Glossary

Aggregated data   means 'information about the population as a whole', such as the Census.

Big data   means 'high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight, decision making, and process optimization'[1] An example would be an analysis of the 153 million records from six databases required to understand the unplanned hospital stays of Western Australian seniors.

Confidentialised data   See **de-identified data.**

Data linking   means 'the bringing together of two or more data sets to create a new, richer data set.'[2] By bringing together sets of data that were previously isolated, researchers, clinicians and governments can deepen their understandings of the ways people actually use the health care system. This has the potential to inform government policy making and decisions about improving service delivery.[3]

Data custodian   means 'agencies responsible for managing the use, disclosure and protection of source data used in a statistical data integration project. Data custodians collect and hold information on behalf of a data provider (defined as an individual, household, business or other organisation which supplies data either for statistical or administrative purposes). The role of data custodians may also extend to

---

1   Department of Finance and Deregulation, Australian Government Information Management Office, *The Australia Public Service Big Data Strategy*, August 2013, p. 8, www.finance.gov.au/sites/default/files/Big-Data-Strategy_0.pdf (accessed 8 January 2016).

2   The National Statistics Service is a network of Australian Government and State and Territory entities led by the Australian Bureau of Statistics that works together to improve Australia's statistics system. National Statistics Service, *Data Linking: What is data linking?*, Information Sheet 1, p. 1, www.nss.gov.au/nss/home.nsf/533222ebfd5ac03aca25711000044c9e/91242a5a14b12e26ca257 ba8007b0819/$FILE/data%20linking%20w.pdf (accessed 9 December 2015).

3   National Statistics Service, *Data Linking: What is data linking?*, Information Sheet 1, p. 1.

	producing source data, in addition to their role as a holder of datasets.'[4]
Data linkage key	means a data linkage key is a code that is constructed to replace identifying information, such as name, date of birth and address on a linked record in order to protect the privacy of the subjects of the study. By using a linkage key, researchers can link records that belong to the same person from multiple datasets without needing to know who the person is.[5]
De-identified data	means 'a process by which a collection of data or information (for example, a dataset) is altered to remove or obscure personal identifiers and personal information (that is, information that would allow the identification of individuals who are the source or subject of the data or information).[6]
Enduring linkage	means that once links between datasets are created, the links are maintained and data, as it is progressively added to the dataset, is checked against the existing data to create an enduring research resource.[7]
Ethics approval	means 'review of research by the Human Research Ethics Committee of the National Health and Medical Research Council or another body'.[8]
Integrating Authority	means 'an authority responsible for the ongoing management of integrated data, ensuring it is kept secure, confidential and fit for the purposes of the approval process'.[9] The Australian Institute of Health and Welfare are an integrating authority.

---

4   National Statistics Service, 'Rights and responsibilities of data custodians', www.nss.gov.au/nss/home.NSF/533222ebfd5ac03aca25711000044c9e/59fd060543b4e9e0ca25 7a4e001eacfe/$FILE/Rights,%20responsibilities%20and%20roles%20of%20data%20custodian s_Dec2013.pdf (accessed 22 January 2015).

5   National Statistics Service, *Data Linking: Deterministic linking and linkage keys*, Information Sheet 3, p. 2.

6   Department of Finance, *Australian Public Service Better Practice Guide for Big Data*, January 2015, p. 30.

7   Dr Merran Smith, Chief Executive, PHRN, *Committee Hansard*, 11 December 2015, p. 23.

8   National Health and Medical Research Council, *National Statement on Ethical Conduct in Human Research (2007)*, May 2015, p. 88.

9   Cross Portfolio Statistical Integration Committee, *Data Integration Involving Commonwealth Data for Statistical and Research Purposes: Governance and Institutional Arrangements*, 6 October 2010, p. 5.

Metadata	is 'data about data'.[10] It can be of three types: definitional (provides definitions about aspect of the data), procedural operational (drives the collation of the data) or conceptual (dealing with concept sources and methods).[11]
Perturbed data	See **confidentialised data**
Public interest certificate	means an application made under the *Privacy Act 1988* for a certificate to engage in an act or practice that breaches, or may breach, an Australian Privacy Principle.[12]
Unique identifier	means 'a number or code that uniquely identifies a person, business or organisation, such as a passport number or Australian Business Number'.[13]

---

10 Prof Louisa Jorm, Director, Centre for Big Data Research in Health, *Committee Hansard*, 11 December 2015, p. 13.

11 Mr Warren Richter, Head, Chief Information Officer Group, AIHW, *Committee Hansard*, 11 December 2015, p. 73.

12 See *Privacy Act 1988*, s. 72.

13 National Statistics Service, *Data Linking: What is data linking?*, Information Sheet 1, p. 1.