Submission to the Review of the Privacy Act 1988

Dr Chris Culnane, Associate Professor Ben Rubinstein[1]

Executive Summary

At the core of effective privacy protection is sound and technically coherent definitions of both personal information and the techniques used to protect such information. This submission examines the current state of definitions within the *Privacy Act 1988*, finding ambiguity and contradictions that permit unwanted exploitation of individual's information. It is *absolutely essential* that the reform of the *Privacy Act* addresses these problems. Without redress, the privacy protection framework within Australia will continue to sit atop a shaky foundation.

Once appropriate definitions have been introduced a framework for explicit consent should be established. Much can be learnt from the challenges the EU has faced in advancing consent; in Australia, similar development should accompany legislative changes with appropriate technical regulations to ensure consent is obtained fairly and in an efficient manner for the consumer.

The right to deletion, whilst seen by some as controversial, should be viewed as a natural consequence of consent withdrawal. Rather than trying to formulate specific deletion requirements, development of stronger consent withdrawal methods, accompanied by addressing the erroneous equivalence in the option to delete or de-identify, will provide the necessary balance between consumer rights and practical application.

As part of providing a stronger consent framework, increased transparency about the transfer and sharing of data needs to be enforced. Today's opaque data economy prevents consumers from exercising what little power they have. Organisations should not fear providing details of their privacy protection techniques, or the nature of the data they hold and use. Such transparency is essential to building confidence.

Reformation of the *Privacy Act* should address the adversarial and exploitative nature of today's data ecosystem. Consumers are ill equipped to protect their privacy and enforce what few rights they have. Expanded rights from a reformed *Privacy Act* must be accompanied by greater capability and resourcing of the OAIC. The effectiveness of privacy legislation is dependent on effective and efficient enforcement, with expertise spanning both legislative and technical, something that has been lacking in Australia. To provide further redundancy in this regard stronger direct action rights need to be legislated, through specific legislation or a privacy tort.

---

[1] The author is with the School of Computing and Information Systems, University of Melbourne. The opinions expressed in this submission are the authors' own and do not reflect the views of The University of Melbourne. The authors are researchers in data privacy. With Teague they reidentified the 2016 Medicare/PBS and 2018 Myki card releases.

Summary Recommendations. This submission makes the following recommendations:

1. Adopt “related” in favour of “about” as a threshold for personal information.
2. Privacy protection should not rely on eliminating uniqueness of individuals in a dataset.
3. The sharing of data, including data deemed to be no longer identifiable should require consent.
4. Privacy protections should not rely on de-identification, as defined by stripping of PII or eliminating uniqueness of individuals’ records in a data release.
5. Retire the qualifier of “reasonably” as threshold for identifiability.
6. Consistency of related definitions should be treated as paramount, to avoid serious, unintended consequences.
7. Where technical measures are taken to protect privacy in data sharing or release, the *Privacy Act* should prefer those that provide security properties that assert protection against an identified threat model.
8. Wherever possible, implementations of privacy enhancing systems should be made open.
9. Government guidance should not recommend Five Safes in its current form.
10. De-identified data should be considered to remain personal information until sufficient proof is provided that demonstrates adequate privacy protection.
11. Stronger rights for direct action should be afforded to individuals.
12. The government should clarify the retroactive state of the now expired *Privacy Amendment (Re-identification Offence) Bill 2016*.
13. A review of government data releases between 29th September 2016 and 1st July 2019 should be undertaken to evaluate if any assumed operation of the *Privacy Amendment (Re-identification Offence) Bill 2016* in their evaluation of “reasonably identifiable”.
14. Follow the EU GDPR lead in framing the provision of consent, but accompany it with further technical regulations covering the obtaining of such consent.
15. Require the adoption of an automated consent provision framework that allows consumers to set defaults on a browser or device level.
16. A collecting party must maintain records of data sharing, including derived data deemed no longer identifiable, so that consent withdrawal can propagate to all parties in possession of the data or derived data. A consumer need only notify the original collecting party to have consent withdrawn throughout the chain.

17. Consumers must be able to access the record of who their data has been shared with, and be able to withdraw consent for the usage of their data, or their derived data, from those downstream parties directly.

18. Transfer of ownership of data resets consent, requiring the new owners to reobtain consent from the individual. By default, where no consent can be obtained or the consumer does not respond, consent is assumed to have expired and the data must be deleted.

19. End the *Privacy Act* exemptions to small businesses and political parties.

20. Remove references to “deletion or de-identification”, replacing with only “deletion”.

21. Withdrawal of consent should lead to deletion by default, with decisions on whether consent can be withdrawn used as a threshold for deletion.

22. Establish a data provenance scheme with associated public artefacts. Such a scheme would require the publishing of meta-data about datasets held, providing reference back to the original collecting party to facilitate granular consent decisions by the individual.

Table of Contents

1. Definition of Personal Information

The definition of personal information is naturally at the core of the *Privacy Act*. Existing limitations hamstring the Act, leading to its inability to adequately protect individual privacy. The current definition is too narrow in what information it encompasses, and too broad in determining identifiability. This leads to data either not falling within the definition, or being moved out of the definition through the use of naive de-identification techniques that offer little technical protection and superficial compliance with the *Privacy Act*.

1.1. What is personal information?

Traditionally the notion of personal information has been closely tied to identity and identifiers, as if such types of information are prerequisite to identifiability. We view this framing to be too narrow: the notion of identifiability is not tied to identity but recognition. Specifically, is it possible to recognise the same individual in a future data set, time, or context? Any data that distinguishes an individual from the crowd can act as a pseudo-identifier, causing recognition of the individual, leading to their being subject to profiling, targeting, and ultimately invasion of their privacy.

Attempts at categorising data as either technical, personal, or non-identifiable are doomed to fail, as the methods for analysis, in particular big data analytics, no longer need to use such categorisations, as we demonstrated when we re-identified the Myki travel data in 2018[2], where touch on, touch off locations and times were sufficient to re-identify people. The information that causes someone to be identifiable can be technical information. In the Myki case, such data points in isolation were not unique but when taken as a set were. Information need not even be about an individual - in this instance it was about an object the person possessed, their Myki smart card.

