIHW, it may be desirable for reasons of consistency to use an established linked data set for a particular project. The use of such a linked data set for further analysis projects would (like its initial creation) be subject to any terms and conditions agreed with the data providers. In this case, access to the linked data for analysis other than that specified in the original Ethics Committee submission requires a new submission to the Ethics Committee, and agreement from the data custodian of the linked data and the data custodians of the original data sets (which may require consultation with the data provider). The form which accompanies the submission to the Ethics Committee showing such agreement by data custodians is given in Appendix 5.

Data linkage principles

In terms of data-handling, one of the prime ways to preserve privacy is to ensure that identifying information (such as name, date of birth and address) is not transferred from one data set to another. The procedures presented below were developed to ensure that in the course of linking two data sets held within the Institute, such identifying information is not transferred from one set to the other, and is not kept in the linked data set. The procedures were developed with reference to the frameworks suggested by the National Health and Community Services Information Management Groups for statistical linkage in data collections (NHIMG 2002,¹ NCSIMG 2004), and to more recent consideration of data linkage by the Statistical Information Management Committee (SIMC).

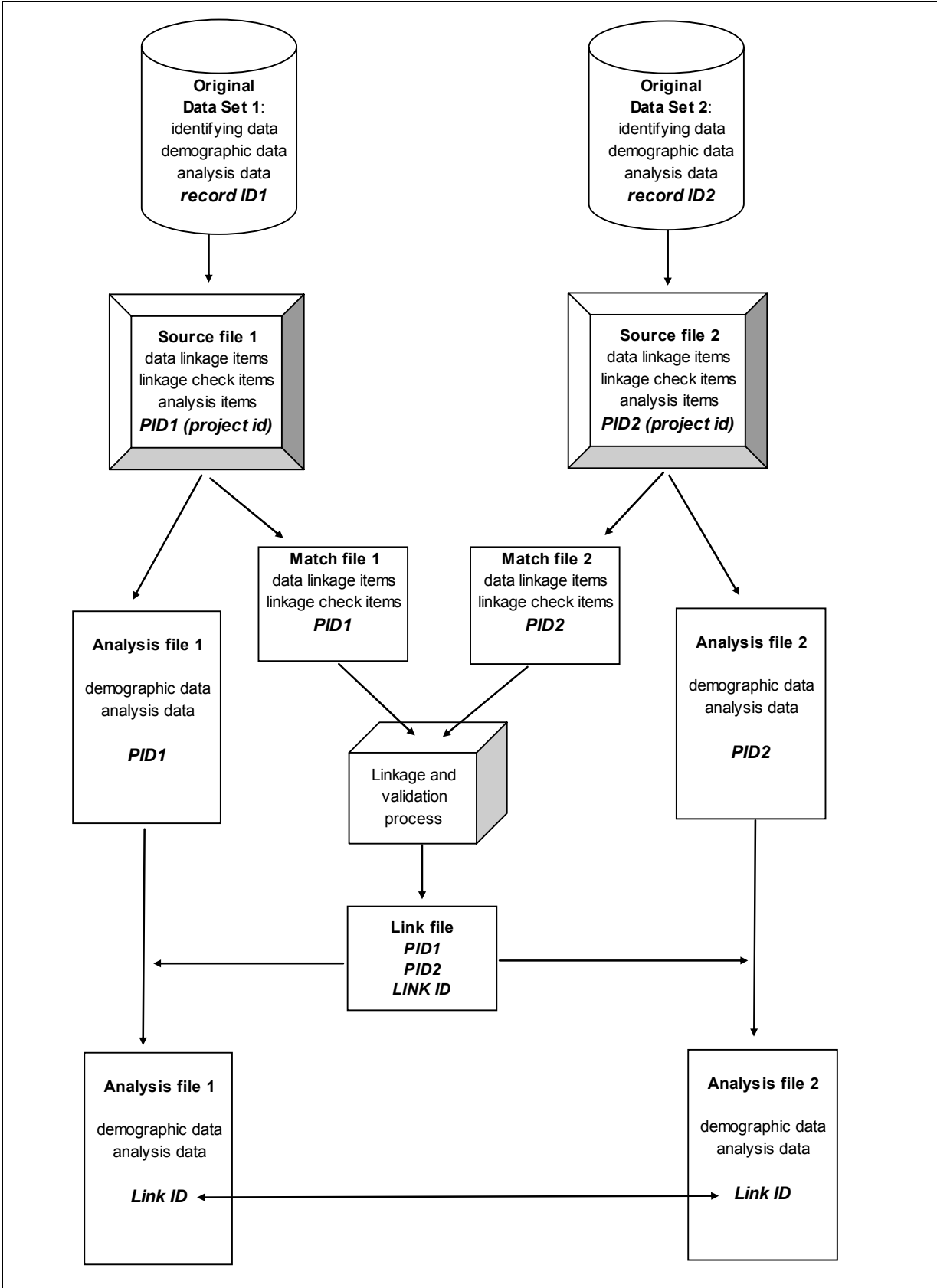
The principles underlying the data linkage protocol for internal AIHW linkage are that:

