veda
applied intelligence

18 July 2012

Ms Julie Dennett
Committee Secretary
Legal and Constitutional Affairs Committee
Parliament House
CANBERRA ACT 2600

Dear Julie,

Veda is an information economy company, best known as Australia's leading provider of consumer credit reports and related business risk information and analysis.

Veda welcomes the long-awaited introduction of comprehensive credit reporting legislation and urges support for its passage. While there are areas of the *Privacy Amendment (Enhancing Privacy Protection) Bill 2012* (Bill) that may be improved, overall it is a very positive step in giving Australia a robust and privacy enhancing credit reporting regime.

Veda notes the submission by industry association ARCA and supports further consideration of the four priority issues nominated.

The purpose of this submission by Veda is about the proposal to regulate de-identified data, where, for research purposes, data is stripped of information which would otherwise allow the identification of the person who is the data source.

Importantly, no recommendation was made by the Australian Law Reform Commission for regulation of de-identified data in the Privacy Act and the introduction of these provisions appear to be without precedent in modern economies.

De-identified data is critical for creating data series, accurate statistical modelling and developing insights into historic trends. It helps ensure the accuracy of credit risk models and the insights it can contribute are also provided to key financial pillars such as the Reserve Bank.

Veda is also concerned at the proposal to only retain current plus two previous addresses on a credit file. This will have substantial impact on the accuracy of 2.4 million Australian credit files and diminish the ability of credit reporting bureaus to accurately match a credit inquiry to a file.

In an era of increasing mobility and transience, there is a strong argument to be allowed to collect current address plus two previous *or all addresses over the previous five years*, whichever is the greater. Attached is a copy of a letter sent to the Attorney General's department on this critical issue.

My team and I would welcome an opportunity to meet and discuss any questions or points of clarification that may arise in the Committee's consideration of these issues.

Yours sincerely

NERIDA CAESAR

Recommendations on de-identified data

1. Retain the *purpose* for which de-identified data may be used, without prescribing rules

Amend 20M to delete all parts after 20M (2) (a) and the corresponding complaint related provision in section 23A (4). This keeps the original Australian Law Reform Commission (ALRC) recommendation on use and disclosure of credit information and other types of personal information in the credit reporting system. It also retains the integrity of the regulatory structure that is established to protect the privacy of individuals by focusing on regulating 'personal information' as defined.

2. Productivity Commission referral

If the Committee supports introducing rules on use, collection or disclosure of de-identified information, then any such rules should only be made after consideration and advice from the Productivity Commission on the economic value of de-identified data and the potential implications of depriving the economy of the insight currently being gained from the de-identified use of information derived from the credit reporting system and the information exchanges in that system.

3. Appropriate penalties for misuse of personal information (of any kind) to be retained

Address concerns about subsequent re-personalisation with substantial penalty provisions as apply elsewhere in the Bill. This addresses the fact that once information identifies an individual it comes within the scope of privacy regulation by virtue of the fact that the information is personal information as defined.

Recommendation on two addresses

1. **Insert an additional provision allowing for credit reports to include, for the purpose of record management, the greater of:**

(i) The current plus last two addresses;
(ii) All addresses over the previous five years.

KEY FEATURES OF DE-IDENTIFIED DATA

- De-identified data is stripped of information needed to identify individuals;
- It is used in the aggregate as part of statistical modelling;
- Depersonalised data models are used in fraud and identity verification services, as well as for assessing improving or deteriorating credit risk.
- De-identified credit reporting data is not regulated elsewhere in the OECD; debate on de-identified data has focussed on health research;
- Re-personalisation of de-identified data has occurred in the United States, where other, large scale, public record data sources are easily available.

INTRODUCTION

This part of the submission provides a history of uses of de-identified information. These uses go to the core of the value being provided to the Australian economy, the political process and how the Australian economy considers and treats matters of credit risk and economic hardship.

In regulating de-identified data Australia would be the only jurisdiction in the OECD to take this approach. Veda submits this goes against and undermines the role this data plays in the economy, the regulatory structure and that the perceived risk associated with this type of information does not justify the level of regulatory intervention being proposed.