The definition adopted by the EU in the GDPR goes a long way to addressing weaknesses in definitions of personal information, crucially shifting to a notion of being “related” to, as opposed to “about”. Adopting such an approach in Australia would not only go a long way to correcting current deficiencies, but could also facilitate greater trading opportunities with the EU through a compatible and equivalent privacy protection regime.

Recommendation 1. Adopt “related” in favour of “about” as a threshold for personal information.

[2] Chris Culnane, Benjamin I. P. Rubinstein, and Vanessa Teague. “Stop the Open Data Bus, We Want to Get Off”. Aug. 2019. arXiv: 1908.05004 [cs.CR]. https://arxiv.org/pdf/1908.05004

1.2. Inferred Information

The tensions created through inferred information are rooted in the failure to consider the notion of group privacy. By permitting identifiable data to be derived into data about groups, which exists outside the protection of the *Privacy Act*, avenues for exploitation are created.

Group privacy is an issue if all members of a statistical group share the same attribute, then merely knowing an individual is a member of that group permits certain inference of information. Whilst protecting group privacy can be a challenge, since at its extreme it could prevent sharing of any data, the current approach is too weak. When evaluating the privacy risk associated with group privacy one must consider what could be learnt if membership of the group can be determined. This could be innocuous information, or it could be a highly sensitive attribute.

For example, if aggregate group data comprised a single aggregate statistic such as the number of individuals with a particular medical condition, then the only way to determine group membership would be to know that the individual suffers from that condition. As such, the information gained through association is nil. However, if the aggregate data includes multiple attributes, for example, the number of people with that condition who live in a particular street in a certain age bracket, association might be achieved without needing to know all the attributes. There is the possibility that through the use of street and age, the group membership can be determined, allowing the association and inference of the sensitive medical attribute. It is not possible to say in what circumstances such inferences will be able to be drawn, since in isolation the data released as de-identified[3] did not uniquely identify anyone. Any inferences will be context and data specific, making evaluation and regulation of such approaches challenging, particularly since the group as a whole may not share the same views on privacy and how their data can be used.

Although related to privacy risks, uniqueness in a data release does not directly relate to privacy. Homogeneity attacks like those described above have largely discredited the framework of *k*-anonymity[4] which measures levels of uniqueness. While uniqueness in a subsample does not imply with certainty uniqueness in a population.

Recommendation 2. Privacy protection should not rely on eliminating uniqueness of individuals in a dataset.

Even when data has been derived into a form considered not to be about identifiable individuals, it is important to recognise that such data was derived from identifiable data and as such only exists because of the original collection of identifiable data. There should not be a default right for the collector to derive such group statistics that then exist outside the *Privacy Act*.

[3] We adopt here a definition of “de-identified” as meaning a process going beyond just removal of PII such as name and birthdate, but eliminating uniqueness of individuals’ data in a release.
[4] Ashwin Machanavajjhala, Daniel Kifer, Johannes Gehrke, and Muthuramakrishnan Venkitasubramaniam. "l-diversity: Privacy beyond k-anonymity." *ACM Transactions on Knowledge Discovery from Data (TKDD)* 1(1), 2007.

Attempting to regulate the group statistics themselves is likely to present a significant challenge. However, one measure that could *in theory* better equip consumers to mediate the sale and transfer of data derived from their personal information, is to require appropriate consent, even if such data is deemed not to be re-identifiable. This would not limit those who are happy to share their information with a provider in exchange for services, but it would empower those who do not wish to engage in such trade to refuse consent - albeit with the possible consequence of either having to pay for associated resources or not having access at all. Why would such an approach only work in theory? Because any such approach is going to be dependent on effective enforcement by a regulator, and be dependent on consent being obtained freely and fairly, which is regularly not the case. We discuss the issues around consent in more detail in Section 2.

Recommendation 3. The sharing of data, including data deemed to be no longer identifiable should require consent.

1.3. De-Identified Data

The first thing to realise about de-identified data is that it often isn't actually de-identified in any meaningful or technical sense. This is not a new phenomenon[5], with the fallacy of de-identification being discussed and known about 10 years ago[6]. This predates the introduction of the definition of de-identification into the *Privacy Act*.[7] As such, at the point of introduction the definition was not fit for purpose. This problem has been compounded by inexplicable interpretation by the OAIC which not only does not follow the legislation, but is clearly not in the spirit in which the legislation was introduced.[8]

While a new data release might not enable new re-identifications, it might aid *reconstruction* of sensitive attributes for already identified individuals in a privately-held dataset, causing harm nonetheless.[9]

[5] Office of the Victorian Information Commissioner, "De-identification and privacy: Considerations for the Victorian public sector", 2018. See (Nov 2020) https://ovic.vic.gov.au/resource/de-identification-and-privacy-considerations-for-the-victorian-public-sector/

[6] Paul Ohm, "Broken promises of privacy: Responding to the surprising failure of anonymization", *UCLA Law Review*, Vol. 57, 2010. See (Jan 2020) http://paulohm.com/classes/techpriv13/reading/wednesday/OhmBrokenPromisesofPrivacy.pdf

[7] House of Representations, *Privacy Amendment (Enhancing Privacy Protection) Bill 2012*. See https://www.aph.gov.au/Parliamentary_Business/Bills_Legislation/bd/bd1213a/13bd020 (Nov 2020)

[8] Office of the Australian Information Commissioner, "Publication of MBS/PBS data: Commissioner initiated investigation report", 2018. See (Nov 2020) https://www.oaic.gov.au/privacy/privacy-decisions/investigation-reports/mbspbs-data-publication/

[9] Simson Garfinkel, John M. Abowd, and Christian Martindale. "Understanding database reconstruction attacks on public data." *Communications of the ACM* **62**(3), (2019): 46-53.

Recommendation 4. Privacy protections should not rely on de-identification, as defined by stripping of PII or eliminating uniqueness of individuals' records in a data release.

