



**Guidelines  
for the  
Disclosure  
of  
Secondary Use  
Health Information  
for  
Statistical Reporting, Research  
and Analysis**

Endorsed 15 June 2017

## Introduction

### Development

These guidelines are an update of the version developed by the National Health Information Standards and Statistics Committee (NHISSC) in 2015. They set out principles and techniques regarding the disclosure of *secondary use health information* for statistical reporting, research and analysis. Terms that are defined in Appendix A: Glossary are *italicised* and *bolded* the first time they appear in the body of the Guidelines

### Scope

These guidelines are appropriate for health information held by National Health Information Agreement (NHIA) signatories such as National Minimum Datasets (NMDs) and Dataset Specifications (DSSs) disclosed for purposes of statistical reporting, research and analysis. It applies to *unit record* and *aggregate data* and to individuals and health services and includes the output of data linkage.

Out of scope for these guidelines are data sharing arrangements between governments including the specification and provision of NMDs & DSSs; specified agreements or data outside the NHIA scope; data security and data reliability. Also excluded are data linkage protocols and data for payment or linkage purposes, which by necessity contain *re-identifiable* or *identifiable* information.

### Context

This document provides general guidance to assist in the management of risks regarding the identification of individual patients/clients and health service providers, where legislative provisions do not provide sufficient detail about the release of data (Although the case studies included in this document use hospital admitted patient data to illustrate the principles and techniques, those principles and techniques are intended to be broadly applicable to other health information).

These guidelines are intended to be used in conjunction with other more specific agreements or arrangements, including both existing agreements between parties to the NHIA and agreements with regard to the subsequent release to third parties of data owned by jurisdictions which are parties to the NHIA. For example, the National Health and Medical Research Council's (NHMRC) National Statement on Ethical Conduct in Research Involving Humans (2007, updated May 2015, s5.1.22) states that institutions may choose to exempt from ethical review research that (a) is negligible risk research; and (b) involves the use of existing collections of data or records that contain only *non-identifiable data* about human beings.

Although health privacy legislation and policies vary between Australian jurisdictions, their common purpose is to govern the collection, use and disclosure of

*personal information* about the health of, or health services provided to, individuals whose identity is apparent or can be reasonably ascertained.

This means that, before providing a health dataset (in either unit record or aggregate form) to other agencies or individuals, the providing agency must satisfy itself that:

- if the dataset is identifiable or re-identifiable data, it will only be disclosed where there is patient consent or for purposes for which the use or disclosure of personal information is permitted by its policies and legislation, or
- the dataset is non-identifiable data, in the sense that the identity of individual patients is not apparent, and cannot reasonably be ascertained from the dataset either on its own or in combination with any other information to which the user may reasonably be considered to have access.

The focus of these guidelines is on the second requirement. The Australian Institute of Health and Welfare's (AIHW) *Policy on reporting to manage confidentiality and reliability*, Nov. 2013 contends that a cell in a table is identifiable if, as well as being able to identify the entity, other details are revealed (i.e. 'attribute disclosure'). Similarly, NHIA Clause 26f states that aggregate data suppression rules should be case specific and only enacted where there is a risk that an individual could have information disclosed that was previously unknown to the recipients of the data. NHISSC supports the AIHW and NHIA approaches but also recognises that minimising the risk of identification is the safest way to prevent attribute disclosure.

The Health Statistics NSW paper: *Privacy Issues and the Reporting of Small Numbers* also refers to community disclosure where a data release has the potential to disclose information about small communities. It is NHISSC's view that managing the risk of identification and attribute disclosure for individuals will generally also ensure that small communities are similarly protected. However additional action may be necessary, e.g. the Northern Territory department does not always release information for certain Statistical Area level 2s (SA2s) because those SA2s, in combination with Indigenous Status, would immediately identify specific Aboriginal communities.

## **Purpose**

The aim of these guidelines is to assist data custodians to manage the risk of identification of individual patients/clients (the simple fact that a person attended a specific health service may raise privacy issues for the person concerned) and the disclosure of previously unknown information about that patient (e.g. their diagnosis or procedures) by defining a number of underpinning principles and recommending some of the available techniques.