The Economy - Unemployment and early indicators of shifts in the Australian Economy

In early 2008, Veda received a phone call from Treasury. At that time, unemployment hovered at around four per cent, GDP was strong and mortgage defaults were down to 0.2 per cent.

But elsewhere, the nine month old US sub-prime crisis was rolling across the Atlantic and while the European banking system was under pressure, only the United States economy was in recession.

The question was "*Did Veda's data show any insights into adverse shifts in the Australian economy*?"

The answer was yes. Veda was able to produce for Treasury a regular default data series, broken down by state and default type[1]. These series were compliant with the Privacy Act 1988 and were produced quickly and efficiently. In all cases, Veda took steps to make sure that no re-engineering was possible by the recipients.

Veda produces a range of other data insight series and is a regular commentator on consumer credit appetite, trends in defaults by age and demographic and changes to the quality of credit applicants.

RECENT EXAMPLES OF THE INSIGHTS DE-IDENTIFIED DATA PROVIDES

- A 2008 Veda bankruptcy study used 2003-2006 ITSA Statements and credit bureau information to show 95 per cent of bankrupts had applied for more credit knowing they were unable to meet their current credit commitments.
- A 2009 Veda study using credit bureau data showed applicants for credit were of better quality now than in 2007 (ie applicants are less likely to have negative information on their file)
- A 2011 study for the Reserve Bank is tracking credit stress amongst first home borrowers during the GFC versus those who borrowed in 2006.

If the Bill proceeds as drafted, these types of studies would not happen in the future.

The Regulatory model - De-identified data and Australia's standing as an information economy.

The question of privacy and de-identified data had been extensively considered by the Australian Law Reform Commission (ALRC) in its review of Australia's privacy law.

The ALRC considered secondary use of data over two years, through an issues paper, an options paper and finally a position paper. Its eventual findings made no recommendation for the regulation of de-identified information.

[1] for the information of the Committee, it was the motor vehicle default series that proved the most insightful, pointing to early signs of credit stress in Western Australia and Queensland

In Veda's view, this is self-evident. **De-personalised, data cannot identify or disclose personal or sensitive information (by definition) about the individual** and therefore has no personal information and hence no scope for protection as a human right, in this case privacy.

However in the Government's October 2009 response to the ALRC's recommendations, the Government firstly rejected the ALRC's findings on secondary use, proposing much more tightly defined secondary uses and then also proposed privacy control be extended to cover:

"De-identified credit reporting information for research purposes...research would also be required to be conducted in accordance with rules developed by the Privacy Commissioner."[2]

Data series, statistical modelling and historic trends are reliant on de-identified data. They contribute to insights into credit risk, fraud detection and identity authentication.

De-identified data was critical to a 2008 Access Economics study[3] that pointed to the likely benefits of comprehensive reporting to the Australian economy. More recently, a Veda pilot study with leading credit providers was able to replicate the Access Economics findings, providing the first 'hard' evidence that the big winners in comprehensive credit reporting are consumers who, under the current credit reporting laws, are denied access to credit despite being credit worthy.

Credit reporting information is extremely widespread in any modern economy. The holding of an active credit card, application for loan or a telecommunications or utility account contributes to a data holding that covers around 15.5 million credit active Australians.

Of itself, its wide spread pervasiveness gives de-identified credit reporting data a very high level protection, as opposed to health data, where specific medical conditions, by virtue of their rare occurrence, may be more easily identified in a de-identified series.

During the 2011 Senate Inquiry into the exposure draft bill, the extent of this overreach was recognised by the Office of the Privacy Commissioner, whose submission stated:

"Significantly, this approach would be the first time that the Privacy Act would regulate the use of de-identified information. Ordinarily, such information falls outside the Privacy Act's coverage as it is does not meet the definition of 'personal information'."[4]

---

[2] Enhancing National Privacy Protection, the Hon Senator Joe Ludwig, Special Minister of State, October 2009

[3] Access Economics "The benefits of Broadening Access to credit via Comprehensive Credit Reporting" commissioned by Veda Advantage 2008

[4] Para 73 of the OAIC submission 39a to the Senate Finance and Public Administration Inquiry into credit reporting provisions

The proposed move potentially robs Australia of the economic and social benefit data analysis can bring, including statistical modelling, developing trends and historic insights. Lack of access to de-identified data and its prescriptive regulation creates arbitrary distinctions in the information economy, stifling innovation and research.

