Submission to the Productivity Commission

Data Availability and Use

July, 2016

InFact Decisions Pty Ltd

www.infactdecisions.com

and

Verifier Australia Pty Ltd

www.verifier.me

Submission

Lisa Schutz Managing Director of InFact Decisions and Verifier, 24 July, 2016.

A Much Needed Inquiry

From my point of view, as a person who is both building new products that are classified as FinTech and runs a data and analytics consulting business, the current Productivity Commission inquiry into the role of data in driving innovation in the economy is a welcome initiative.

Data (as much as money) is the fuel of financial services. While the Commission's mandate is broader than just financial services, this submission will focus purely in this sector – and specifically the focus will be on consumer credit data.

Why Focus This Submission On Consumer Credit Data?

This is the area I know best. However, I would also suggest that data sharing in financial markets, with its considerable history, is a good sector to learn lessons that can be applied elsewhere.

This submission draws extensively on a paper presented at the Australian Centre for Financial Studies, 21st Melbourne Money and Finance Conference – 18-19 July 2016 "Fintech and Financial Innovation" conference. A copy of the paper, co-authored by myself, is provided in Appendix A of this submission.

I. In Short – A Framework to Consider

The goal of credit data sharing is to have the most effective domestic financial services sector, one that can continue to evolve in the use of credit data which is crucial for that effective functioning – and always has been.

In turn, a highly functional domestic financial services system is likely to create organisations that can compete globally – either as data sharing infrastructure providers or as lenders. Certainly, since most jurisdictions aim for high levels o credit data sharing it makes sense for Australian lenders to get their learnings locally before they attempt to move offshore. And this helps them protect against new entrants to the domestic market.

The challenge is that data sharing affects significantly the competitive balance for the incumbents and their challengers. While FinTech is a new sector and is making demands for access to more data from the Government, the clamour of challenger lenders to access incumbent's customer data is as old as consumer credit markets.

The intent of the suggestions made in this submission is to support a consumer led, permission based overlay to the current credit data sharing framework. This, with additional participation (telecommunication and utilities firms) in the CCR regime will ultimately, I believe, pave the way for more voluntary lender driven sharing.

For Consideration – five foundational steps required to achieve dynamic, comprehensive consumer credit data sharing:

1. **Secure Courier** access to consumer data: Access to consumer data must be via secure courier style methods (not impersonation methods) – this is in line with the OpenID Financial API working group goals and the Attorney General's Document Verification Service

identification gateway.

2. **Extensions to Part 5 of the Australian Privacy Principles**: The individual's right of access to their data (Principle 12) is clarified to require that data sources provide:
   - access **at scale**
   - access via a **secure courier** methodology
   - **authentication** of the information

   Supplying data directly via API achieves all these three requirements. What would no longer be acceptable would be data sources declining to provide API linkages to data when requested (and offering paper as a substitute).

3. **Enable Full CCR via either**

   Option 1: Mandatory Comprehensive Credit Reporting; or

   Option 2: Support voluntary CCR by the following three steps:

   - Include telecommunication and utility data in the Voluntary CCR model; and
   - Mandate partial CCR to enable consumer opt-in
   - Support consumer opt- in via Steps One and Two

4. **Appoint a data supervisor - use of data in credit model supervision the responsibility of one regulatory agency.** This agency in turn reviews the modelling from prudential, consumer protection and privacy perspectives.

5. **Creation of an anonymous credit risk analytics data pool to support ongoing credit model innovation**. This analytics data pool (or multiple, competitive ones) is designed to enable more rapid innovation in credit models and to test the case for admission of new types of data into credit models. If in test cases credit departments are able to request new classes of data - say, for instance, LinkedIn profiles - then they can be tested in the sandpit to see if the meta data is useful in credit assessment process. If the case is made, and the credit model supervision agency judges that the privacy issues are outweighed by the value of less incorrect credit decisions, then lenders can start asking consumers to opt in to sharing that extra data. Bearing in mind that there is a social dynamic to using data for consumer credit assessment and that CCR data is likely to be the most predictive in all cases (and sharing it a necessary reputational mechanism).

II. The Context

Consumer Data Sharing Is More a Domestic Issue Than an Export Opportunity

While FinTech might have export potential, if nothing else, the Government's support for the FinTech sector is sensible to promote an innovative domestic sector and enable that sector to be resilient to overseas new entrants (possibly FinTech as well as conventional financial services operators).