1.3.1. Reasonably Identifiable

The notion of reasonable identifiably was introduced by the *Privacy Amendment (Enhancing Privacy Protections) Act 2012*, modifying the definition of personal information and introducing the definition of de-identification. The explanatory memorandum to that legislation explains how "reasonably identifiable" should be interpreted:

> *"The new definition will refer to an individual who is, 'reasonably identifiable'. Whether an individual can be identified or is reasonably identifiable depends on context and circumstances. While it may be technically possible for an agency or organisation to identify individuals from information it holds, for example, by linking the information with other information held by it, or another entity, it may be that it is not practically possible. For example, logistics or legislation may prevent such linkage. In these circumstances, individuals are not 'reasonably identifiable'. Whether an individual is reasonably identifiable from certain information requires a consideration of the cost, difficulty, practicality and likelihood that the information will be linked in such a way as to identify him or her."* [10]

This interpretation is problematic. Distinguishing between something that is technically possible and practically possible is highly subjective and fraught with danger when practicality rapidly evolves. The first problem this creates is that data released as "de-identified" may become identifiable through nothing more than greater access to more powerful computers. Since no mechanisms are specified for the recall of such data, the legislation has created a ticking time bomb of data that has been shared under the pretence of being de-identified on the day of its release, but that has a non-trivial likelihood of re-identifiability in the future.

The second problem is the example of what might prevent re-identification: "*logistics or legislation may prevent such linkage".* These are not considered barriers by the computer security community, because they impose an implicit honesty assumption on everyone, which we know to be unreasonable. We fit locks on doors, despite legislation prohibiting burglary and trespass. Logistical barriers are easily overestimated. The notion of cost, difficulty, practicality and likelihood are so broad and subjective as to render the notion of reasonably identifiable useless, or less charitably, a legal loophole.

Recommendation 5. Retire the qualifier of "reasonably" as threshold for identifiability.

[10] Privacy Amendment (Enhancing Privacy Protections) - Explanatory Memorandum. Available from: https://www.legislation.gov.au/Details/C2012B00077/Explanatory%20Memorandum/Text (Accessed 28th November 2020)

There are further inconsistencies in the definitions of personal information and de-identified information. Putting to one side the intent behind reasonably identifiable, its use in the definition of personal information is at least logically sound. The definition reads:

> *personal information means information or an opinion about an identified individual, or an individual who is reasonably identifiable:*

The definition effectively covers two states, one in which an individual has been identified, the first part of the proposition, and one in which an individual could be identified at some point in the future, subject to the notion of reasonableness. It is important to notice the distinction between the use of the word "identified" and "identifiable" in the two parts of the proposition.

The definition of de-identified does not follow the same pattern:

> *de-identified: personal information is de-identified if the information is no longer about an identifiable individual or an individual who is reasonably identifiable*.

In particular, the word "identifiable" is used for both parts of the proposition. As such, if read strictly, the first part of the proposition "an identifiable individual" supersedes the second part of the proposition, "an individual who is reasonably identifiable". The first part has no reasonableness constraint and as such should be applied on the basis of technical ability, not practical ability. It is unlikely this was the intent of the drafters.

In effect, this results in the *Privacy Act* having a definition of personal information that has a reasonableness constraint, and a definition of de-identified that effectively does not. It creates a third state for data to be in. One in which it is not personal information due to the reasonableness constraint, but is also not de-identified because it is technically possible for an individual to be identified. The data exists in limbo, devoid of any meaningful protection.

The impact of the poor definition on the application of the *Privacy Act* is relatively limited, due to the APP's being primarily written in terms of personal information, with only occasional reference to de-identified. However, due to the *Privacy Act*'s position as the standard for Australian privacy, the definition has found itself copied elsewhere, for example, in the Consumer Data Right legislation which includes the following definition:

> *"de-identifying CDR data, including so that it no longer relates to:*
> *(i) an identifiable person; or*
> *(ii) a person who is reasonably identifiable;"*[11]

Whilst it could be argued that this is a stronger definition of de-identified, which would benefit privacy, this interpretation depends entirely on the correct application of the definition by the regulator. History suggests that this is a charitable interpretation.

[11] Treasury Laws Amendment (Consumer Data Right) Act 2019. Available from https://www.legislation.gov.au/Details/C2019A00063 (Accessed 28th November 2020)

Recommendation 6. Consistency of related definitions should be treated as paramount, to avoid serious, unintended consequences.

1.3.2. Interpretation of De-identification in Relation to Re-identification

The best indicator of the interpretation of personal information and de-identification comes from the OAIC's report into the MBS/PBS data release[12]. In which the Department of Health made publicly available a "de-identified" dataset containing a 10% sample of the population's MBS/PBS records over a period of 30 years.

The OAIC report, for some inexplicable reason, appears to interpret the propositions in personal information and de-identified as conjunctive and evaluates not only whether an individual has been identified but whether that was reasonable. Such an interpretation is not only unconventional grammatically, acknowledging that conjunctive and disjunctive terms have a long history of ambiguity in legislation, but appears to be inconsistent with the spirit of the legislation. Once an individual has been identified, as was the case in the MBS/PBS release, the reasonableness of it is moot. It would be a considerable stretch to suggest that it is impractical to do something that has already occurred, yet that is the position that the OAIC took with regards to clearly and repeated demonstration of patient re-identification.

As such, there is little confidence that the OAIC would interpret de-identified in a strict manner. Fundamentally, the ambiguity created by the definitions, and their interpretations, needs to be resolved. Failure to do so will permit organisations to continue to claim data is de-identified under a spurious cover of being reasonable, allowing the sale and sharing of such data in compliance with the *Privacy Act*, yet still being a clear affront to individual privacy and exposing such organisations to serious reputational risk.