It also appears to be a world first.

On 25 January 2012 the European Commission announced comprehensive reform to the existing data protection rules, promising to strengthen online privacy rights and boost Europe's digital economy[5]. No proposal was made in relation to de-identified data and neither the USA, New Zealand nor Canada have restrictions on de-identified credit reporting information.

The perceived risks

Much of the academic and policy research on de-identified or de-identified data focuses on the health industry and the risk of 're-personalisation' of the data[6]. Veda submits that while these concerns may be valid in the given contexts, they are not valid in the credit reporting environment.

This is so for a number of reasons:

1. **Anonymity - process of de-identification of information creates anonymity, and hence protection.**

Even if the information were to be re-identified, that information would become the subject of strict regulation, either under the credit reporting provisions or the generic protections afforded by the Australian Privacy Principles (APPs) under the Privacy Act. Non-compliance with credit reporting provisions carry a heavy sanction and a right of compensation by the affected individual[7]. For completeness, we also note that various State and Federal consumer protection regimes have a role to play. In almost 45 years of Veda's operation (as Veda and it predecessors) we have never re-engineered de-identified information collected as part of the credit reporting system to convert into personal information. Similarly, the matter has not been the subject of any complaints to Veda or its predecessors.

---

[5] http://ec.europa.eu/justice/newsroom/data-protection/news/120125_en.htm

[6] ucsdhc-web1.ucsd.edu/.../BULLETIN-HIPAA-FactSheet, digital.law.washington.edu/dspace-law/...1/.../vol5_no1 EPIC - Re-identification - Electronic Privacy Information Center, epic.org/privacy/reidentification/ - United States

[7] Division 7 of the Bill

2. **Credit reporting (by definition) relies on a closed user group environment and is therefore different to health**

Much of the concerns about use of de-identified data are discussed and debated in relation to health data. In that context, re-personalisation occurs by reference to external data sets or by the uniqueness of medical information.

Veda submits that these are very particular concerns and do not apply to credit reporting in Australia.

In credit reporting, unlike health, information sharing occurs for a very specific purpose (i.e. credit worthiness assessment) with a defined group (i.e. credit providers). Information is collected for the purposes of sharing amongst known users, all parties to the stringent regulatory controls of the credit reporting environment. There is no secondary use or disclosure outside the given environment. There is also no risk to the consumer.

3. **Dealing with the risk of re-personalising de-identified data**

Australia's first privacy laws were introduced more than twenty years ago, including credit reporting information, which was perceived to warrant greater protections. Subsequently, the rise of the internet and social media has created pools of information that offer many times more powerful insight into a person (eg google search history).

Instances of de-identified data being re-personalised have been reported in health (see Dr Latanya Sweeney[8]) and most recently examples of AOL Online in 2006, Netflix in 2007[9].

A common thread is the capacity for re-personalisation exists where there is largely unregulated access to public record information. For example in the USA access may easily be had to voter lists, including party preference; criminal records, vehicle ownership and rental records. Professor Paul Ohm in a study on re-identification found:

"Would-be re-identifiers will find it easier to match data to outside information when they can access many records indicating the personal preferences and behaviours of many people."[10]

---

[8] http://dataprivacylab.org/people/sweeney/

[9] http://www.cs.utexas.edu/~shmat/shmat_oak08netflix.pdf

[10] http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1450006 pg 1766

By contrast, Australia has very little that is available on public records and most have very strict access, use and disclosure requirements. Appendix 1 provides a summary and an illustration of some of these.

However it is prudent to recommend the inclusion in the legislation of substantial penalties for subsequent re-personalisation with substantial penalty provisions as apply elsewhere in the Bill.