## Principles

There are a number of principles that apply to disclosure of health information for statistical reporting, research and analysis:

1. Information is a strategic national asset and agencies should provide as much public access as possible whilst minimising the risk of personal information disclosure.
2. Data disclosure must comply with legislation and interagency agreements, including but not limited to information privacy principles, secrecy, consent, commercial-in-confidence, freedom of information and Commonwealth data integration principles.
3. All *data custodians* should have, adopt, or develop their own guidelines which conform to legislation specifically applicable to them and utilise the principles and techniques outlined in these guidelines.
4. Original data custodians retain primary responsibility for their data at all times. For example, where states and territories provide national minimum datasets to the AIHW, Australian Department of Health, Independent Hospital Pricing Authority, etc. the state and territory is the original data custodian and all other parties provided with that dataset are also custodians of the data.
5. Where other data custodians seek to publish or disseminate health information for research or analysis purposes, they need to reach agreement with all of the original data custodians. The AIHW's policy guideline 6 says that where data suppliers require the application of additional suppression rules to an AIHW release in order to manage confidentiality – the supplier's proposed suppression rules should be applied.
6. In regard to unit record data for research and analysis purposes, custodians should aim to provide data that satisfies the purpose of the request whilst seeking to prevent individuals or an organisation's commercial interests being identified or having information disclosed about them that was previously unknown to the recipients. Further, only those data items essential to the user's purpose should be released. It is not good practice to provide more information than is needed for a specific project (data minimisation principle).
7. Aggregate data suppression rules should be case specific and only enacted where: (i) there is a risk that an individual could be identified possibly leading to information about them being disclosed that was previously unknown to any possible recipient of the data; or (ii) there is a risk of exposing an organisation's commercial operations. Data custodians must protect the 'commercial in confidence' nature of private hospitals in both aggregate and unit record data releases, unless they have received explicit approval from all relevant private organisations. The provision of aggregate information about the performance of individual public hospitals is the prerogative of State/Territory health authorities.

## Techniques

The Australian Bureau of Statistics (ABS) provides a comprehensive range of techniques, including alternative approaches for managing the risk of personal identification that could be applied to health information by data custodians. The ABS' Confidentiality Series can be accessed at:

<http://www.nss.gov.au/nss/home.NSF/pages/Confidentiality+Information+Sheets>

NHISSC has endorsed some specific techniques. In all cases professional judgement is required to assess the privacy implications of the request, to utilise one or more of the following techniques, and to assess whether an individual or an organisation's commercial operations can be identified and have previously unknown information about them disclosed:

1. In seeking to minimise the risk of identification and attribute disclosure in unit record data releases, custodians should anonymise data by:
  - removing and/or modifying personal identifiers such as a person's name, address, date of birth and unit record number. For example, if it is an essential requirement of the data request to know that multiple episodes relate to the same person in the same hospital, then the unit record numbers provided should be encrypted (see Case Study 1);
  - not providing other specific dates unless absolutely necessary (see Case Study 2). In most cases the request can be satisfied by using one or a combination of the following:
    - (i) provision of month and year of admission/separation, etc;
    - (ii) provision of day of the week and time of day (generally for emergency department data);
    - (iii) provision of dates encrypted as the number of days from the date of first event (day zero) or other selected starting date which is not known to the user (see Case Study 1). This is in order to enable the user to identify episode chronology and calculate intervals between events;
    - (iv) include sequence number and days between the sequential episodes.

NB Where an organisation is providing one of the two datasets for a linkage project, and where both datasets contain temporal information, the data custodian will need to provide specific dates/times to the linkage agency. Once linkage is performed, the linkage agency then calculates and provides to the user only the temporal information from the two datasets as per (i) to (iv) above. If the linkage agency is also the organisation of the end user, date and times should be supplied in a separate file to the rest of the content data with a common project specific record ID. The linkage agency is able to link the two datasets at their end and

calculate duration variables and then must destroy the date/time dataset. The data custodian should maintain a copy of the date/time dataset for re-supply in the event that the investigator identifies any data quality concerns.

- aggregating variables wherever possible: e.g. provide 5 year age groups rather than date of birth; a metropolitan/rural indicator or SA2 rather than postcode and locality of residence, diagnosis related group instead of individual diagnosis and procedure codes, etc.

This technique, which can also be applied to aggregate data, is based on the data minimisation principle and addresses the concern that if the 'denominator' population (i.e. the population in the community) is too small it can provide a risk of individual identification and information disclosure.

Custodians should ensure that the pool of people who could potentially have contributed to unit record data or to a cell in aggregate data is as large as possible while still enabling the user to do their job. This approach could be assisted by a numerical test, i.e. unit record data would not be provided for sub-groups where their estimated population is less than a value set by the custodian. (See Case Study 3).