1.3.3. Practical Exploitation

The issue of organisations claiming data as de-identified when it is actually identifiable is not uncommon. Evidence of such behaviour can be found in privacy policies and terms. For example at the time of writing, Westfield's WiFi privacy policy states:[13]

> *"If you access or log-in to the Westfield Wi-Fi Service, and we hold other personally unidentifiable information that can be associated to you or the device on which you are accessing the Wi-Fi Service (including, but not limited to a device ID number (MAC address)), then that information may be linked with personal information we hold about you as set out in the Wi-Fi Privacy Terms or*

[12] Office of the Australian Information Commissioner, "Publication of MBS/PBS data: Commissioner initiated investigation report", 2018. See (Nov 2020) https://www.oaic.gov.au/privacy/privacy-decisions/investigation-reports/mbspbs-data-publication/
[13] Westfield's Wi-Fi Terms of Use and Privacy Terms, See (accessed: 28/11/2020), https://www.westfield.com.au/terms-and-conditions#wi-fi-terms-of-use-and-privacy-terms

> *the Scentre Group Privacy Policy, and will be treated in the same manner as the personal information to which it has been linked."*

The ability to potentially link "personally unidentifiable" information to identifiable information is a roundabout way of saying "re-identification". Any personally unidentifiable information an organisation holds could have been collected, stored, sold, or shared on the basis of it not being personal information. If re-identification were known to be feasible, then the data should not be classified as unidentifiable. Such terms result in the consumer consenting to re-identification of their data. They are not in a position to determine what unidentifiable data the organisation may hold, nor where it came from, and as such cannot possibly be expected to make a reasonable decision. Such clauses demonstrate the *Privacy Act*'s propagation of the fallacy of de-identification and the unreasonable consent conditions currently operating within the Australian market.

1.4. Importance of Verifiable Security Properties

As claimed by so many privacy policies and privacy impact assessments, the above Westfield Wi-Fi privacy policy excerpt asserts "unidentifiability" in the same way that the *Privacy Act* introduces "de-identified", without a definition that is falsifiable and verifiable. Such approaches to privacy protection - devoid of learnings from the scientific discipline of computer security - are inadequate. "De-identification" invites ad hoc solutions as witnessed in the 2018 Myki and 2016 MBS/PBS releases, along-side (obsolete but once) peer-reviewed proposals like *k*-anonymity[14]. Common to these approaches is a false self of intuition that a dataset has been rendered somehow anonymised without any consideration of what *security property* is being asserted, nor consideration of a *threat model* of the kinds of attacks that the "de-identification" process mitigates. Technical measures that do not provide a security property cannot be argued to provide *any measurable privacy*. Most well documented failures of technical measures can be traced back to a failure of threat model thinking.

Recommendation 7. Where technical measures are taken to protect privacy in data sharing or release, the *Privacy Act* should prefer those that provide security properties that assert protection against an identified threat model.

A common but flawed counter argument is that formal approaches to privacy, by contributing stronger protection (or measurable protection at all), must necessarily sacrifice benefits of data sharing or release. Cryptographic protocols are commonplace and protect against computationally-bounded adversaries when storing data on untrusted storage devices, transmitted data through untrusted networks, or processing data on untrusted cloud infrastructure. Differential privacy[15] has successfully been employed at scale by the U.S.

[14] Pierangela Samarati, Latanya Sweeney, "Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression", Harvard Data Privacy Lab, 1998.
[15] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. "Calibrating noise to sensitivity in private data analysis." In *Theory of Cryptography Conference*, 2006, pp. 265-284.

Census Bureau for all releases derived from the 2020 U.S. Census[16] and Google's COVID-19 Community Mobility Reports[17] among many organisations releasing data or sharing data with untrusted third parties. Just as the risk of dying from overdose does not unilaterally prevent prescription of pain medication by a skilled physician, practicality of technical privacy measures have been demonstrated repeatedly even while misuse can destroy public-good benefits of data sharing. The risk of misuse of formal approaches to privacy rarely justify adoption of ad-hoc techniques without security properties, just as potential for side effects of scientifically proven medicine does not warrant use of homeopathy.

Technical privacy measures such as cryptography and differential privacy adopt threat models that eschew "security through obscurity" - they do not require opaque, secret, implementations, and provide protection when code implementations are published, following so-called Kerckoff's Principle. An important benefit is transparency. Transparency of technical privacy measures, in the case of differential privacy, permits post-release improvement to accuracy of data analysis on shared data[18]. In all cases transparency engenders trust from data subjects and an increased social license to collect and share data. Transparency also leads to increased reliability of implementation through the "many eyes" phenomenon - Linus's Law.

Recommendation 8. Wherever possible, implementations of privacy enhancing systems should be made open.

1.4.1. Poor Guidance and Policies

The issues paper claims that to "...support robust de-identification practices and the management of re-identification risks, the OAIC and CSIRO's Data61 have released a non-binding de-identification decision-making framework." However, as has been discussed before[19] the guidance provided in the De-identification Decision-making Framework is poor and potentially misleading. The framework fails to adequately consider the challenges presented by longitudinal data; it fails to even acknowledge that the techniques described should not be applied on longitudinal data without significant pre-processing. The definition of *k*-anonymity provided in the framework is incorrect, crucially, failing to include the part of the definition requiring each individual to be represented in a single tuple.

[16] John M. Abowd, "The US Census Bureau adopts differential privacy." *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2018
[17] Google LLC, "Google COVID-19 Community Mobility Reports", https://www.google.com/covid19/mobility/ Accessed: 27/11/2020
[18] Raj Chetty and John N. Friedman, "A Practical Method to Reduce Privacy Loss When Disclosing Statistics Based on Small Samples", in *American Economic Review Papers and Proceedings*, 109, pp. 414-420, 2019.
[19] Chris Culnane and Kobi Leins. "Misconceptions in Privacy Protection and Regulation". LiC [Internet]. 2020 Apr.16 [cited 2020Nov.28];36(2):49-0. Available from: https://journals.latrobe.edu.au/index.php/law-in-context/article/view/110

1.4.2. An Inadequate Alternative: Five Safes

The Five Safes (also referred to as the Data Sharing Principles[20] by the ONDC) is a risk governance framework[21,22] for data sharing and release. While the framework is growing in popularity in the UK and Australia, and is presently positioned as the foundation for privacy in the *Data Available and Transparency Bill 2020*, it has undergone no substantive peer review by technical privacy experts or legal scholars. Our recent analysis of Five Safes[23,24] has found the framework to be inherently inadequate.