Data availability, in terms certainly of consumer credit data, has more impact on the FinTech sector's ability to disrupt **incumbent domestic suppliers** of financials services, since that data confers on incumbents an informational advantage that translates to revenue opportunity (marketing/cross selling) and cost opportunity (particularly the ability to minimise credit risk).

Data Sharing in Financial Services – The Usual Suspects

The debate around data sharing in this particular sector revolve around the following players:

1. The consumers
2. The incumbent lenders
3. The challenger lenders (be they standard types of new entrants or ones from the emergent FinTech sector. In terms of domestic competition challengers are virtually identical in their data demands.)
4. The regulators
5. The community – including advocates who seek to represent at risk segments of society who might be harmed by the functioning of the market
6. Challenger data providers – including social media organisations who might wish to challenge the regulated consumer credit data paradigm.

The Goal – Create A More Effective Financial Services Sector Through Use of Data

Regulation of financial services is unquestionably a high wire act. The normal goals apply – the need for the operators in the sector to earn reasonable returns and to do so as flexibly as possible so that competition and innovation is encouraged, leading to better outcomes for consumers, society and shareholders. Likewise, there is the need for customers to be protected from predatory/problematic corporate behaviour and specifically with regards to financial services (as is the case for health care and government data) have their private information treated as such. Additionally there are the systemic/stability issues that are specific to Financial Services. As the US sub-prime mortgage crisis showed, consumer credit can create systemic problems. On a global scale.

Australian financial services has a history of technological innovation and likewise of avoiding the level of banking failures experienced elsewhere.

III. The Problem

The Problem – Information Asymmetry

Financial services is a sector, particularly lending, that is very susceptible to the costs of information asymmetry. Good credit assessment and allocation of loan funds is good for society as it reduces the costs associated with repayment uncertainty. When lenders cannot distinguish good borrowers from bad borrowers, all borrowers are charged an average interest rate that reflects their pooled experience[i]. This type of cross-subsidisation means that borrowers with productive uses for loans are cross subsidising less qualified borrowers who are less likely to put their loan to productive purposes. In addition, lenders charge a premium to cater for a level of uncertainty in the risk rating process adding unproductive cost to the economy.

However, we live in an increasingly intrusive world and there must be limits to the intrusion of industry into the lives of their customers. Where to draw the line and how to balance the legitimate interests of all the stakeholders is the core question of this paper.

The Problem – A Constantly Changing Data Landscape – Future Proofing Regulation

We live in a world that, as a result of smart devices, the internet of things technology, better tools for handling large amounts of disparate data, more emphasis on statistics/machine learning and the advent of servers as a service (e.g. "the cloud"), there is simply more ability to digest and generate insight from data than ever before.

As the volume of data and the analysis of that data grows exponentially, we do not really know what is likely to be the frontier of activity in twelve months.

In my view, what is needed is a system that can continuously redraw boundaries in response to the constantly changing landscape of data and analytics, that balances all the interests of the sector and the community.

Boring but True - The Best Predictor of Future Behaviour Is Past Behaviour

In the context of answering the willingness and capacity to pay questions that are the core of credit assessment, credit reports and income data are crucial. That is not to say that new types of data (including social media information) could not be helpful. They could in certain circumstances – mostly when other data is not available. But let us start the conversation with the basics about credit reporting. If we get that data flowing it will make the most significant impact on the sector.

Key credit data fields are relatively constant – the best indicator of whether people are **willing** to repay debt is whether they have before (credit reporting) and the best indicator of whether people have the **capacity** to repay debt is their income. NCCP does require a view on expenses, but it should be noted that expenses are far more consumer controlled and can fluctuate dramatically, so it is less helpful .

Role of Credit Reporting – Reputational Mechanism and Assessment Tool

Data sharing in consumer credit markets is primarily about keeping bad debt costs low and avoiding, not just the economic but, the social costs implied from misallocation of credit. As such, while the means is data sharing, the motive is not just credit assessment, there is an element of social engineering involved. Sharing credit information acts as a reputational mechanism. Consumers are careful not to get overextended because this will backfire in later loan applications.

Which brings us to the consequences of consumer credit data used in credit applications – you apply, data is shared about you, these days predictive models not people are the main way lenders assess your risk of not repaying the loan. On the basis of this assessment, you either get your loan approved or you find yourself declined.

