Australian Government
Office of the Australian Information Commissioner

Privacy business resource 4

De-identification of data and information

April 2014

De-identification of personal information can enable information to be shared or published without jeopardising personal privacy. This enables organisations to maximise the utility and value of information assets while safeguarding privacy and confidentiality.

Many organisations collect and retain data and information (known as information assets) that include personal information.[1] In doing so they must comply with the Australian Privacy Principles (APPs) in the *Privacy Act 1988*. The APPs regulate how organisations collect, use, disclose and store personal information. Importantly, APP 6 limits the disclosure of personal information.

This resource provides general advice about de-identification, to assist businesses and other organisations to protect privacy when using or sharing information assets containing personal information. Guidance is provided on when de-identification may be appropriate, how to choose appropriate de-identification techniques and how to assess the risk of re-identification. This resource is not legally binding, and should be considered in the context of other requirements of the Privacy Act and other legislation.

General obligations under the Privacy Act

The APPs support de-identification of personal information in specified circumstances. For example, if an entity to which the Privacy Act applies no longer needs personal information for any purpose for which it was collected or may be used, the entity must take reasonable steps to destroy or de-identify the information (APP 11.2).[2]

The obligation to destroy or de-identify does not apply to information that an organisation is legally obliged to retain. Nor does it apply to personal information that is part of a Commonwealth record — which may be relevant if an organisation that has entered into a contract with an Australian Government agency.[3]

Organisations should also be aware of the obligation imposed by the *Tax File Number Guidelines* issued under s 17 of the Privacy Act. The Guidelines require an organisation to take reasonable steps to securely destroy or permanently de-identify tax file number information that is no longer required by law to be retained, or is no longer necessary for a purpose under taxation law, personal assistance law or superannuation law.[4]

---

[1] 'Personal information' is defined in s 6(1) of the Privacy Act, as information or an opinion about an identified individual or an individual who is reasonably identifiable. For more information, see www.oaic.gov.au/privacy/what-is-covered-by-privacy.

[2] Other APPs that refer to de-identification are APP 4.3 (unsolicited personal information) and APP 6.4 (disclosure of personal information). See also APP 13.1 on correction of personal information. Further information about the APPs is available in the OAIC's *Australian Privacy Principle Guidelines*, available at www.oaic.gov.au.

[3] 'Commonwealth record' is defined in s 6(1) of the Privacy Act as having the same meaning as in the *Archives Act 1983*. For information about the Archives Act, see www.naa.gov.au.

[4] Office of the Australian Information Commissioner, *Tax File Number Guidelines 2011*, published December 2011, Comlaw website, www.comlaw.gov.au/Details/F2011L02748; and *Privacy Fact Sheet 6 — The binding Tax File Number Guidelines 2011 and the protection of tax file number information*, last modified March 2012, Office of the Australian Information Commissioner website, www.oaic.gov.au/publications/privacy_fact_sheets/privacy_fact_sheet6_TFN_guide_2011.html.

De-identification obligations for credit reporting bodies and credit providers also apply in relation to credit-related personal information.[5] Credit reporting bodies must also comply with s 20M of the Privacy Act, which prevents the use and disclosure of de-identified credit reporting information except when that use and disclosure is for the purpose of conducting research in accordance with the Privacy Commissioner's *Privacy (Credit Related Research) Rule 2014*.

Organisations cannot collect health information about individuals for one of the research or public health or safety purposes permitted under s 16B(2) of the Privacy Act if de-identified information would serve the same purpose (s 16B(2)(b)). If de-identified information would not serve the same purpose (and if other conditions imposed in s 16B(2) have been met) the organisation can only collect the information in accordance with guidelines approved by the Information Commissioner or the Privacy Commissioner under s 95A about use of health information for research or public health or safety purposes.

More information about an organisation's obligations under the Privacy Act is available at www.oaic.gov.au.

What is de-identification?