The framework's naming and adopted language encourages a false sense of security without substance. It is appropriate to consider a range of risks, from people accessing data, and the appropriateness of projects using data, to risks derived from the kind of outputs from collected sensitive data. While Five Safes draws the data holder's attention to various dimensions of *risk*, there is no guarantee that these dimensions are adequate. Indeed where the original framework comprised four safes[25] until 2007 where "Safe Data" was post hoc added, a recent ACS working group on data sharing has considered adding "Safe Organisations", "Safe Outcomes", "Safe Lifecycle"[26], while Data Republic has labelled five safes as "not enough" and proposed the addition of "Audit" and "Legal"[27].

While the governance framework makes no attempt to prefer rational definitions for privacy protection, security properties, or indeed privacy by design. Five Safes has the potential to encourage trading off dimensions of risk for one another, where in many cases, appropriate governance would demand that all dimensions be appropriate - "defence in depth" instead of "perimeter defence". Merely adopting the Five Safes framework, or fulfilling compliance with a "Five Safes assessment", do not necessarily render data sharing or release safer.

Recommendation 9. Government guidance should not recommend Five Safes in its current form.

---

[20] Office of the National Data Commissioner, "Data Sharing Principles", March 2019. See https://www.pmc.gov.au/resource-centre/public-data/data-sharing-principles (January 2020)
[21] Tanvi Desai, Felix Ritchie, and Richard Welpton. "Five Safes: designing data access for research", University of the West of England Bristol, *Economics Working Paper Series* 1601, p5. See https://www2.uwe.ac.uk/faculties/BBS/Documents/1601.pdf (January 2020)
[22] Australian Bureau of Statistics, "Managing the Risk of Disclosure: The Five Safes Framework", *1160.0 – ABS Confidentiality Series*, August 2017. See (accessed Jan 2020) https://www.abs.gov.au/ausstats/abs@.nsf/Latestproducts/1160.0Main%20Features4Aug%202017
[23] Chris Culnane, Benjamin I. P. Rubinstein, and David Watts. "Not fit for Purpose: A critical analysis of the 'Five Safes'". arXiv:2011.02142 [cs.CR]. Nov. 2020. https://arxiv.org/pdf/2011.02142
[24] Chris Culnane and Ben Rubinstein, "Consultation Response 'Data Availability and Transparency Bill'", Nov 6, 2020. See https://www.datacommissioner.gov.au/exposure-draft/submissions (Nov 2020)
[25] Wikipedia contributors. "Five safes" — Wikipedia, the free encyclopedia, 2019. [Online; accessed 7-September-2019].
[26] Ian Oppermann. "Data sharing in an Australian context", 2018. Speech at FutureData 2018. https: //www.futuredata.events/resources/ [Accessed 2019-09-08.]
[27] Data Republic. "Why five 'safes' aren't enough for inter-organizational data exchanges: Introducing data republic's seven governance controls", 2019. https://www.datarepublic.com/blog/ five-safes-inter-organizational-data-exchange, [Accessed 2019-09-08.]

1.4.3. A Modern Definition

The current definition of personal information and de-identified is not fit for purpose. It introduces inconsistencies and ambiguities that cause it to fail to protect individual privacy. Common de-identification techniques do not guarantee that an individual is not identifiable. It is of note that the National Statement on Ethical Conduct in Human Research shuns such terms[28]

> *"The National Statement does not use the terms 'identifiable', 'potentially identifiable', 're-identifiable', 'non-identifiable' or 'de-identified' as descriptive categories for data or information due to ambiguities in their meanings"*

The *Privacy Act* should follow the same approach and avoid such terms.

Data that has been de-identified could potentially remain personal information. As such, the burden of proof that the data is protected should be on the releasing or sharing party to prove that the methods the privacy protection mechanisms used are robust and adequate. Information about the methodology should be made publicly available. Claims that doing so would undermine security is in itself evidence that the methods are inadequate privacy protections.

Recommendation 10. De-identified data should be considered to remain personal information until sufficient proof is provided that demonstrates adequate privacy protection**.**

In addition to improving the definitions within the *Privacy Act* improved enforcement needs to occur as well as new rights to take direct action. The necessity for the right to direct action has been underlined by incidents like the MBS/PBS release and the subsequent OAIC investigation. The failure to adequately understand the risk presented by releasing 30 years of medical records for over 2.4 million Australians indicates a significant capability gap within the regulator. If a regulator is not equipped to evaluate mitigations in its purview, they should not act as sole arbitrator.

The OAIC should be funded sufficiently so it may build up internal technical capability sufficient to understand, provide guidance, and enforce the privacy protections offered to individuals within the Act. However, a backstop should be provided so that an individual who believes the OAIC is not capable of acting on a perceived breach, may pursue that matter independently. Whether that is achieved through stronger direct action methods or a privacy tort we leave up to legal scholars to decide.

Recommendation 11. Stronger rights for direct action should be afforded to individuals.

[28] NHMRC, ARC, and Universities Australia. "National Statement on Ethical Conduct in Human Research 2007 (updated 2018)", 2018. See (Accessed: 2020-11-28) https://www.nhmrc.gov.au/about-us/publications/national-statement-ethical-conduct-human-research-2007-updated-2018

1.5. Re-Identification Amendment

Following the 2016 re-identification of the MBS/PBS dataset the government attempted to introduce a criminal offence for performing re-identification of government-held datasets that were released as de-identified through the *Privacy Amendment (Re-identification Offence) Bill 2016*. The government chose to take the extremely unusual step of retroactively introducing the amendment. As such, in keeping with precedent on retroactive legislation, the re-identification amendment became law on the 29th September 2016 after the Attorney General George Brandis announced it in a media release[29]. As such, organisations covered by the *Privacy Act* were obliged to comply from that date onwards. The amendment went further than just criminalising the act of re-identification, by making it "...an offence to counsel, procure, facilitate, or encourage anyone to do this, and to publish or communicate any re-identified dataset"[30].

