office of the privacy commissioner
new south wales

Fact Sheet: Reasonably Ascertainable Identity

January 2017

FS2017/001

NSW privacy legislation regulates the handling of personal information by public sector agencies. Personal information in NSW privacy legislation is any information or opinion (including information or an opinion forming part of a database and whether or not recorded in a material form) about an individual whose identity is 'apparent or can reasonably be ascertained'.

This fact sheet offers interpretation and guidance on the meaning of 'reasonably ascertainable identity'. It offers practical tips to help users to determine if an individual's identity can be reasonably ascertained and discusses de-identification of data.

1. Key points

- A person's identity can be apparent or ascertained even if they are not directly named.
- The test is whether identification is possible, by any person (or machine) other than the subject themselves.
- The surrounding context, and other available information sources, can enable a person's identity to become apparent or ascertainable from the information or opinion, if no more than moderate steps are required to combine the data sources.
- If information has been de-identified to the point where re-identification is not possible, it is no longer 'personal information'.
- There are differing techniques for attempting de-identification. Care should be taken to choose the best method and prevent re-identification.
- When in doubt, assume that your data will meet the definition of 'personal information' and apply privacy protections accordingly.

2. No names needed

The test for identifiability includes not only if the person's identity is "apparent" from the information in question, but also if their identity can "reasonably be ascertained". Because of this, the absence of a name does not mean a set of information is not "personal information".[1]

In one case, a document about a staff member referred to a person only by his first name. As there was only one person with that first name out of the 400 possible staff, the individual was identifiable.[2] Most cases which have reviewed the concept of identifiability have involved information which does not include any name, but which includes enough other details to identify the subject.[3]

The NSW Civil and Administrative Tribunal (NCAT) has found that even if identification is not "likely", the standard is simply whether or not it is "possible".[4]

The dictionary meaning of 'ascertainable' is "able to be found out by trial, examination or experiment".[5]

3. Identification by whom?

The legislation does not make clear who is supposed to be able to ascertain the subject's identity – the holder of the information, the subject themselves, a particular third party audience or the world at large?

We suggest that the test is whether a person can be identified by any individual, entity or machine, other than themselves.[6] This could be the organisation holding the data, or any third party.

A previous NSW Privacy Commissioner applied this test in an investigation reported to Parliament. Specifically, the Education Minister had disclosed information about an unnamed student: the student's gender, age, the year in which he was enrolled, a description of an event involving the student, the date on which the

---

1 Privacy NSW, Special Report to Parliament, *Student A and the Minister for Education*, 7 May 2002

2 *AQK v Commissioner of Police, NSW Police Force* [2014] NSWCATAD 55

3 *AFW v WorkCover Authority of New South Wales* [2013] NSWADT 133

4 *Field v Commissioner of Police, New South Wales Police Force* [2015] NSWCATAD 153 at [75]. Although this case was brought under the *Government Information (Public Access) Act 2009* (GIPA Act), the GIPA Act adopts a similar definition of personal information, including the phrase "whose identity is apparent or can reasonably be ascertained from the information or opinion".

5 *Ben Grubb and Telstra Corporation Limited* [2015] AICmr [35] 1 May 2015, at [68]. Note that although this determination was set aside on appeal, the meaning of 'ascertainable' was not at issue in the appeal; see *Telstra Corporation Limited and Privacy Commissioner* [2015] AATA 991, 18 December 2015.

6 This is similar to the simple test proposed by a former New Zealand Assistant Privacy Commissioner: "if the person can be recognised by others than themselves"; see Katrine Evans, "Personal information in New Zealand: Between a rock and a hard place?", paper for *Interpreting Privacy Principles: Chaos or Consistency*?, Symposium, Sydney, 17 May 2006.

student was removed from the school, and the date on which a school assembly was planned.

