Submission to the ACCC Consultation on the Exposure Draft of CDR draft rules (banking) May 2019

Dr Chris Culnane, A/Prof Benjamin Rubinstein, Cynthia Sear, A/Prof Vanessa Teague
Contact Email: [redacted]

This submission was prepared by the authors in a personal capacity. The opinions expressed in this submission are their own and do not reflect the official position of The University of Melbourne.

This submission will focus on the privacy concerns related to Consumer Data Requests made both by the consumer and on behalf of the consumer. Our analysis focuses on computer science issues, but we broadly agree with the legal issues raised by Anna Johnson and others.[1]

Introduction

We support the principle of consumer access to their data, and their right to access and use their data as they see fit. However, in the absence of strong privacy protection legislation, the greater access rights afforded to consumers have the potential to be misused by predatory data collectors, to identify consumers and to adversely impact a consumer's ability to access types of financial products and services. To strengthen the proposed draft rules against these undesirable and potentially dangerous outcomes we therefore recommend the following four changes:

1. Greater restrictions on unaccredited organisations regarding requests for consumer data.
2. Addition of a consumer right to deletion instead of de-identification in Part 7.
3. Removal of references to de-identification in Division 7.10.
4. Clarification and tightening of definition as regards the prohibited use of data.

The rationale for each of these changes is outlined below. Please feel free to contact us should you require any further explanation.

Recommendation One: Relevant to Part 3 of the "Draft Rules": Consumer Data Requests made by CDR Consumers

Stronger protections are needed for unaccredited companies who may attempt to solicit CDR data indirectly. By way of an example, an unaccredited organisation could request consumer data by providing instructions on how the consumer themselves can request such data under Part 3, before requiring the consumer to upload the data to the unaccredited organisation. As such, the unaccredited organisation could acquire consumer data without being covered by the consent and privacy protections provided by the CDR.

It is clearly desirable to give consumers access to their data, but it cannot be expected that consumers will immediately become sophisticated data guardians. Rather than limiting consumer access it would be more desirable to impose restrictions on the collection and use of such data outside of the CDR framework. As such, consumers would be empowered to access, review and utilise their data within a safe environment away from potentially predatory data collection businesses.

[1] https://www.salingerprivacy.com.au/2019/04/27/ai-ethics/

Recommendation Two: Related to Part 7 Privacy Protection - 7.8 Rules relating to privacy safeguard 12

At some point, a draft option to allow a customer to demand the deletion of their data has been watered down to an option to demand that it be "deleted or de-identified." This is a much weaker protection, because de-identification of detailed unit-record level data about individuals doesn't work. Providing an option to de-identify data instead of deleting it presents a serious risk of privacy harm to any consumers utilising the CDR.

The CDR rules recommend Data61's De-identification decision-making (DDM) framework, but this is not a scientifically tested or rigorously reviewed standard for de-identification. It should not be the mandated approach in legislation or associated statutory rules. It contains a number of weaknesses and problems which could mean data could be easily re-identified despite being supposedly de-identified in accordance with the framework.

More fundamentally, de-identification—if it is meaningful at all—is inherently based on the notion of being lost in a crowd. (Alternative privacy-protection frameworks such as Differential Privacy can apply to single records.) De-identification methods require multiple individuals to be in the dataset to work, if they work at all. In the situation where the consumer revokes consent and requests deletion, the dataset to be de-identified will consist of just that single consumer's data. If that crowd, as in this case, consisted of only one person, then it is impossible to be lost in it.

The misunderstanding could stem from the DDM framework's incorrect definition of k-anonymity, the method on which many de-identification techniques are based. The DDM framework incorrectly defines k-anonymity in terms of records instead of correctly defining it in terms of individuals. To correctly apply the technique would require the pre-processing of the dataset to render all records related to one individual into a single tuple or row. If this had been correctly specified in the DDM framework it would be immediately obvious that de-identification cannot be performed, since the dataset would contain only a single row and therefore k can never grow to being more than 1. Finally, k-anonymity is not regarded as best practice by privacy experts.[2]

A third factor to acknowledge is that even if aggressive aggregation were performed on the data to provide only summary statistics for an individual, it is difficult to see how this could be considered as de-identified. Those statistics will again refer to a single individual, and even if they were summed or averaged with previous summary statistics from other individuals, the change in those average at that point in time could still be tracked to an individual. In fact, such tracking is inherently possible by the requirement in 7.8.2.c to make a record of evidence of de-identification.

It is not clear what the intention is behind the inclusion of de-identification as an alternative to deletion. If there is a legitimate use-case in mind it should be explicitly stated. Weakening of privacy protections should be the exception, not the norm.