Personal information is 'de-identified' if the information is no longer about an identifiable individual or an individual who is reasonably identifiable.[6]

De-identification involves removing or altering information that identifies an individual or is reasonably likely to do so. Generally, de-identification includes two steps:

1. removing personal identifiers, such as an individual's name, address, date of birth or other identifying information, and
2. removing or altering other information that may allow an individual to be identified, for example, because of a rare characteristic of the individual, or a combination of unique or remarkable characteristics that enable identification.

De-identification can be effective in preventing re-identification of an individual, but may not remove that risk altogether. There may, for example, be a possibility that another dataset or other information could be matched with the de-identified information.[7] The risk of re-identification must be actively assessed and managed to mitigate this risk. This should occur both before an information asset is de-identified and after disclosure of a de-identified asset (see 'Assessing the risks of re-identification' below).

Confidentialisation

The National Statistical Services' (NSS's) *Confidentiality Information Series* provides guidance about a process called 'confidentialisation'. Confidentialisation involves both de-identifying data and then taking the additional step of assessing and managing the risk of indirect identification occurring in the de-identified dataset.[8]

---

[5] See the OAIC's *Privacy Business Resource 3: Credit reporting — what has changed* (www.oaic.gov.au/privacy/privacy-resources/privacy-business-resources/privacy-business-resource-3-credit-reporting-what-has-changed) for an overview of Part IIIA of the Privacy Act, which regulates credit reporting in Australia.

[6] This definition of 'de-identified' is given in s 6(1) of Privacy Act.

[7] Eg, Anna Cavoukian and Khaled El Emam, *Dispelling the myths surrounding de-identification: anonymisation remains a strong tool for protecting privacy*, (June 2011), Information and Privacy Commissioner Ontario website, p 4, www.ipc.on.ca/images/Resources/anonymization.pdf; and Paul Ohm, 'Broken promises of privacy: responding to the surprising failure of anonymisation' (2010) 57 *University of California Los Angeles Law Review* 1701, p 1744.

[8] National Statistical Service, *Confidentiality Information Series,* National Statistical Service website, www.nss.gov.au/nss/home.NSF/pages/Confidentiality+Information+Sheets.

In practice, ensuring that an information asset has been 'de-identified' for the purpose of s 6 of the Privacy Act will require organisations to adopt a risk assessment approach similar to the one involved in confidentialisation.

Why should personal information be de-identified?

De-identifying information:

- is required by the Privacy Act in specified circumstances
- enables an information asset that includes personal information to be disclosed in a de-identified form that makes it available for use by researchers and others[9]
- helps protect confidential information and data
- can lessen the risk that personal information will be compromised when an information asset is exposed to unauthorised access, use or distribution (that is, a data breach) — for example, where:
    - an intruder gains unauthorised access to organisation information
    - an employee internally accesses information without authorisation
    - data or an information storage device is lost or stolen.[10]

When to de-identify

Personal information does not need to be de-identified if it is required for a purpose permitted under the APPs. An example is a business using personal information about customers to deliver services requested by those customers.

De-identification may need to be considered if:

- particular sections of the organisation do not require access to the full information asset, but could make use of a de-identified version
- the information asset is being shared with another entity
- the information asset is being published.

For example:

- A real estate agency holds the personal information of clients, and provides access to the database to agents in the firm. The firm's public relations team is preparing a media release about recent property sales prices. The team requires access to information from the database, but not necessarily to information about the

[9] *Autism Aspergers Advocacy Australia and Department of Families, Housing, Community Services and Indigenous Affairs* [2012] AICmr 28 (www.oaic.gov.au/freedom-of-information/applying-the-foi-act/ic-review-decisions/2012-aicmr-28) the Australian Privacy Commissioner found that de-identification can be used to protect an individual's privacy in response to a request under the *Freedom of Information Act 1982* (Cth). The Commissioner has also released the *Privacy Guidelines for the Medicare Benefits and Pharmaceutical Benefits Programs* (www.oaic.gov.au/privacy/applying-privacy-law/legally-binding-privacy-guidelines-and-rules/privacy-guidelines-for-the-medicare-benefits-and-pharmaceutical-benefits-programs-issued-march-2008-effective-from-1-july-2008) under s 135AA of the *National Health Act 1953*, which establish when identifiable Medicare Benefits or Pharmaceutical Benefits claims information can be disclosed for the purposes of medical research.