Including offences to facilitate, encourage, and counsel, had a chilling effect on the ability to discuss, educate and research de-identification, and by extension re-identification. Whilst it may have been a coincidence that the amendment was retroactively introduced on the evening prior to the date we had agreed with the Department of Health that we would publish our findings into their release of the MBS/PBS re-identification, it clearly had a chilling effect on what we said the following day.

Such legislation serves little purpose in actually preventing re-identification. First, it only covered government-held data, yet from a privacy perspective, vastly more data is held and exchanged by businesses in a claimed de-identified state. Second, it effectively provided a blanket protection for inadequate government de-identification, since the threshold for it to apply was merely that the dataset had been released as "de-identified" - an arbitrary self-assessment about identifiability - not that it had been meaningfully de-identified. This would have created something of an emperor's new clothes scenario. Provided parties pretended the data was de-identified, all would be OK (the privacy of ongoing damage to data subjects might not be so OK). Failure to maintain the pretence would be punishable by up to two years in jail.

Third, the legislation did not consider that the act of re-identification produces no public artefacts, and as such, detecting an incidence of re-identification would be next to impossible without either a whistleblower or someone self-reporting. As such, the group most likely to fall under the legislation were security researchers who would be obliged by public interest to responsibly disclose such vulnerabilities and discoveries.

[29] https://web.archive.org/web/20160930092155/https://www.attorneygeneral.gov.au/Mediareleases/Pages/2016/ThirdQuarter/Amendment-to-the-Privacy-Act-to-further-protect-de-identified-data.aspx
[30] *Ibid.*

Fourth, as was mentioned in the Australian Bankers Association submission[31] to the Senate Legal and Constitutional Affairs committee inquiry into the proposed amendment, re-identification can take place accidently.

Possibly due to the many problems with the proposed legislation it failed to pass, with it not even being brought for a vote in the Senate where it was introduced. However, it remained as part of government business, occasionally reappearing in the order of business in the Senate. Eventually the bill expired on the 1st of July 2019.

Due to the unprecedented nature of its retroactive introduction, the Bill's expiration created yet more ambiguity. Did the expiration of the Bill cause the retroactivity to cease to be valid? Is the original statement still in effect? Could the government re-introduce the Bill and enforce it retroactively? We are unaware of any statement unwinding or clarifying the original position to have been made by the Attorney General. All records of the original media release appear to have been expunged from the Attorney General's website at time of writing. As early as 2018 - notably before the Bill had expired - the original media release announcing the retroactivity had been removed from the website, requiring someone to browse the National Library of Australia archives to find it[32].

There are undoubtedly organisations and individuals who remain under the impression that the re-identification amendment became law, when it did not. Even if the expiration of the proposed legislation did cause the retroactivity to cease, its initial retroactive introduction had the effect of causing organisations to comply with the re-identification amendment as if it was law for a period of nearly 3 years. In effect, the Attorney General was able to introduce a law and require compliance outside of parliament.

Far from providing protection, or removing ambiguity, the re-identification amendment managed to add to the ambiguity of an already highly ambiguous concept and definition. The unprecedented nature of the introduction has caused unnecessary confusion. Crucially, due the *Privacy Act*'s use of the poorly defined term "reasonably identifiable", and the corresponding evaluation of the term considering legislative constraints, it is theoretically possible that datasets were released during the period before expiration that evaluated the reasonableness of identifiability on the assumption that the legislation existed.

The government needs to clarify its position with regards to the retroactivity of the original legislation, and should review whether the assumed existence of the legislation impacted on the risk evaluation of any releases between 2016 and 2019. Furthermore, the debacle of the re-identification amendment should stand as a lesson against retroactive introduction of

[31] Australian Bankers' Association, Submission to the inquiry into the Privacy Amendment (Re-identification Offence) Bill 2016. See (accessed: 29 Nov 2020) https://www.aph.gov.au/Parliamentary_Business/Committees/Senate/Legal_and_Constitutional_Affairs/PrivacyReidentification/Submissions

[32]https://webarchive.nla.gov.au/awa/20161101002032/http://pandora.nla.gov.au/pan/21248/20161101-1107/www.attorneygeneral.gov.au/Mediareleases/Pages/2016/2016MediaReleases.html

potentially controversial and unsound legislation, particularly where it covers contentious and ambiguous topics like de-identification.

Recommendation 12. The government should clarify the retroactive state of the now expired *Privacy Amendment (Re-identification Offence) Bill 2016.*

Recommendation 13. A review of government data releases between 29th September 2016 and 1st July 2019 should be undertaken to evaluate if any assumed operation of the *Privacy Amendment (Re-identification Offence) Bill 2016* in their evaluation of "reasonably identifiable".

2. Consent

Too often consent is viewed as an inhibitor to data sharing, when in fact, in a well calibrated privacy protection mechanism, consent is the enabler of sharing. By default data should be protected and its use restricted to the purpose for which it was provided. Such collection must be transparent and conducted with the consent of the individual. Such a principle is becoming ever more important with the advancement of technology and the wider distribution of IoT devices, which are capable of collecting vast quantities of data related to individuals in a covert manner. As we have already noted, such consent should be extended to cover data derived from the original collected data as well.

However, consent is not a silver bullet, it is only effective if it is freely given and not subject to methods such as dark patterns[33]. It should also not be underestimated that the power consent has. As such not only is obtaining consent important, but the regulation of methods of obtaining consent should also be a focus of the *Privacy Act* review. Failure to fully treat consent negatively impacts individual privacy, through organisations obtaining consent through subterfuge, coercion or manipulation. Such consent could be far reaching, see for example the Westfield WiFI example given above, and lead to an overall reduction in individual privacy protection.