A fourth factor to recognise is that 7.8 is currently ambiguous as regards the rules that will apply to de-identified data. Would such de-identified data remain constrained by the prohibited use clause, or would the de-identified data become a free for all, outside of any protections, as it apparently is free from the constraints of the Privacy Act? Given the repeated failures of de-identification,[3] and the repeated warnings about its fallacy,[4] it is disappointing to see it appear in frameworks being written in 2019 that are intended to enhance consumer rights and be of benefit

[2] Ashwin Machanavajjhala, Johannes Gehrke, Daniel Kifer, and Muthuramakrishnan Venkitasubramaniam. "ℓ-diversity: Privacy beyond k-anonymity." In *22nd International Conference on Data Engineering (ICDE'06)*, pp. 24-24. IEEE, 2006.

[3] Chris Culnane, Benjamin I. P. Rubinstein, Vanessa Teague. "Health Data in an Open World," 2017. https://arxiv.org/abs/1712.05627

[4] Broken Promises of Privacy:Responding to the Surprising Failure of Anonymization - Paul Ohm - https://www.uclalawreview.org/pdf/57-6-3.pdf

to consumers. There is no consumer advantage to permitting the de-identification of data, and as such, 7.8 should be strengthened to require the deletion of data.

As a final consideration, it should be noted that any push back from industry to the requirement to deletion should be tempered by the fact that the CDR involves a contract between the consumer and the CDR recipient. Equivalent contracts in business frequently contain clauses requiring the deletion of data provided under such contracts at their termination, often going even further to include any instance of the data in back-ups. Businesses are perfectly capable of complying with such requirements, it is an anachronism that the same protections and capabilities are not afforded to consumers today.

In short, deletion should mean deletion.

Exactly this issue forced one of the amendments made to the My Health Record Act as a result of public pressure.[5] Previously, deletion of My Health Records didn't really mean deletion. Clearly, ordinary consumers (and a majority of federal MPs) decided that it should.

Recommendation Three: 7.10 Rules relating to privacy safeguard 13 — steps to be taken when responding to correction request

References to de-identification should be removed from this rule, permitting correction or deletion. Aside from the concerns regarding the effectiveness of de-identification, it seems odd to permit the de-identification of information which is incorrect. Furthermore, the ability to *add to* instead of correct also seems counter to consumer benefit. If incorrect information is held, simply adding correct data in addition to incorrect data is not sufficient. This is particularly relevant where data will be utilised for decision making or pricing. The mere presence of some values could adversely impact on the rates or products a consumer is given access to. If such data is incorrect it should be removed or replaced, not left in the record to persist in causing harm. Whilst a statement detailing a correction request would make it possible to discern the accurate data, unless there is a legal requirement for the entity processing the data to respect such statements it is of little value in protecting the consumer.

Correction should mean correction. Otherwise it is inevitable that the uncorrected persistent version may be used.

Recommendation Four: Clarification of Prohibited Use

The definition of prohibited use would appear to prevent the usage of Machine Learning and AI to build statistical models from received CDR data. We would welcome such a restriction as such models can be used to derive secondary value from consumer data without having obtained consent for that usage. If that is the intention, we recommend the addition of an explicit statement in the rules to avoid any doubt.

Second, we recommend that the definition of prohibited use be tightened. It currently restricts building profiles or insights about any person, whereas most models will build profiles or insights about groups or classes of people, where class is a particular combination of attributes. This could lead to a loophole being exploited to allow a CDR recipient to build broad models about classes of people, but not a person. This could then be used to deny products and services to individual members of a particular class or group. As such, we recommend a stipulation prohibiting the building of profiles and insights for groups of persons.

We note here that it is unclear how the prohibited use restriction could be enforced. The building of any profile or insight model could be conducted entirely in private, without any public

[5] My Health Records Amendment (Strengthening Privacy) Bill 2018 https://www.aph.gov.au/Parliamentary_Business/Bills_LEGislation/Bills_Search_Results/Result?bId=r6169

artefacts being created. Once built, the original data would no longer be required, since it would have been codified into the model. As such, unless volunteered, establishing the provenance of the data that was used to build such a model would be extremely difficult, if not impossible. As a result there would be a very low risk that a malicious recipient who built such a model would either be caught, or could be pursued for breaching the prohibited use restriction, unless the act was divulged by a whistle-blower. For this reason we recommend that the ACCC consider ways to raise the salience of such wrong-doing. In making it clear and widely known what appropriate and lawful use of data is and is not, law-abiding citizens and workers will be empowered to speak out in the advent of unlawful uses.