- Data linkage is not carried out directly between original complete data sets;
- Data linkage is undertaken using purpose-specific linkage data sets that contain only the data required for establishing and validating links;
- Links between data sets are recorded using *project specific* unique record identifiers so that links identified for a particular project (including longitudinal analyses) cannot be used to establish links between data sets outside the scope of the project using a chain of links ('consequential' linking);
- Analysis files do not contain identifying data (such as name, date of birth and address, or the record number from the original data set); and
- Intermediate data sets and the project specific record identifiers are deleted following completion of the final linked analysis data sets.

The key feature of this protocol is specifying processes that ensure the separation of personal identifying information from general data to be used in analysis, and the absence of any record identifiers in the final linked data which would allow linkage back to the source data (see Figure 1).

¹ NHIMG was the precursor of the Statistical Information Management Committee (SIMC) – a national committee that advises the National Health Information Group (NHIG) on national health statistics.

Figure 1: Schematic example illustrating the data linkage principles of the protocol



Linkage procedures

The above linkage principles are put into effect by following a set of linkage procedures. Guideline linkage procedures are described below; however, it is recognized that the requirements of specific projects may result in some modification to these procedures.² Any differences between these guidelines and the procedures to be used for a particular project are to be described in the submission to the AIHW Ethics Committee.

The linkage procedures are as follows:

- Step 1. Obtain approval for the linkage project from the Institute's Ethics Committee, including signing of the linkage undertaking forms.
- Step 2. Derive the separate data sets to be included in the project, adding a unique project record identifier to each record in each data set. Data contained in these unlinked *source files* include:
 - (A) a project- and data set-specific unique record identifier (or number);
 - (B) data to be used to establish links (i.e. *data linkage items*, such as name, address, sex, date of birth, event date);
 - (C) data items to be used to clarify or check links (i.e. linkage check items, such as country of birth, language spoken, place of relevant event);
and
 - (D) data items to be used in the final analysis (i.e. analysis items, such as care needs, health conditions, service data including services delivered).

(Note: there may be overlap in the data items included in these; for example, country of birth could be included in both C and D).
- Step 3. For each data set separately, derive a *match file*, which contains only A, B and C above:
 - (A) the data-set specific unique project record identifier
 - (B) data linkage items
 - (C) linkage check items.
- Step 4. Undertake linkage of the match files, and produce a *link file* identifying the links using the project record identifiers. For a particular pair of match files, the *link file* contains only three data items:
 - (A) the unique project record identifier from the first match file
 - (A) the unique project record identifier from the second match file
 - (E) a unique *link number* for each record in the link file.

² As noted earlier, these procedures relate to the linkage of internal AIHW databases within the AIHW, and not to linkage between AIHW databases and data sets provided by external researchers undertaken under Ethics Committee clearance (for example, those involving linkage to the National Death Index).

No other data are contained on the link file.