2.1. Regulation of Consent Provision

The regulation of the provision is hugely challenging and no best practice has yet been established. Looking at the situation in the EU, with regards to Privacy and Electronic Communications Regulations (PECR)[34] and the GDPR, the mechanisms defined there are

[33] Midas Nouwens, Ilaria Liccardi, Michael Veale, David Karger, and Lalana Kagal. 2020. "Dark Patterns after the GDPR: Scraping Consent Pop-ups and Demonstrating their Influence". In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. DOI:https://doi.org/10.1145/3313831.3376321

[34] Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications). Available from https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32002L0058 (Accessed 28th November 2020)

currently not working effectively. Not only is there a serious problem of consent fatigue[35] but enforcement of clear breaches of the consent regulations, for example, through consent bundling or default consent options, neither of which is allowed, have so far been ineffective. Some of the largest online providers are currently in breach of regulations. That may change in the future, with a recent ruling by the The Court of Justice of the European Union (CJEU) re-iterating the constraints defined in the GDPR[36]. Time will tell whether the ruling leads to stronger enforcement by regulators.

Whilst we would encourage following the EU's lead in framing consent we would recommend a more proactive stance be taken on regulation of the provision of consent. There should be a safe by default standard enforced, that should a consumer do nothing, their privacy is protected and they have not accidently or implicitly provided consent to the collection or use of their data. Note, this is arguably the intention in the EU, however, the lack of technical regulation to accompany it has created the potential for ambiguity and interpretation.

Recommendation 14. Follow the EU GDPR lead in framing the provision of consent, but accompany it with further technical regulations covering the obtaining of such consent.

To counter consent fatigue we would recommend the provision of an automated consent mechanism standard, which would allow a consumer to set their privacy and consent provisions once in their browser or device and these would be sent as the default options to the service provider in a standardised format. Any additional consent requirements needed by the service provider would need to be asked for explicitly outside of the default flow of access. In effect the easiest and quickest option would be sticking to the default provisions. Australia should not be afraid to lead in this area as its internal market lends itself to technical innovation due to its size and make-up and that should be leveraged not just by tech firms for experimentation, but by regulators to innovate in regulation and enforcement as well.

Recommendation 15. Require the adoption of an automated consent provision framework that allows consumers to set defaults on a browser or device level.

2.2. Revocation

Whenever consent is provided it must be accompanied by a provision and method for the subsequent withdrawal of consent. Currently such mechanisms are clunky and differ between providers. In line with the previous recommendation, a standard method and framework for the

[35] Guidelines on Consent under Regulation 2016/679 (wp259rev.01). Available from https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=623051 (Accessed 28th November 2020)

[36] Judgment of the Court (Second Chamber) of 11 November 2020 Orange Romania SA v Autoritatea Națională de Supraveghere a Prelucrării Datelor cu Caracter Personal (ANSPDCP), C-61/19 - Orange Romania. Available from http://curia.europa.eu/juris/liste.jsf?language=en&num=C-61/19 (Accessed 28th November 2020)

revocation of consent should also be introduced. Furthermore, such revocations should propagate to those with whom data has been shared with, including, where appropriate, data that has been derived and was no longer deemed identifiable. As such, a new burden should be introduced on the party that collected the data to be able to track how such data is shared, with the ability to revoke consent throughout the entire chain. The consumer should not be required to undertake a treasure hunt for their data in order to enforce a revocation.

Whilst organisations may object to this additional burden, it should be noted that there is no requirement for them to share data. They are undertaking such sharing in order to generate profit, and as such, should expect there to be associated costs with undertaking such actions.

Recommendation 16. A collecting party must maintain records of data sharing, including derived data deemed no longer identifiable, so that consent withdrawal can propagate to all parties in possession of the data or derived data. A consumer need only notify the original collecting party to have consent withdrawn throughout the chain.

The requirement for the collecting party to maintain such records should not negate the rights of the individual to revoke consent on a piecemeal basis. The records held by the collecting party of who has received the data downstream should be accessible to the individual to facilitate the data subject in revoking consent directly with those downstream parties as well. As such, a consumer should be empowered to revoke consent from the use of their data or their derived data from any party downstream should they wish to.

Recommendation 17. Consumers must be able to access the record of who their data has been shared with, and be able to withdraw consent for the usage of their data, or their derived data, from those downstream parties directly.

A standard for the structure and sharing of such information should be specified in associated regulations to facilitate automated processing of the access, monitoring, and revocation by the consumer.

2.3. Transfer vs. Sharing

The *Privacy Act* does not currently provide provisions for when data is transferred, through for example, acquisition of the business that collected it. It is wrong to assume that the consent provided for its collection and use should by default transfer with the data. A consumer will have made such a decision based on a number of factors, including trust in the organisation, history in protecting data, and what other data that organisation may hold. If transfers can occur without restriction, such a state of affairs encourages trade in data through business acquisition. This is particularly dangerous in the digital platforms environment, where a few oligopolistic companies are able to acquire businesses just for their data.

Transfers of ownership should by default reset consent, requiring the new owner to obtain fresh consent from the individual. Where such consent cannot be obtained the data must be deleted.

Recommendation 18. Transfer of ownership of data resets consent, requiring the new owners to reobtain consent from the individual. By default, where no consent can be obtained or the consumer does not respond, consent is assumed to have expired and the data must be deleted.

2.4. End the Small Business and Political Party Exemptions

Currently, with a small number of exceptions, businesses with annual turnover of $3 million or less are exempt from the *Privacy Act*. Similarly political activity carried out by representatives of registered political parties are exempt from the *Privacy Act*. Political parties have access to unprecedented online tracking of voters; there is no valid reason that they should be exempt from proper management of the privacy of Australia's citizens - indeed they should lead by example. The privacy rights of individuals whose data is collected by small businesses are no different to the rights of individuals represented in medium and large business collections.

Recommendation 19. End the *Privacy Act* exemptions to small businesses and political parties.