[10] The OAIC *Guide to information security* (www.oaic.gov.au/privacy/privacy-resources/privacy-guides/guide-to-information-security) provides guidance on the reasonable steps entities are required to take under the Privacy Act to protect the personal information they hold.

identity of individual clients. Consequently, the team should only have access to a de-identified version of the database records.

- A medical research team creates a dataset about participants in a study. The dataset includes participant address information that could be used to identify or contact them. If the researchers want to share a de-identified version of the dataset with other researchers, the information should be removed or otherwise obscured as part of the de-identification process prior to sharing.

In determining the necessary level of de-identification organisations and researchers should consider:

- what kind of information or data is contained in the information asset
- who will have access to the information asset, and why
- whether the information asset contains unique or uncommon characteristics (quasi-identifiers) that could enable re-identification
- whether the information or data will be targeted for re-identification because of who or what it relates to
- whether there is other information or data available that could be used to re-identify the de-identified data or information
- what harm may result if the information or data is re-identified.

Further information on identification risk factors and methods for assessing identification risks is provided in the National Statistical Service's *Confidentiality Information Sheet 5 — Managing the risk of disclosure in the release of microdata*.[11]

In some cases de-identification of an information asset may reduce the utility of the asset. Nevertheless, de-identification may be necessary to minimise the risk of wrongly disclosing personal or confidential information.

How to de-identify

De-identification techniques should be carefully chosen, based on a risk assessment, to ensure that personal information is protected and that a de-identified information asset will still be useful for its intended purpose.

Removing or modifying personal identifiers such as a person's name, address and date of birth is an essential component of de-identification.

Other de-identification techniques that may be considered include:

- Removing or modifying quasi-identifiers (for example, significant dates, profession, income) that are unique to an individual, or in combination with other information are reasonably likely to identify an individual.
- Combining information or data that is likely to enable identification of an individual into categories. For example, age may be combined and expressed in ranges (25-35 years) rather than single years (27, 28). Extreme values above an upper limit or below a lower limit may be placed in an open-ended range such as an age value of 'less than 15 years' or 'more than 80 years'.

[11] Available at www.nss.gov.au/nss/home.nsf/pages/Confidentiality+-+Managing+the+risk+of+disclosure+in+the+release+of+microdata.

- Altering identifiable information in a small way such that the aggregate information or data is not significantly affected — a tolerable error — but the original values cannot be known with certainty. A variety of methods for rounding data are described in the NSS's *Confidentiality Information Sheet 4 – How to confidentialise data: the basic principles*.[12]
- Swapping identifying information for one person with the information for another person with similar characteristics to hide the uniqueness of some information. For example, a person from a particular town in Australia may speak a language that is unique in that town. Information about that individual's spoken language could be swapped with the spoken language information for another individual with otherwise similar characteristics (based on age, gender, income or other characteristics as appropriate) in an area where the language is more commonly spoken.
- Manufacturing 'synthetic data', which can be generated from original data and then substituted for it, while preserving some of the patterns contained in the original data.[13] For example, a synthetic dataset could be used to test a program that detects fraud. The synthetic data could be built to enable a test environment that replicates patterns from an authentic dataset of normal use and fraud but does not allow individuals from the original data to be identified. This allows systems to be tested with data that is realistic but poses less risk of re-identification.
- Suppressing data, which involves not releasing particular information that may enable re-identification, or deleting that information from the dataset. Data suppression may impair the utility of an information asset. Businesses and researchers may wish to consider first whether other de-identification techniques will adequately reduce the risk of re-identification.