- Step 5. Derive an *analysis file* from each of the source files by retaining only data items to be used in the analysis. Each analysis file will therefore contain:
- (A) the data-set specific unique project record identifier
 - (D) data items (including demographic items) to be used in the final analysis (such as age, sex, country of birth, language spoken, care needs, health conditions, event dates, service data including services delivered).

Identifying data such as name, date of birth and address data are not included on any of the analysis files.

- Step 6. Using the unique project record identifiers, add the link number to the individual analysis files and delete the dataset-specific unique project record identifiers (A) so that the final data sets have no record identifiers that can allow linkage back to the source files. The analysis files will therefore now contain items D and E:
- (E) the link number of each record
 - (D) data items (including demographic items) to be used in the final analysis.

Note that such a data set can still contain identifiable data, because rare characteristics in each data set taken together can make data identifiable. It is for this reason that the linked data set, like other AIHW data sets, has a designated data custodian, and can only be accessed by those who have signed the linkage undertaking for the project.

- Step 7. Following finalisation of the linked analysis data sets, delete the match, link and source files.
- Step 8. Link the analysis files using the *link number*.
- Step 9. Undertake analysis.

Appendix 1: List of other bodies carrying out Institute functions

Australian Centre For Asthma Monitoring (ACAM)
(Woolcock Institute Of Medical Research, Sydney)

Dental Statistics And Research Unit (DSRU)
(University of Adelaide, Adelaide)

Australian General Practice Statistics and Classification Centre (AGPSCC)
(Westmead Hospital, Sydney)

National Injury Surveillance Unit (NISU)
(Flinders University, Adelaide)

National Perinatal Statistics Unit (NPSU)
(University of New South Wales, Sydney)

Public Health Information Development Unit (PHIDU)
(University of Adelaide, Adelaide)

National Centre for Classification in Health (NCCH)
(University of Sydney, Sydney and Queensland University of Technology,
Brisbane)

National Centre for Immunisation Research and Surveillance of Vaccine
Preventable Diseases (NCIRS)
(Westmead Children's Hospital, Sydney)

National Centre in HIV/ Aids Epidemiology and Clinical Research (NCHECR)
(University of New South Wales, Sydney)

Centre for Burden Of Disease and Cost Effectiveness
(University Of Queensland)

Source: AIHW 2005:110-111.

Appendix 2: Australian Institute of Health and Welfare Act 1987, Section 29: Confidentiality

“(1) Subject to this section, a person (in this subsection called the informed person) who has:

- (a) any information concerning another person (which person is in this section called an *information subject*), being information acquired by the informed person because of:
 - (i) holding an office, engagement or appointment, or being employed, under this Act;
 - (ii) performing a duty or function, or exercising a power, under or in connection with this Act; or
 - (iii) doing any act or thing under an agreement or arrangement entered into by the Institute; or
- (b) any document relating to another person (which person is in this section also called an *information subject*), being a document furnished for the purposes of this Act;

shall not, except for the purposes of this Act, either directly or indirectly:

- (c) make a record of any of that information or divulge or communicate any of that information to any person (including an information subject);
- (d) produce that document to any person (including an information subject); or
- (e) be required to divulge or communicate any of that information to a court or to produce that document in a court.

Penalty: \$2,000 or imprisonment for 12 months, or both.

- (2) Subject to subsections (2A) and (2B), nothing in this section prohibits:
- (a) person from divulging or communicating information, or producing a document, to the Minister if it does not identify an information subject;
 - (b) a person from divulging or communicating information, or producing a document, to a person specified in writing by the person (in this subsection called the information provider) who divulged or communicated the information, or produced the document, directly to the Institute;
 - (c) a person from divulging or communicating information, or producing a document, to a person specified in writing by the Ethics Committee if to do so is not contrary to the written terms and conditions (if any) upon

