Comments on the proposed amendments to the Privacy Act to criminalise re-identification

Justin Warren

December 2016

What Is Personal Information?

The discussion of the proposed Amendments (the Amendments) makes an important error regarding personal information: in the Explanatory Memorandum, the Government refers to *government datasets*. These datasets may be held by government entities, but the data in them is about citizens. It belongs to them, not to the government.

The information is personal, therefore it belongs to the person it is about. The person has (one hopes) consented to providing the personal information to the responsible agency in exchange for services or to enable that person to receive a benefit. The responsible agency does not own the information.

Privacy Implies Control

An important aspect of privacy that is often overlooked is that of control. In order to maintain privacy, I must be able to decide to whom I disclose information, when, and why. If I am not able to control who can see my private information, I do not have privacy.

For example, I disclose personal information about myself and my health to my doctor, because they need that information in order to help me manage my health. This is a choice I make. Without that information my doctor would not be able to do their job, and I would not be able to receive the benefit of improved health.

However, my doctor is not then free to do whatever they like with my information, and it is not available for anyone to see. I provided it to them for a limited purpose, and I expect that if they wish to share that information with anyone else that they must first seek my consent. If my doctor can share my information with whomever they like without my consent, how can I be said to have privacy?

The responsible agency owes a duty to each individual to keep their personal information secure and private. That was part of the implied (and sometimes explicit) contract between the two parties when the information was entrusted to the agency. If they do not owe this duty to me, then I cannot be said to have privacy in respect of that information.

The agency to which the information is entrusted should bear the burden for keeping that information private. Indeed, why should it be otherwise?

Impact of a Privacy Breach

The explanatory memorandum makes clear that the Government is keenly aware of the impact of a breach of privacy.

- Paragraph 21: *"the Government also recognises that the privacy of citizens is of paramount importance"*
- Paragraph 27: *"Unlike some other offences, breaches of privacy cannot be easily remedied because the publication of private information cannot be reversed."*
- Paragraph 48: *"...releases of private information can have significant consequences for individuals beyond their privacy and reputation, which cannot be easily remedied."*

Given the dire consequences of a privacy breach, we must understand how such a breach occurs and could be prevented.

Who Is Harmed, and How?

A breach of privacy as discussed by the Amendments is that of individuals whose personal information has been disclosed. The privacy of the responsible agency has not been breached. Therefore it is the individuals whose privacy has been breached that have been harmed.

The harm has been caused by the publication of private information without that individual's consent. Who has published the information, and to whom could consent have been given?

Consent

If individuals supplying personal information consent to its publication with them being identified, then no privacy breach can occur if that information is published, and there would be no need to de-identify the dataset. The very fact that a dataset has attempted to be de-identified implies that consent for personal identification has not been provided by at least some individuals.

If consent to identify them was not given, then a privacy breach will have occurred if that information is published.

It is possible that individuals could provide their consent to the Noticer, but given the dataset is published to the general public, usually on the Internet, this implies that all individuals in the dataset would need to have given their consent to each and every possible Noticer throughout the world. This is nonsensical, as it is functionally equivalent to giving consent to the responsible agency to identify individuals when publishing the dataset, and removes the need to de-identify the dataset in the first place.

The private information was entrusted to the responsible agency by the individual. Therefore, the responsible agency is the one to whom consent was either provided, or not, and they are responsible for maintaining individual privacy.

What Caused The Harm?

If no attempt at de-identification was made, then it is clear that the harm was caused when the dataset was published by the responsible agency. What if a trivial and ineffective attempt at de-identification was made before publication? Is the harm caused by publication of a not-really-de-identified dataset, or when someone notices that the dataset is trivially re-identified?

I argue that the harm is caused by the publication of the dataset by the responsible agency, not whoever notices that the dataset is trivially re-identified (the Noticer). The potential risk of re-identification is well known, therefore to publish a dataset that has been trivially de-identified could be construed as negligence. Further, the personal information was entrusted to the responsible agency, not the Noticer. Thus the duty is owed to an individual by the agency, not by the Noticer. The agency has failed in its duty to adequately safeguard the personal information entrusted to it by publishing a dataset that was not de-identified.

What if a great deal of effort and care is taken by an agency to attempt to de-identify a dataset, yet that attempt proves ineffective? The end effect is that a privacy breach has still occurred, because personal information about an individual has been disclosed without that person's consent.

The Government acknowledges this fact in its explanatory memorandum:

> *"However, with advances in technology, methods that were sufficient to de-identify data in the past may become susceptible to re-identification in the future."*

Here the situation becomes more complex, because what is 'sufficient' is ill-defined, as is the time-frame of

'the future'. Is a method sufficient if it prevents re-identification for a week after publication? A year? A decade? Is an individual's privacy less breached if the unauthorised disclosure of personal information happens after a delay?