2. Custodians should require external users of unit record data to sign 'conditions of release' covering specific confidentiality requirements such as the purpose for which data may be used, requirements for the secure storage and retention of data, restrictions on the publication of data, the provision of data to a third party and any attempt to re-identify individuals. The Conditions should indicate the applicable laws covering release and the penalties that apply for a breach of the conditions.
3. To maintain the anonymity of individual private hospitals or private hospital owners in aggregate tables derived from hospital morbidity data, cells should be suppressed if:
  - there are fewer than three (3) separately owned private hospitals; or
  - there are three or more separately owned private hospitals and one private hospital owner contributed more than 85% of the total separations; or
  - there are three or more separately owned private hospitals and two private hospital owners contributed more than 90% of the total separations. (see Case Study 4)

4. Small cell suppression in aggregate data is commonly used in national statistical reporting and by many of the individual states/territories. It is an easy test to apply and detects cells with potential identification problems possibly leading to the release of previously unknown information, e.g. age, diagnosis or procedure (see Case Study 5).

Small cells (e.g. containing values between 1 and 4) may be avoided by aggregating variables, e.g. age group ranges 65-74, 75-84, 85+ are replaced with 65+, data from small areas or communities are aggregated over a number of years, etc. If this is not possible, then the small cells may be suppressed.

5. Cells in aggregate data where the value of the cell is the same as a row/column total should be suppressed if it is considered that it could lead to disclosure of an additional attribute.
6. The application of the private hospital, small cell, and cell = row/column total techniques may require the suppression or amalgamation of several cells in a table, possibly including some with values of zero or greater than 4, in order that a cell not be derivable by subtraction. In these circumstances, it is advisable that the compiler of the table choose a method of confidentialisation that maintains the column and row totals and results in the loss of the least amount of useful information (see Case Study 6).
7. As the original data custodians, if individual states/territories want to apply particular rules for a national statistical publication or other specific aggregate data releases involving their data (e.g. some requiring no restrictions, some requiring small cell suppressions with different threshold values e.g. <3, <5, <10, etc.) the most conservative option will apply in line with the principles espoused in this document.

## **Case Studies**

These Case Studies are provided to illustrate the techniques

### **Unit Record Data**

**Case Study 1: A contractor working for a Government Department on a service planning project requests unit record data, including amongst other data items:**

- Campus
- Patient ID
- Admission date
- Separation date
- Date of Birth
- All diagnoses codes
- All procedure codes
- Postcode
- Locality

After discussion with the contractor, who confirms that: (i) identifying the campus; (ii) knowing that multiple episodes relate to the same patient in that campus and their chronology; and (iii) calculating the interval between episodes; are essential to the work, a dataset is negotiated that seeks to prevent individual identification, but still enables the contractor to complete the required job. Any data items that are not essential to the user's purpose are also not provided. The dataset released after conditions of release were signed included:

- Campus
- Encrypted Patient ID
- Admission and Separation Month and Year plus encrypted Admission and Separation dates (number of days since a specified date unknown to the user)
- Length of Stay
- Age in 5 year groups
- Diagnosis Related Group
- Statistical Local Area of Usual Residence



**Case Study 2: A University researcher investigating the link between air pollution and admission for respiratory conditions requests unit record data, including amongst other data items:**

- Campus
- Patient ID
- Admission date
- Separation date
- Date of Birth
- All diagnoses codes
- Postcode
- Locality

After discussion with the researcher, who confirms that knowing the admission date and a more detailed usual residence than SLA is essential to the work, a dataset is negotiated that prevents individual and health service identification, but still enables them to complete the required job. Any data items that are not essential to the user's purpose are also not provided. The dataset released after conditions of release were signed included:

- Encrypted Campus code
- Encrypted Patient ID
- Admission Date
- Length of Stay
- Age in 5 year groups
- Diagnosis codes (Only those episodes with respiratory condition as principal or additional diagnosis)
- Locality of Usual Residence

**Case Study 3: A consultant working for a community action group and looking at self-sufficiency issues for a specific hospital and its catchment (as defined by a list of 8 postcodes), requests unit record data, including amongst other data items:**

- Campus
- Admission date
- Separation date
- Age
- Sex
- Diagnosis Related Group
- Postcode
- Discharge Status
- Insurance Status
- Weighted Separations

Examination of underlying population data grouped by five year age groups, sex and postcode reveals that a significant number of cells are less than 1000 ( a pre-set minimum value chosen by this particular custodian). Indeed for some of the postcodes within the prescribed catchment, the total for all ages for males and females is under 1000.