- which the information provider divulged or communicated the information, or produced the document, directly to the Institute; or
- (d) the publication of conclusions based on statistics derived from, or of particulars of procedures used in, the work of the Institute, if:
- (i) to do so is not contrary to the written terms and conditions (if any) upon which an information provider divulged or communicated information relevant to the publication, or produced a document relevant to the publication, directly to the Institute; and
 - (ii) the publication does not identify the information subject.
- (2A) Paragraph (2)(c) applies only to information that is health-related or welfare-related information and statistics.
- (2B) Paragraph (2)(c) applies to a document only to the extent to which the document contains health-related or welfare-related information and statistics.
- (3) A person to whom information is divulged or communicated, or a document is produced, under paragraph (2)(a), (b) or (c), and any person under the control of that person is, in respect of that information or document, subject to subsection (1) as if the person were a person exercising powers, or performing duties or functions, under this Act and had acquired the information or document in the exercise of those powers or the performance of those duties or functions.
- (4) In this section:
- (a) court includes any tribunal, authority or person having power to require the production of documents or the answering of questions;
 - (b) person includes a body or association of persons, whether incorporated or not, and also includes:
 - (i) in the case of an information provider – a body politic; or
 - (ii) in the case of an information subject – a deceased person;
 - (c) produce includes permit access to;
 - (d) publication, in relation to conclusions, statistics or particulars, includes:
 - (i) the divulging or communication to a court of the conclusions, statistics or particulars; and
 - (ii) the production to a court of a document containing the conclusions, statistics or particulars; and
 - (e) a reference to information concerning a person includes:
 - (i) a reference to information as to the whereabouts, existence or non-existence of a document concerning a person; and
 - (ii) a reference to information identifying a person or body providing information concerning a person.”

Source: Parliament of the Commonwealth of Australia 1987a.

Appendix 3: The undertaking required for Ethics Committee approval

Undertaking made in pursuance of Section 29 of the Australian Institute of Health and Welfare Act 1987

WHEREAS:

- (a) Subsection 29 (1) of the *Australian Institute of Health and Welfare Act 1987* ('the Act') provides for the disclosure of information to a person specified in writing by the Ethics Committee;
- (b) The Ethics Committee has agreed to release information to you;

NOW I, _____
Full name and position of Responsible Officer

in the _____
Name of Department or Organisation

HEREBY UNDERTAKE that the above mentioned organisation will use the information in accordance with the following conditions.

1. The unit record file will not be matched, in whole or in part, with any other information for the purposes of attempting to identify individuals, nor will any other attempt to identify an individual be made.
2. The person/organisation will not disclose or release the information to any other person or organisation, except as statistical information that does not identify an individual.
3. Access to the unit record file will be restricted to only those employees of the organisation who are directly responsible to the Principal Investigator. The Principal Investigator will explain to any employees granted access to the information the provisions of the AIHW Act prohibiting release of the information to others.
4. Access will not be granted to any other organisation without specific approval of the AIHW Ethics Committee.
5. The information will be used for statistical purposes in health and/or welfare research.
6. The information will not be used as a basis for any legal, administrative or other actions that could directly affect any particular individuals or organisations as a result of their identification in this project.
7. The identifying information will be used only for the project proposed and described in this application. Use of any of this information in any other project will not be undertaken until a separate application form has been submitted to, and approved by, the Ethics Committee.
8. The recipient will cooperate with any surveillance procedures established by the Institute or its Ethics Committee and advised to the recipient in writing.
9. Results of the project will be made available for consideration by the Ethics Committee, if it so requests prior to any public release.
10. The Institute will be acknowledged in all reports and publications resulting from this project, and will be provided with a copy of all such reports and publications.
11. The recipient will comply in all respects with the requirements of section 29 of the AIHW Act, as attached (and of Part III of *The Privacy Act 1988*).
12. Copyright in all data is vested in the Commonwealth and contributing States and Territories.
13. Any publication which uses the data must identify the AIHW as the source.

In providing this undertaking I understand and accept on behalf of the above mentioned organisation that subsection 29(1) of the *Australian Institute of Health and Welfare Act 1987* provides that a person who receives information or a document relating to another person and makes a record of, or divulges that information to any person, is guilty of an indictable offence punishable on conviction by a fine of \$2,000 or imprisonment for 12 months, or both.

Signature: _____

This application must be signed by a responsible officer with the authority or delegation to commit the above mentioned organisation to the terms and conditions in section 9.