Other steps that can be taken to manage and minimise the risk of re-identification include:

- Requiring the data or information receiver to sign a contract limiting the use and distribution of the information or data, and enforcing the terms of that contract. For example, the contract could include an assurance that the receiver will not attempt to re-identify the information, and will destroy any information re-identified unintentionally.
- Limiting access to a de-identified information asset, for example, allowing other organisations or researchers to view the data rather than providing a copy.
- Assisting analysis of data rather than providing access to it, for example, running an analysis of the data and providing the result rather than the raw data.

Further advice and examples of different methods of de-identification and methods of restricting access to data are provided in the NSS's *Confidentiality Information Sheet 4 — How to confidentialise data: the basic principles* and *Confidentiality Information Sheet 5 — Managing the risk of disclosure in the release of microdata*.[14]

[12] Available at www.nss.gov.au/nss/home.nsf/pages/Confidentiality+-+How+to+confidentialise+data:+the+basic+principles.
[13] United Kingdom Information Commissioner's Office, *Anonymisation: managing data protection risk code of practice*, published 2012, United Kingdom Information Commissioner's Office, Wilmslow, Appendix 2, p 53, www.ico.org.uk/for_organisations/data_protection/topic_guides/anonymisation.
[14] Available at www.nss.gov.au/nss/home.nsf/pages/Confidentiality+-+How+to+confidentialise+data:+the+basic+principles and www.nss.gov.au/nss/home.nsf/pages/Confidentiality+-+Managing+the+risk+of+disclosure+in+the+release+of+microdata.

Assessing the risks of re-identification

Before releasing a de-identified information asset, organisations and researchers should confirm whether de-identification has been successful. Potential risk assessment processes may include:

- Applying the 'motivated intruder' test — this test considers whether a reasonably competent motivated person with no specialist skills would be able to identify the data or information (the specific motivation of the intruder is not relevant). It assumes that the motivated intruder would have access to resources such as the internet and all public documents, and would make reasonable enquiries to gain more information.[15]
- Looking at re-identification 'in the round' — that is, assessing whether any entity or member of the public could identify any individual from the data or information being disclosed, either in itself or in combination with other available information or data.[16]

The risk of re-identification will depend on the nature of the information asset, the de-identification techniques used and the context of the disclosure. Relevant factors to consider when determining whether an information asset has been effectively de-identified could include the cost, difficulty, practicality and likelihood of re-identification.

Depending on the outcome of the risk analysis and the de-identification process, information and data custodians may need to engage an expert to undertake a statistical or scientific assessment of the information asset to ensure the risk of re-identification is low.

The risk of re-identification may shift as technologies develop and a greater amount of information assets are published or obtained by an organisation. Businesses and researchers should regularly re-assess the risk of re-identification and, if necessary, take further steps to minimise the risk. This may include:

- assessing whether a higher level of de-identification is required
- assessing whether the release of further de-identified information assets could potentially facilitate re-identification of already-published information.

Further Resources

The NSS, representing all government agencies, and led by the Australian Bureau of Statistics, has produced a *Confidentiality Information Series* which is designed to explain, and provide advice on, a range of issues around confidentialising data, including basic techniques to confidentialise data and manage risks. [17]

The Australian National Data Service also provides materials on techniques for de-identification and ethical considerations, including:

- *De-identifying your data*
- *Ethics, consent and data sharing*.[18]

[15] United Kingdom Information Commissioner's Office, p 22.
[16] United Kingdom Information Commissioner's Office, p 19.
[17] National Statistical Service, *Confidentiality Information Series,* National Statistical Service website, www.nss.gov.au/nss/home.NSF/pages/Confidentiality+Information+Sheets.
[18] Available at www.ands.org.au/resource/data-deidentification.html and www.ands.org.au/guides/ethics-working-level.html.

The information provided in this resource is of a general nature. It is not a substitute for legal advice.

For further information

telephone: 1300 363 992
email: enquiries@oaic.gov.au
write: GPO Box 5218, Sydney NSW 2001
Or visit our website at **www.oaic.gov.au**