In discussion with the consultant the need for an identified campus was established and agreed and it was confirmed that a patient identifier of any kind was not required. Also agreed was separation month and year as opposed to admission and separation dates. After discussion about compacting either the geographic area, the age dimensions or both, it was agreed that the eight postcodes would be grouped together as the "catchment" and that the area outside the catchment would be Statistical Area Level 2. Additionally the oldest age category would be "70+".

The dataset released after conditions of release were signed included:

- Campus
- Separation month and Year
- 5 Year Age Groups ending with 70+
- Sex
- Diagnosis Related Group
- Catchment as a whole/Non-catchment by SA2
- Discharge Status
- Insurance Status
- Weighted Separations

## Aggregate Data

### Case Study 4: Private Hospital data

An external requestor asks for the following aggregate table where Parameter X could be a sub-group of hospitals, defined age group, geographical location, etc. or a combination thereof.

For the purposes of illustration only a subset of diagnosis are shown.

<b>Separations by Parameter X by Principal Diagnoses, Private Hospitals</b>	
<b>Principal Diagnosis</b>	<b>Seps</b>
G47 Sleep Disorders	491
I20 Angina Pectoris	159
J35 Chronic Diseases of Tonsils and Adenoids	188
K40 Inguinal Hernia	196
K80 Cholelithiasis	175
M17 Gonarthrosis (Arthrosis of knee)	253
M75 Shoulder Lesions	234
O80 Single Spontaneous Delivery	179
O82 Single Delivery by Caesarian Section	185
R07 Pain in Throat and Chest	148
Other diagnoses with less than 10 separations	1124
<b>Total</b>	<b>3332</b>

Step 1 - Suppress numbers where there are less than 3 separately owned private hospitals or where 1 private hospital owner has more than 85% of separations or two have more than 90%. In this example, there were only two private hospitals providing maternity services.

<b>Separations by Parameter X by Principal Diagnoses, Private Hospitals</b>	
<b>Principal Diagnosis</b>	<b>Seps</b>
G47 Sleep Disorders	491
I20 Angina Pectoris	159
J35 Chronic Diseases of Tonsils and Adenoids	188
K40 Inguinal Hernia	196
K80 Cholelithiasis	175
M17 Gonarthrosis (Arthrosis of knee)	253
M75 Shoulder Lesions	234
O80 Single Spontaneous Delivery	np
O82 Single Delivery by Caesarian Section	np
R07 Pain in Throat and Chest	148
Other diagnoses with less than 10 separations	1124
<b>Total</b>	<b>3332</b>

np Commercial-in-confidence

Step 2 - Further action to prevent deduction. From the table above you can deduce that the "np"s total 364. Whilst it is not possible to accurately split between O80 and O82, this number provides a good indication of the level of obstetric activity. By including the np numbers in with "Other" this information is protected as follows:

<b>Separations by Parameter X by Principal Diagnoses, Private Hospitals</b>	
<b>Principal Diagnosis</b>	<b>Seps</b>
G47 Sleep Disorders	491
I20 Angina Pectoris	159
J35 Chronic Diseases of Tonsils and Adenoids	188
K40 Inguinal Hernia	196
K80 Cholelithiasis	175
M17 Gonarthrosis (Arthrosis of knee)	253
M75 Shoulder Lesions	234
O80 Single Spontaneous Delivery	np
O82 Single Delivery by Caesarian Section	np
R07 Pain in Throat and Chest	148
Other - commercial-in-confidence data and diagnoses w ith less than 10 separations	1488
<b>Total</b>	<b>3332</b>

np Commercial-in-confidence

**Case Study 5: Attribute Disclosure**

An external requestor specifies the following table which contains a small cell (a value between 1 and 4). In order to identify a person in that or any other cell, all of the following detail would need to be already known: that they were admitted to hospital in that time period, their indigenous status and their specific principal diagnosis. No additional information is provided to that already known; i.e. there is possible identification but no attribute disclosure, so the custodian may decide to release the table without alteration.