Date: _____

Witness

Name: _____

Position: _____

Organisation/Unit: _____

Signature: _____

Date: _____

All employees of the above organisation who will be granted access to the information must be listed and must agree to comply with the conditions included in the undertaking.

Principal Investigator

Name: _____

Position: _____

Organisation/Unit: _____

Signature: _____

Date: _____

Details of any other person/s who will have access to the data

Name: _____

Position: _____

Organisation/Unit: _____

Signature: _____

Date: _____

Name: _____

Position: _____

Organisation/Unit: _____

Signature: _____

Date: _____

Please attach details of any other person who will have access to the data.

Appendix 4: The Linkage Undertaking

Project name: _____

The undertaking

I, _____

Full name and position of Officer

in the Australian Institute of Health and Welfare

HEREBY UNDERTAKE that the linkage between the following data sets:

Data set: _____
Data set: _____
Data set: _____
Data set: _____

will be done in accordance with the Data Linkage Protocol described in the AIHW policy document *Data Linkage and Protecting Privacy within the Australian Institute of Health and Welfare* (http://www.aihw.gov.au/dataonline/privacy_of_data.cfm). Where there are significant departures from this protocol, they have been drawn to the attention of the Ethics Committee in this application.

Signature and date _____

AIHW data custodian of linked data

I have sighted the Ethics Committee submission for this project involving linkage between data sets, and agree to be the data custodian of the resulting linked data.

Data set _____
Full name of data custodian _____
Position in AIHW _____
Signature and date _____

AIHW data custodian notification

As data custodian, I have sighted the Ethics Committee submission for this project involving linkage between data sets, and agree that the above officer be given access to the specified data set(s) for linking on condition of gaining Ethics Committee approval, and of final approval being granted by data providers (where appropriate).

Data set _____
Full name of data custodian _____
Signature and date _____

Continued

Data set _____
Full name of data custodian _____
Signature and date _____

Data set _____
Full name of data custodian _____
Signature and date _____

Data set _____
Full name of data custodian _____
Signature and date _____

Appendix 5: Analysis of previously linked data

Project name: _____

Description of linked data: _____

Requesting officer

Full name _____
Position in AIHW _____
Signature and date _____

AIHW data custodian of linked data

I have sighted the Ethics Committee submission for this project involving analysis of previously linked data, and agree that access to the specified data set(s) be given to the project officers on condition of gaining Ethics Committee approval, and of final approval being granted by data providers (where appropriate).

Data set _____
Full name of data custodian _____
Position in AIHW _____
Signature and date _____

AIHW data custodian notification

As data custodian, I have sighted the Ethics Committee submission for this project involving previously linked data, and agree that access to the specified data set(s) be given to the project officers on condition of gaining Ethics Committee approval.

Data set _____
Full name of data custodian _____
Signature and date _____

Data set _____
Full name of data custodian _____
Signature and date _____

Data set _____
Full name of data custodian _____
Signature and date _____

Data set _____
Full name of data custodian _____
Signature and date _____

References

AIHW 2001. AIHW information security and privacy: Policy and procedures. Canberra: AIHW.

AIHW 2003. Guidelines for custody of AIHW data (April 2003). AIHW, Canberra. Viewed 14 September 2004 < AIHW intranet <<http://intranet/index.cfm/15,496,99,183,doc/496.doc>>>.

AIHW 2005. AIHW Annual report 2004–05. Cat. no. AUS 66. Canberra: AIHW.

NCSIMG 2004. Statistical data linkage in community services data collections: a report prepared by the Statistical Linkage Key Working Group. Canberra: AIHW.

NHIMG 2002. Issues for the use of unique patient identifiers in statistical collections. Canberra: AIHW.

Parliament of the Commonwealth of Australia 1987a. Australian Institute of Health and Welfare Act 1987.

Parliament of the Commonwealth of Australia (House of Representatives) 1987b. Australian Institute of Health Bill 1987, Explanatory Memorandum. pp. 11264/87.

Revisions

Date	Authority	Comments
19 June 2006	Executive Committee	Endorsed
22 June 2006	Board	Approved