3. Deletion

The issue of deletion of data is closely related to that of the provision of consent. Deletion should be the default action where consent for the collection, storage, or use of data is revoked. Crucially, the use of the phrase "deletion or de-identification" should cease. The two are not equivalent and by allowing de-identification as an option it prevents meaningful consent withdrawal, and creates a perpetual transfer between the consumer and organisation. Some residual value will reside within the de-identified data and as such, some loss will persist for the consumer.

3.1. Deletion vs. De-identification

The use of "deletion or de-identification" as an approach has found its ways into the C*onsumer Data Right*, potentially perpetuating the problems it causes outside of the *Privacy Act*. Opposition was made to a blanket deletion requirement in the CDR, and one would assume the same opposition would oppose its introduction to the *Privacy Act*. That opposition was based on the technical difficulty of deleting data from back-ups and archives. However, such an argument is completely fallacious, if it is difficult to delete data it is considerably harder to de-identify it. Deletion only requires destruction, as such, overwriting, or trimming of data is all that is required. De-identification could involve fundamental changes to fields and the structure of the data itself, through methods like cryptographic mask or encryption. Such changes are not possible in many back-up mediums.

As such, the argument was not sound to begin with, and appears to be motivated to exploit the ambiguity around de-identification to allow the recipient of the data to retain and extract value

from it even after consent has expired. This falsehood needs explicit redress in the *Privacy Act* and for the correction to flow on to the CDR as well.

Recommendation 20. Remove references to “deletion or de-identification”, replacing with only “deletion”.

The debate around a right to deletion should not focus on deletion as a separate action to the revocation of consent. By focusing on consent, with deletion as a consequence of a withdrawal, a more nuanced approach can be derived, avoiding the challenges associated with determining whether a deletion request is reasonable, since it becomes a decision as to whether consent can be withdrawn.

Recommendation 21. Withdrawal of consent should lead to deletion by default, with decisions on whether consent can be withdrawn used as a threshold for deletion.

4. Transparency

A number of the above recommendations aim to improve transparency so as to empower the consumer or citizen to more effectively enforce their rights. The current data economy takes place almost entirely behind opaque doors. There is no oversight of the nature or scale of most sharing. A consumer is not aware of how their data has been shared, in what form, with what safeguards and under what premise, i.e. consent or de-identification. This lack of transparency almost entirely negates the powers of the consumer to monitor and enforce their privacy rights beyond the first collecting party.

As such, considerable effort must be placed to improve the transparency of the data sharing market. In particular, the methods of privacy protection being used to legitimise such sharing need to be open to public scrutiny, as well as accounts of where data has come from and what data organisations hold. It should be noted that when discussing what data an organisation holds, this does not mean publishing the data itself, but rather the metadata:

- A list of fields contained within the dataset;
- The number of individuals contained within the dataset;
- Where the dataset came from, i.e. the party one iteration back;
- The original collecting party, i.e. the original source; and
- What privacy protections methods were used.

The aim of such transparency is not only to facilitate the stronger consent framework discussed above, but also to increase awareness of what data is being fed into the decision-making processes by organisations in receipt of data.

In effect a data provenance scheme must be developed that provides public accountability. Such a scheme will require organisations that hold data to be able to trace back where their data flowed from, and prove that they have the necessary consent. Additionally, individuals will be able to use the public artefacts to better track and monitor their data and where necessary revoke consent.

Recommendation 22. Establish a data provenance scheme with associated public artefacts. Such a scheme would require the publishing of meta-data about datasets held, providing reference back to the original collecting party to facilitate granular consent decisions by the individual.

5. Privacy Environment

The current privacy environment in Australia is weighted heavily towards the exploitation of an individual's data, with a steady stream of legislation and regulation that have either diluted individual privacy protections or facilitate mass sharing of data, for example, data sharing legislation and the consumer data right. Legislated privacy protections have not kept apace. This seems to largely be driven by an unsubstantiated argument about productivity, influenced by the Productivity Commission's inquiry.[37] Compounded by poor privacy legislation and ineffective enforcement, it has resulted in a data environment that is adversarial towards the individual consumer, as has been discussed in the ACCC Digital Platforms inquiry.[38]

In a rush to share data and exploit it justified by phrases such as "the new oil", the mistakes of past exploitation of resources have not been heeded. This is not solely a problem Australia faces, many countries are grappling with the power and exploitative practices that large data driven digital platforms are utilising. However, where a government is embarking on widespread legislation to encourage greater sharing of data, it presents a particular challenge to simultaneously strengthen privacy protections.

The consequence is that the true productivity costs associated with privacy protection are overlooked and externalised, and rather than the sharing of data generating productivity benefits for society, they are generating vast profits for a few large organisations.

If an individual wishes to effectively protect their privacy today they will incur significant resource costs in doing so, in both time and money. For example, they will need to purchase products or

[37] Productivity Commission, "Data Availability and Use", Inquiry Report, No. 82, 31 Mar 2017. See https://www.pc.gov.au/inquiries/completed/data-access/report [Accessed: 29 Nov 2020]
[38] ACCC, "Digital Platforms Inquiry: Final Report", June 2019. https://www.accc.gov.au/publications/digital-platforms-inquiry-final-report [Accessed: 29 Nov 2020]

services that protect them from surveillance capitalism[39], like a VPN. They will need to spend time to counter the adversarial privacy invading practices that exist online, for example, by clearing cookies per session, using containers, blocking scripts, etc. All of which will require expending time learning how to perform such actions and the time to apply them. The opportunity cost is varied: web histories that aid recall, rapid searching, disabling WiFi or Bluetooth in public, not using phone based digital wallets, avoiding some electronic transactions in favour of cash. All such behaviours have an impact on that individual's productivity.

Data has the potential to benefit society and individuals, but it also has the potential to facilitate exploitation and manipulation on an unprecedented scale. Implementing a robust privacy protection framework is essential in delivering the former and avoiding the latter. Such a framework will allow individuals to safely engage in productivity enhancing services, whilst rejecting those primarily designed to profit at the individual's expense. As such, a strong privacy framework is not a hindrance to productivity enhancing data sharing, it is foundational to its success.

[39] Zuboff, Shoshana. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: PublicAffairs, 2019.