<b>Hospital Separations for Cardio-Vascular Disease by sub-type and indigenous status</b>		
<b>Principal Diagnosis</b>	<b>Indigenous status</b>	
	<b>Non-indigenous</b>	<b>Indigenous</b>
Acute rheumatic fever I00-I02	22	1
Chronic rheumatic heart diseases I05-I09	474	13
Hypertensive diseases I10-I15	1,696	5
Ischaemic heart diseases I20-I25	40,204	212
Pulmonary heart disease and diseases of pulmonary circulation I26-I28	2,548	18
Other forms of heart disease I30-I52	35,337	73
Cerebrovascular diseases I60-I69	11,250	42
Diseases of arteries, arterioles and capillaries I70-I79	7,350	22
Diseases of veins, lymphatic vessels and lymph nodes, not elsewhere classified I80-I89	22,021	49
Other and unspecified disorders of the circulatory system I95-I99	2,491	12
<b>Total</b>	<b>123,393</b>	<b>447</b>

A second table specified contains small cells. In order to identify a person in those or any other cell, all of the following would need to be already known: that they were admitted to hospital in that financial year, their indigenous status and their specific principal diagnosis. It is possible that this would enable the person’s age group to be determined; i.e. there is possible identification and attribute disclosure. Techniques outlined in this document and in Case Study 6 would need to be applied to prevent attribute disclosure.

<b>Hospital Separations for Hypertensive diseases (I10–I15)</b>			
<b>by age group and indigenous status</b>			
<b>Age Group</b>	<b>Indigenous status</b>		
	<b>Non-indigenous</b>	<b>Indigenous</b>	<b>Total</b>
00-24	58		58
25-29	16		16
30-34	38	1	39
35-39	49		49
40-44	59	1	60
45-49	82		82
50-54	123	1	124
55-59	124	1	125
60-64	128		128
65-69	160		160
70-74	171	1	172
75-79	200		200
80-84	243		243
85+	245		245
<b>Total</b>	<b>1,696</b>	<b>5</b>	<b>1,701</b>

## Case Study 6: Small Cell Suppression

An external requestor asks for the following aggregate table where Parameter X could be a group of hospitals, defined age group, geographical location, etc or a combination thereof.

Separations for Parameter X by principal diagnosis of mental health related conditions and sex			
Principal diagnosis	Males	Females	Total
Mental & behavioural disorders due to psychoactive substance use (F10–F19)	183	130	313
Schizophrenia, schizotypal and delusional disorders (F20–F29)	169	204	373
Mood disorders (F30–F39)	136	146	282
Neurotic, stress-related disorders (F40–F49)	59	114	173
Disorders of adult personality and behaviour (F60–F69)	14	34	48
Behavioural and emotional disorders (F90–F98)	6	9	15
Organic, including symptomatic, mental disorders (F00–F09)	6	7	13
Behavioural syndromes assoc. w ith physiological disturbances (F50–F59)	1	16	17
Unspecified mental disorder (F99)	1	7	8
Mental retardation (F70–F79)	1	2	3
Disorders of psych. Development (F80–F89)	1	4	5
Other <sup>(a)</sup>	37	39	76
<b>Total</b>	<b>614</b>	<b>712</b>	<b>1326</b>

(a) list of codes included in other

It is determined that action is required because the small cells may lead to identification and attribute disclosure

Step 1 - Suppress all numbers between 1 and 4

Principal diagnosis	Males	Females	Total
Mental & behavioural disorders due to psychoactive substance use (F10–F19)	183	130	313
Schizophrenia, schizotypal and delusional disorders (F20–F29)	169	204	373
Mood disorders (F30–F39)	136	146	282
Neurotic, stress-related disorders (F40–F49)	59	114	173
Disorders of adult personality and behaviour (F60–F69)	14	34	48
Behavioural and emotional disorders (F90–F98)	6	9	15
Organic, including symptomatic, mental disorders (F00–F09)	6	7	13
Behavioural syndromes assoc. w ith physiological disturbances (F50–F59)	np	16	17
Unspecified mental disorder (F99)	np	7	8
Mental retardation (F70–F79)	np	np	np
Disorders of psych. Development (F80–F89)	np	np	5
Other <sup>(a)</sup>	37	39	76
<b>Total</b>	<b>614</b>	<b>712</b>	<b>1326</b>

(a) list of codes included in other

np Numbers between 1 and 4 are not published.

Step 2 - Further suppression to prevent deduction

- It is very easy to deduce the value of each of "np"
- The male column adds to 610 therefore each male "np" must be 1, and the female numbers can be determined by deduction.
- The total column adds to 1323 without the "np", therefore the "np" must be 3
- To prevent identification by deduction you would also suppress the female numbers for F50-F59 and F99; plus the male and female numbers for F00-F09\*; and the Total for F80-F89. \*F00-F09 was chosen as it had the lowest numbers
- Note how Column totals and all Row totals bar two are preserved.

