Office of the Information Commissioner
Queensland

Applying the legislation

GUIDELINE *Information Privacy Act 2009*

Dataset publication and Privacy

Queensland government agencies hold large amounts of information about Queensland and the community. There are a number of laws which apply to this information, including the *Right to Information Act 2009* (Qld) (**RTI Act**), the *Information Privacy Act 2009* (Qld) (**IP Act**) and the *Public Records Act 2002* (Qld).

Government-held information should be proactively disclosed unless doing so would be contrary to the public interest.[1] This principle recognises that government manages its information holdings on behalf of the Queensland community.

The recently launched Open Data portal[2] involves Government agencies publishing datasets and information holdings on an ongoing basis.

Privacy and datasets

Many datasets held by government contain personal information. This personal information must be protected in the way set out in the IP Act and the privacy of individuals taken into account when publishing data.

The privacy principles in the IP Act set out how government agencies can collect, store, use or disclose personal information both within and outside Australia. Publishing personal information is a 'disclosure'[3] and posting personal information online can result in the transfer of that information outside Australia[4].

What is personal information?

Personal information is any information associated with an identifiable living individual.

The IP Act defines personal information as:

> *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 reasonably be ascertained, from the information or opinion*.

[1] *Right to Information Act 2009* (Qld)
[2] http://data.qld.gov.au/
[3] Disclosure is defined in section 23(2) of the IP Act as giving personal information, or the means to obtain personal information, to an entity that did not know the personal information or is not in a position to be able to find it out.
[4] See section 33 of the IP Act.

Datasets containing personal information

The fact that a dataset contains personal information about individuals should not in itself prevent publication. The dataset can be published once appropriate steps are taken to de-identify the data.

What is de-identification?

De-identification refers to various mechanisms for removing or manipulating data so that individuals are no longer apparent**,** or reasonably ascertainable, from the data. Data is de-identified when it can no longer be linked to an identifiable individual.

Dataset de-identification allows the release of useful information while respecting individual privacy as the information ceases to be personal information. Once it ceases to be personal information the IP Act no longer applies to it.

When is information personal information?

For information to be personal information two criteria must be satisfied.[5]

- it must be *about* an individual
- the individual's identity must be apparent or reasonably ascertainable from the information.

When is information *about an individual*?

Information is about an individual if it reveals something about that individual. A connection between the information and the individual may not be enough to make it *about* the individual. Some information will obviously be about an individual, such as names, contact details, medical records; other information, such as information about a property which does not reveal anything specific about the owner, may not be. Agencies will need to consider the context in which information appears to decide if it *about* an individual.

When is identity *apparent or reasonably ascertainable*?

An individual's identity is apparent if you know who the individual is simply by looking at the information, for example the individual's name is included.

However, stripping out obviously identifying information, such as full names, addresses, or e-mail addresses may not be sufficient; the identity of the individual may still be 'reasonably ascertainable'.

An individual's identity is ascertainable if the individual could be identified by using or cross-referencing other available information. Whether the identity is

[5] See *Mahoney and Ipswich City Council* (Unreported, Queensland Information Commissioner, 11 June 2011)

reasonably ascertainable will depend on factors such as the number of steps required to work out who the individual is.[6]

> **De-identification in an online world**
>
> The more information about an individual that is readily available, the greater the chance that the individual's identity could be reasonably ascertained. Given the large amounts of personal information available online, including some published by the individual over which the government has no control, agencies will need to carefully consider whether the datasets they are considering for publication can be de-identified to the point that individual identities are no longer reasonably ascertainable. Determining whether de-identified information can be re-identified will not always be a simple task.

Re-identification risk

The IP Act does not require an absolute guarantee that individuals will never be re-identified from de-identified data, only that the identity not be *reasonably ascertainable*. How involved the cross-referencing would need to be to qualify as 'reasonable' would need to be determined as part of the agency's risk assessment processes when it makes the decision to publish a dataset.

Factors which can impact on 'reasonableness' include:

- the amount of alternative information about the individuals contained in the dataset that is publicly available - for example census data and public registers
- the ease of access of the alternative information – digital vs paper records
- the level of detail provided – for example, date of birth is more specific than month of birth or year of birth
- the number of steps and the associated amount of time, resources and effort required to identify an individual
- how up-to-date the information is – more current information can be more identifying
- intimate knowledge – the extent to which only people with personal knowledge of individuals such as family or close friends would be able to identify an individual.

De-identification techniques and risk assessment

A range of techniques and tools exist for de-identifying datasets and assessing the risk of re-identification. Some of these are outlined in the OIC Guideline: *Publishing Datasets and De-identification Techniques.*

[6] For more information on the risks of re-identification refer to OIC's guideline: *Publishing Datasets and Risk Assessment* at www.oic.qld.gov.au

For additional information and assistance please refer to the OIC's other guidelines on dataset publication, or contact the Enquiries Service on 07 3234 7373 or email enquiries@oic.qld.gov.au.

This guide is introductory only, and deals with issues in a general way. It is not legal advice. Additional factors may be relevant in specific circumstances. For detailed guidance, legal advice should be sought.

If you have any comments or suggestions on the content of this document, please submit them to feedback@oic.qld.gov.au.

Published 18 February 2013 and Last Updated 18 February 2013

Changes to legislation after the update date are not included in this document