This was enough information for a particular third party audience – the "school community" – to be able to ascertain the student's identity, even though the Minister himself did not know the student's identity.[7]

NCAT has found that identification can be performed even by a machine, in circumstances in which a human could not make the requisite link. A document published online, which had names and other details redacted such that they could not be read by the human eye, was nonetheless readable by search engine webcrawlers, and thus identification of the individual was established.[8]

4. Refer to other sources – within reason

The then Privacy Commissioner advised that:

> *"Constructive identification does not require all the parties to be in full possession of all the details, but results in the public revelation of discrete details which when taken together enable an identity to be constructed".*[9]

Therefore the likelihood of identification should not be considered in a vacuum. A person's identity may be ascertainable with reference to other sources.[10]

The Appeal Panel of the former NSW Administrative Decisions Tribunal (ADT), the predecessor body to NCAT, has confirmed that "depending on the circumstances, sources of information other than the information or opinion which contains the personal information, may be consulted to ascertain the person's identity".[11]

The Australian Privacy Commissioner likewise has made a determination which found that data which "may" link data to an individual, even if it requires some "cross matching … with other data" in order to do so, will meet the test required for "personal information".[12]

For example:

- A photograph of a residential apartment, when taken together with information from a local Council's files, was sufficient to enable the apartment's owner to be identified.[13]
- The publication of an address enabled the identification of the couple holding the lease for the property at that address, by way of "reasonable means", namely "simple internet searches".[14]
- CCTV footage of an incident in a shopping centre, because it could be combined with "publicly available information about the identities of (the individuals filmed on CCTV) for example through social media", was found to meet the test for "personal information".[15]
- The statement "(a named individual) left general practice in order to care for a child who required intensive attention because he had diabetes" was found to be sufficient to ascertain the identity of the child.[16]

Agencies therefore should be mindful of the ease with which data from different sources can be matched, when assessing whether or not a person's identity may be ascertainable from the information held. However the extent of cross-referencing contemplated by the definition of "personal information" is likely limited.

The Tribunal has applied a test of whether or not "more than moderate steps" is necessary to match data from different sources, in order to ascertain an individual's identity.[17]

---

7 Privacy NSW, Special Report to Parliament, *Student A and the Minister for Education*, 7 May 2002
8 *AIN v Medical Council of New South Wales* [2016] NSWCATAD 5
9 Privacy NSW, Special Report to Parliament, *Student A and the Minister for Education*, 7 May 2002
10 *WL v Randwick City Council (No. 2)* [2010] NSWADT 84
11 *Office of Finance and Services v APV and APW* [2014] NSWCATAP 88 at [54]
12 *Ben Grubb and Telstra Corporation Limited* [2015] AICmr [35] 1 May 2015, at [52], [53]
13 *WL v Randwick City Council* [2007] NSWADTAP 58
14 *APV and APW and Department of Finance and Services* [2014] NSWCATAD 10 at [15]
15 *Field v Commissioner of Police, New South Wales Police Force* [2015] NSWCATAD 153
16 *AIN v Medical Council of New South Wales* [2016] NSWCATAD 5
17 *AIN v Medical Council of New South Wales* [2016] NSWCATAD 5

5. De-identification and Re-identification

Although in research contexts, 'de-identified' or 'anonymised' is sometimes used to describe data for which codes or numbers have replaced names, in the privacy context, for information to be truly 'de-identified', re-identification must be extremely difficult, if not impossible.

The then Privacy Commissioner advised that:

> "De-identified information is information from which identifiers have been permanently removed, or where identifiers have never been included. De-identified information cannot be re-identified".[18]

'Identifiers' can include direct identifiers (such as name, address, telephone number or Tax File Number), or indirect identifiers which allow information to be connected until an individual can be singled out (e.g. client number, vehicle registration number, MAC address, or demographic data such as date of birth and sex).

Coded information will remain potentially re-identifiable to a person or body with the means to link the code back to other identifying details. For example, datasets in which each individual's name has been replaced with a statistical linkage key[19] may still enable identification by anyone with access to the dataset, if they are searching for a known individual. Coded information may therefore still be "personal information".