Principal diagnosis	Males	Females	Total
Mental & behavioural disorders due to psychoactive substance use (F10–F19)	183	130	313
Schizophrenia, schizotypal and delusional disorders (F20–F29)	169	204	373
Mood disorders (F30–F39)	136	146	282
Neurotic, stress-related disorders (F40–F49)	59	114	173
Disorders of adult personality and behaviour (F60–F69)	14	34	48
Behavioural and emotional disorders (F90–F98)	6	9	15
Organic, including symptomatic, mental disorders (F00–F09)	np	np	13
Behavioural syndromes assoc. with physiological disturbances (F50–F59)	np	np	17
Unspecified mental disorder (F99)	np	np	8
Mental retardation (F70–F79)	np	np	np
Disorders of psych. Development (F80–F89)	np	np	np
Other <sup>(a)</sup>	37	39	76
<b>Total</b>	<b>614</b>	<b>712</b>	<b>1326</b>

(a) list of codes included in other

np Not published to prevent disclosure of numbers between 1 and 4.

## Step 2 Alternative - Expanding "Other"

You could choose to expand "other" to include suppressed male and female numbers.

Principal diagnosis	Males	Females	Total
Mental & behavioural disorders due to psychoactive substance use (F10–F19)	183	130	313
Schizophrenia, schizotypal and delusional disorders (F20–F29)	169	204	373
Mood disorders (F30–F39)	136	146	282
Neurotic, stress-related disorders (F40–F49)	59	114	173
Disorders of adult personality and behaviour (F60–F69)	14	34	48
Behavioural and emotional disorders (F90–F98)	6	9	15
Other <sup>(a)</sup>	47	75	122
<b>Total</b>	<b>614</b>	<b>712</b>	<b>1326</b>

(a) original list of codes plus F00-F09, F50-F59, F70-F89 and F99



## **Appendix A: Glossary**

### **Aggregate data**

Aggregate data are produced by grouping information into categories and aggregating values within these categories. For example, a count of the number of people of a particular age (obtained from the question 'In what year were you born?'). Aggregate data is typically presented in tables. Aggregate data is also referred to as tabular data or macrodata. (ABS)

### **Anonymisation**

The process of removing identifiers and/or other data items from a dataset with the intention that the dataset content changes from identifiable to re-identifiable or non-identifiable data or from re-identifiable data to non-identifiable data.

### **Data Custodian**

The organisation or agency which is responsible for the collection, use and disclosure of information in a dataset. Data custodians have an obligation to keep the confidential information they are entrusted with secret. (ABS)

### **Health Information**

Includes any data required to inform health research/health status. This includes health data sets as well as data sets linked to data where the primary issue is health (e.g. socio economic status, education, occupational health and safety)

### **Individually Identifiable Data**

Data from which the identity of a specific individual can reasonably be ascertained (NHMRC's National Statement on Ethical Conduct in Human Research (2007, updated May 2015))

### **Non-identifiable data**

Data that have never been labelled with individual identifiers or from which identifiers have been permanently removed, and by means of which no specific individual can be identified. A subset of non-identifiable data are those that can be linked with other data so it can be known they are about the same data subject, although the person's identity remains unknown. (NHMRC's National Statement on Ethical Conduct in Human Research (2007, updated May 2015))

## **Personal information**

Information or an opinion (including information or an opinion forming part of a database), whether true or not, and whether recorded in a material form or not, about an individual whose identity is apparent, or can be reasonably ascertained, from the information or opinion (whether directly from the information or from the information when read in combination with other information held by or available to the organisation).

## **Re-identifiable Data**

Data from which identifiers have been removed and replaced by a code, but it remains possible to re-identify a specific individual by, for example, using the code or linking different data sets.

(NHMRC's National Statement on Ethical Conduct in Human Research (2007, updated May 2015))

It is also possible to re-identify a specific individual via a combination of variables.

## **Secondary Use**

The use of data for any authorised purpose other than the use for which the data was originally collected (primary purpose)

## **Unit Record Data** (Also known as Patient level data or Microdata)

Each record represents observations for an individual or organisation. Unit record data may contain individual responses to questions on a survey questionnaire or administrative form. For example, answers given to the question 'In what year were you born?'. (Microdata - ABS)