Conclusions

For economic and regulatory reasons noted above, Veda submits that:

1. Current law on de-identified data should remain (subject to minor clarifications on purpose)
2. If the matter requires further review, it should be referred to the Productivity Commission
3. Appropriate penalties for misuse of personal information (of any kind) to be retained

Appendix 1

Restrictions on access to types of public records in Australia

DATASET	ACCESS RESTRICTED BY
Tax file number	*Tax File Number Guidelines 2011, Taxation Administration Act 1953* - restrictions & tough sanctions on collection or use of Tax File Number.
Electoral roll	*Electoral Act 2004* closed off general use. Authorised access for authorised organisations for *Anti-Money Laundering and Counter-Terrorism Financing Act 2006*.
Births, deaths and marriages	State registers heavily restrict any access to records
Tenancy (rental bond register) & Tenancy (adverse findings by Tenancy Tribunal)	New State Acts heavily restricts tenant databases, in addition to Commonwealth privacy principles and State Privacy Acts
Criminal record	Access via Crimtrac for prescribed purpose by authorised agencies only
Motor vehicle ownership & motor vehicles	Various State transport Acts restricts disclosure, not only of personal information, but also heavily constrains vehicle information
Business names registered to an individual	*Business Names Registration Act 2011* prohibits use or disclosure of address, DOB

Appendix Two – addresses retained on credit file

6 March 2012

Mr Richard Glenn
Assistant Secretary
Information Law and Policy Branch
Attorney-General's Department

Dear Richard,

I refer to our discussion of Thursday 1 March 2012 regarding Credit Reporting Exposure Draft Bill (the Bill) and the proposal to restrict addresses held on a credit report to current plus two previous addresses.

This restriction, in conjunction with the new prohibition on internal use, will have a serious impact on data-matching, **potentially impacting 2.4 million files** and frustrating the intent of Section 116 (quality of credit reporting information).

Veda understand the rationale rests with the Credit Reporting Determination 1991 No 2, concerning identifying particulars permitted to be included in a credit information file under the current Act. This 21-year-old determination was made in the context of the current Act and does not purport to govern how credit reporting agencies (CRA), as that term is defined, use the information particularly where that use is internal to CRA.

Since the time of the determination, credit reporting agencies have not only had access to the identifying information, but have used that information (internally) for a variety of data management and security related purposes. It can also be critical to the ability to investigate and, if necessary, address particular complaints. This is because there is, typically, a gap between when particular information is added to a file and when the individual makes their complaint.

It is the current capacity for internal use that has enabled Veda to use address information from a range of sources to ensure correct matching, whilst only retaining current plus two addresses on the actual credit reporting information held by Veda.

The newly proposed regulatory regime will make substantial changes in use and collection provisions, specifically in Section 108 which prohibits the use of credit reporting information (including identification information) except in the course of carrying out the agency's credit reporting business. There is no expressly permitted data management and related uses of the kind currently enjoyed by credit reporting agencies.

The percentage of people and the addresses available for matching purposes is:

a) One address – 45.5 per cent
b) Two addresses – 18.2 per cent
c) Three addresses – 19.2 per cent
d) More than three addresses – 17.1 per cent

Essentially, the impact will be on a subset of (d), where the input address received by Veda is not one of the most recent three, but a previous historic address - approximately 2.4 million files.

The potential problem is further compounded by the fact that credit reporting agencies ability to de-identify information will be severely restricted if not entirely removed by the provisions of S115 which expressly prohibits credit reporting agencies from using or disclosing de-identified information other than for very specific purposes identified in S115(2) or as permitted by a legislative instrument made by the Information Commissioner pursuant S115(3).

Options for remedy

1. **A specific provision** allowing credit reports to include, for the purpose of record management, the greater of:
    (i) All addresses over the previous five years; or
    (ii) The current plus last two addresses.

2. **Include a provision under S108**, allowing CRA's to use historic addresses, for the purpose of matching accuracy ("record keeping").

We note a further determination could be sought from the OAIC, but that such a determination would have to comply with the Act itself, which explicitly states the number of addresses to be collected and retained.

Accordingly, the prescriptive use provisions without an express acknowledge as to internal credit reporting agency uses remains a problem that needs to be addressed. The two issues need to be addressed concurrently so that they complement each other.

Richard, of the five issues raised by Veda this one has a very real capacity to undermine the integrity of Australia's credit reporting system.

We would welcome further opportunity to refine any suggested remedies.

Regards

Matthew Strassberg
External Relations