A previous Privacy Commissioner advised that in order to de-identify information, in the context of health information to be used for research, removing just a person's name and address will likely be inadequate, "particularly if there are unusual features in the case, a small population, or there is a discussion of a rare clinical condition".[20]

He also warned that re-identification may arise "in the publication of non-identifying statistical data, which may nevertheless be aggregated with other data to effectively re-identify some individuals".[21]

The re-identification risks posed by public access to large datasets has been evidenced a number of times.

For example:

- A 2000 study linked public 'anonymous' health insurance data of public servants with electoral rolls (name, date of birth, sex, postcode) to identify the Massachusetts Governor's diagnoses and prescriptions.[22]
- In 2006, search engine provider AOL released 'anonymous' web search records for 658,000 users. Journalists linked search terms to identify users and contact them.[23]
- In 2013, a Harvard professor re-identified the names of more than 40% of a sample of 'anonymous' participants in a high-profile DNA study.[24]
- In 2014, 'de-identified' data on 173 million taxi trips made in New York City was released under FOI. Within hours the 'hashed' driver and vehicle numbers were re-identified, and then the GPS data was matched with other publicly available data to identify specific trips taken by known individuals.[25]

There is no 'correct' way to de-identify data. Care should be taken to ensure that the most suitable methodology is chosen, considering the type of data, what it needs to be used for, and what other data sources might be available to the recipient.

---

18 Privacy NSW, *Handbook to Health Privacy*, August 2004, p.50

19 A statistical linkage key enables two or more records belonging to the same individual to be brought together. For example, human services agencies often use a key generated from the 2nd, 3rd and 5th characters of a person's family name, the 2nd and 3rd letters of the persons' given name, the day, month and year when the person was born and the sex of the person, concatenated in that order.

20 Privacy NSW, *Statutory Guidelines on Research*, September 2004, p.8

21 Privacy NSW, *Privacy Contact Officer Newsletter*, June 2002

22 See http://groups.csail.mit.edu/mac/classes/6.805/articles/privacy/sweeney-thesis-draft.pdf

23 "Forget Facebook privacy, your digital life is being monitored", ADITYA CHAKRABORTTY, 26 May 2010, Sydney Morning Herald – available at http://www.smh.com.au/opinion/society-and-culture/forget-facebook-privacy-your-digital-life-is-being-monitored-20100525-wavc.html#poll

24 See http://www.forbes.com/sites/adamtanner/2013/04/25/harvard-professor-re-identifies-anonymous-volunteers-in-dna-study

25 See http://www.salingerprivacy.com.au/2015/04/19/bradley-coopers-taxi-ride-a-lesson-in-privacy-risk/

Techniques could include:

- aggregating data to a high degree
- removing some variables
- coding or pseudonymising (replacing identifiers with unique, artificial codes)
- hashing (one-way encryption of identifiers)
- generalising (for example, by replacing precise date of birth with an age bracket)
- suppressing (for example, by replacing some values with 'missing')
- micro-aggregating (for example, group in fours, so ages 31, 33, 33 and 34 each become 32.75)
- data-swapping (for example, swap salaries for people within the same postcode, so the aggregate is still valid), or
- differential privacy (adding 'noise' to the data, to hide whether or not an individual is present).

The UK has published a comprehensive guide on these and other methodologies for de-identifying data, without affecting the integrity of the data for its intended use.[26]

6. Conclusion

Identifiability is not a black and white concept. There are many shades of grey between data from which individuals are readily identifiable, and data that has been entirely anonymised.

When in doubt, assume that data will meet the definition of "personal information", and apply the relevant privacy principles accordingly.

For more information

Contact the Office of the Privacy Commissioner

PO Box R232, Royal Exchange, NSW 1225

Level 3, 47 Bridge Street, Sydney NSW 2000

Telephone: (02) 8258 0066

Email: privacy@ipc.nsw.gov.au

Website: www.ipc.nsw.gov.au/privacy

NOTE: *The information in this fact sheet is to be used as a guide only.*

Legal advice should be sought in relation to individual circumstances.

[26] UK ICO, *Anonymisation: Managing Data Protection Risk Code of Practice*, November 2012; see https://ico.org.uk/media/for-organisations/documents/1061/anonymisation-code.pdf