The fact remains that it is possible to prevent a privacy breach by simply not publishing the dataset at all. The combination of both the publication of a dataset *and* the de-identification method being insufficient are required for a privacy breach to occur.

Who Is Responsible?

Of key importance is who should be responsible for preventing a breach of privacy. Clearly the responsible agency is the one responsible for publishing the dataset as it is entirely within their control to simply not publish.

However, the Government is keen to publish datasets. Doing so

> *"...enables the government, policymakers, researchers, and other interested persons to take full advantage of the opportunities that new technology creates to improve research and policy outcomes."*

Clearly the privacy of citizens is not of *paramount* importance, but is considered as part of the discussion on how to obtain benefits from publishing datasets. If privacy were paramount, the datasets would not be published at all, as that would reduce the risk of privacy breaches.

A key phrase in the Amendments is *"on the basis that it was de-identified".* The Government appears to think that merely attempting to de-identify the data, regardless of how effective the attempt was, is sufficient to make a dataset 'de-identified'. There is no discussion of how much effort is required to make a dataset 'de-identified'. What constitutes 'sufficient'? An agency could use a trivial, ineffective method, such as converting letters to their position in the alphabet (a→1, b→2, etc.) and that could be construed as having been 'de-identified' simply because the agency says so.

Anyone who then notices that the method was ridiculously ineffective would now be guilty of a crime.

On the assumption that individuals have consented to their personal information being published in a de-identified form, the burden for de-identification should fall on the agency to whom the information has been entrusted. The agency is responsible for preventing the information from being published in a way that can identify individuals, and it is they who should be made to use sufficiently robust methods for keeping individuals from being identified. If they are unable to do so reliably, then they are not sufficiently competent to be publishing datasets containing personal information and should refrain from doing so.

The Amendments provide no deterrents for responsible agencies publishing poorly de-identified datasets. Instead, the Amendments appear structured to prevent any other Australian based entity from noticing—or even attempting to notice—that the responsible agency has failed to adequately protect the privacy of individuals who trusted the agency with their data.

The Government is attempting to make it a crime to cry "The Emperor has no clothes!" when responsible agencies parade our personal information down the street complete naked.

Scale of a Breach

The datasets to which the Amendments refer are frequently published on the Internet, particularly via the data.gov.au portal set up for this purpose. As the Committee is no doubt aware, the Internet is a global network, while Australia's laws are only applicable in Australia. The vast majority of the Internet exists outside of Australia.

If a dataset is not sufficiently de-identified, entities outside of Australia can re-identify them without fear of prosecution in Australia. The threat of Australian law provides zero deterrent for the majority of threats to

individual privacy that Australian citizens face from poorly de-identified datasets being published.

It is clear, therefore, that individuals will suffer irreversible harm if these not-really-de-identified datasets are published, and the deterrent effect is limited at best. The Amendments thus fail to achieve their core goal, while shifting responsibility away from the agencies who are best placed to safeguard the privacy of citizens, and should be required to do so.

Reverse Burden

The reverse burden identified in paragraphs 39-42 of the Explanatory Memorandum ignores the fact that mounting a defence against a criminal prosecution is costly, and that the cost impact on an individual or small business may prove disastrous, even if the defence is successful. Paragraph 42 makes clear that the Bill expects that authorities will not abuse their power by attempting a prosecution that will ultimately fail. This is cold comfort to an individual who successfully defends themselves against a doomed prosecution only to be bankrupted in the process.

To imply that because it is *easy* to prove one's innocence that one should therefore be *forced* to prove one's innocence is offensive, and counter to the very principle that one is innocent until proven guilty. To blithely state that prosecution is unlikely because the authority will not likely succeed ignores the power dynamic between the individual and the state. An unscrupulous authority could use the powers granted by this Bill to financially punish an entity for embarrassing the authority, even if the entity did not commit an offence.

Retrospective Law

The Government is attempting to use the threat of prosecution for an offence that does not yet exist to act as a deterrent for behaviour it has unilaterally deemed unacceptable.

In Australia, laws should not be effective until they have been discussed by the citizens, debated by Parliament, passed or not as the Parliament sees fit, and finally given Royal assent. The ability for a Government department to unilaterally give effect to a law by issuing a hastily written press release is deeply disturbing, and sets a dangerous precedent.

Private Datasets

Despite the broad concern the Government purports to have regarding breaches of privacy, the Amendments only deal with the re-identification of Government datasets and make no mention of individuals being re-identified from private datasets that are published. This is curious.

Private organisations hold a great deal of sensitive and personal information on individuals. If privacy is of such great concern to the Government, why are private datasets not covered by this legislation?