This submission is published at my blog.

Submission to the Inquiry into the Privacy Amendment (Re-identification Offence) Bill 2016.
Author: Rosie Williams BA (Sociology)

I write this submission as an open data/budget transparency and privacy advocate and a citizen journalist, researcher and programmer. I researched and wrote the CensusFail submission to the Inquiry into the 2016 census in which I made salient points regarding privacy and data sharing of administrative data. The Inquiry into the 2016 Census failed to adequately address the privacy risks posed by government agencies sharing personal information without informed consent or appropriate de-identification practices.

The Amendment has had a chilling effect on discussion of re-identification research, discussion and debate. This means that what is publicly known and discussed about re-identification research and risk in Australia may not be the true picture. This can only have a detrimental affect on the general public who will suffer greater risk from practice which is informed by questionable information.

While there is wide dissent to the re-identification amendment, few people at this time of the year with so many other Inquiries and consultations on, have the time to give to a submission to this Inquiry. This should not be taken as disinterest however. I personally organised an event at which Chris Culnane and Vanessa Teague presented their work and the response to (the Bill that is the subject of this Inquiry) in Melbourne which succeeded in attracting 50 people to hear about the topic. Subsequent local privacy themed events in recent days have also seen the re-identification Bill come up for discussion. It is clear that many people would like to do more to demonstrate their disillusionment with the criminalisation of re-identification research but it is a difficult time of year for people to engage with the political process.

As a member of the open data community in Australia and as a Board member of Open Knowledge Australia, it is concerning that the Bill might result in liability to participants and organisers of hackfests. The point of events such as GovHack and HealthHack are to use open public data, often combined with one another to create ‘mashups’ and web projects intended to meet a business or social need.

When datasets are joined together this raises the risk of re-identification however it would be very difficult for participants or organisers to imagine all the ways each dataset could be combined with other data (including data from the private sector) to assess the risk of re-identification. This risk is also unable to be calculated when it is considered the amount of information that is readily available through social media and search engines that relates to individuals.

One way to limit the chances of re-identification is not to publish datasets which contain unit record level information about individuals as open data, rather to limit access to such datasets to trusted researchers as is recommended by the Productivity Commission in its Data Access and Use Inquiry Report. It should be noted here that before one considers the technical risks arising from sharing of administrative datasets, it must first be considered whether there is any legislative authority to share data collected for administrative purposes under specific legislation with datasets collected under other legislation (or share it at all) as there may well be no legislative authority for this.

It would make a lot more sense (both legally and technically) to provide unit record level data only to trusted researchers rather than to spread it far and wide and criminalise re-identification of that data (which should not be possible to begin with). There is the obvious consideration that criminalising the mere observation that a dataset might be easily de-identified is far more likely to result in the government NOT being told that a dataset is at risk than protecting the public. The submission to this Inquiry from the Melbourne University researchers made it clear that the re-identification Bill is likely to impact their willingness to sound the alarm when data published by the government can be used to identify individuals. This result of the Bill would seem counter-intuitive to solving the problem.

In fact it is easy to conclude that the re-identification Bill must be aimed at penalising people for telling the truth about government mistakes in order to save the government from legal liability or embarrassment. Legislation presented to parliament should be designed to protect Australians from criminals (and negligence), rather than protect the government from people telling the truth about the government's incompetence and poor policy.

Re-identification is (one would hope) a technically difficult task. It seems to be common sense that if someone wanted to re-identify data that had been published by the Australian government and avoid legal sanction, they need only to do it outside the reach or Australian law or knowledge. It is unclear how the re-identification Bill can deal with the ease with which adequately equipped persons could get around the Bill as it stands. On balance, the Bill appears to have far more negative than positive implications for the security of personal information that Australians must provide to the government in the course of our daily business.