Data sharing via credit reports, enshrined in Part IIIA of the Privacy Act, is a specific carve out from the general privacy principles that enables lenders to access private information for this purpose from the credit bureau not from the person directly. The reason is obvious – people won't always volunteer the information that they didn't repay a loan – whether they lie or not – they do have a strong disincentive to forget that fact.

The question becomes what data will society allow to be shared about people "behind their backs" (as in it is passed from credit bureau to lender) that might affect their chances of gaining credit. As a society, in Part IIIA of the Privacy Act we have listed the attributes that we will allow lenders to share in this way.

Rebutting The Idea That New Types of Data Are Better Than CCR

The author has been a little startled in recent times to hear the argument that social data, personality type data etc., is predictive of credit risk and supplants comprehensive data. There are two issues for policy makers to consider here which will explain why I am focused on getting the comprehensive data sharing working correctly and why I firmly believe we need to supervise use of data in credit models (Step 5) and why I would strongly **disagree** with the assertion that comprehensive data is no longer necessary.

There are three kinds of data used in risk assessment:

- Data about my personal behaviour in regards to credit **that society enshrines in credit reporting legislation** (Part IIIA of the Privacy Act) – as discussed above.
- Data about me unrelated to credit **that society will not let lenders use decide whether to give me credit or not** – my gender, religion, race etc. – as agreed and implemented via our anti-discrimination laws.
- Data about me that lenders can ask for on applications and model and use for risk assessment – but the individual knows about because they fill in the application – if they leave the field blank that is their choice.

IV. If It Matters - What Went Wrong?

It Matters – So Let's Get the Basics Right

Even without the additional complexities of the data/analytics rich world we live in, the challenges of CCR (Comprehensive Credit Reporting) and NCCP (National Consumer Credit Protection) highlight the issues that face those in support of more data sharing in consumer credit markets. Appendix A goes through the problems regarding the unintended consequences of both the CCR and NCCP legislation and the principles and challenges therefore for regulators.

Below is a bit more detail on those two attempts to address high quality credit data – and then what follows is a generalised set of steps to effect a system that can overcome the problems identified.

Comprehensive Credit Reporting (CCR)

CCR was an attempt at black letter law to legislate for more data sharing to address the **willingness** of consumers to repay credit issue. The rationale and precedent for this kind of legislative intervention is exceptional – and based on a strong base of academic literature that highlights the benefits[ii]. As a result, the World Bank is a big supporter of introducing comprehensive credit reporting.

New Zealand has adopted positive credit reporting, arguably more successfully than Australia. Possibly, the competitive dynamics in Australia (of four large banks) versus New Zealand (one larger bank) had an impact (refer Appendix A). Arguably, also, the use of a code versus legislation model assisted New Zealand as well as involving more than just lenders – extending the scheme to telecommunications and utility providers.

However, the risk of re-examining CCR legislation is the risk that the same lobbying pressures that came to bear originally will likewise dilute the result this time around. Will more data be shared or will the competitive dynamics that are observed globally create the opposite effect[iii].

National Consumer Credit Protection – Serviceability Test

At the same time. NCCP (as per Appendix A) was a case of principles based legislation addressing the issue of **capacity to pay** by requiring lenders to check the ability of consumers to service their loans. This required checking of income and expenses and led to the unintended consequence of irritating both lenders and their customers since data for this process could not be accessed readily online. The result was a rise in manual processing, longer times to yes and higher cost, not to mention consumers dropping out of application processes.

A small cohort of lenders has now adopted third party tools that ask consumers to reveal their banking portal user names and passwords in order to extract bank statement data and thereby conduct these checks in a more automated way. The problem with that, and the reason for the low adoption rate, is that these services use an **impersonation** method. Regardless of where the password is stored and for how long, the consumer has revealed their private credentials in a way that impinges information security and breaches their contractual obligations to keep these passwords secret. They also have no ability to review the material before it is sent – which does not align with Australia's privacy principles. Submissions to the Department of Treasury by FinTech Australia have requested that current methods of data collection if this kind be legitimised[iv].

However, an alternative, and in my view, a superior approach, is to require data access by third parties to consumer data managed through what are loosely called "secure courier" approaches to

data access. The Attorney-General's Department Document Verification Service is a good example of this. The key principle is that third parties seeking access to consumer data on a consumer's behalf **cannot impersonate** the consumer in order to gain access, but must identify themselves in the process. This necessitates a specific API connection and ensures that the data source is aware of who is accessing their data with consequent security benefits. The issues in regards to the **method of access** of consumer data has already been raised as a concern by ASIC in two instances[v].

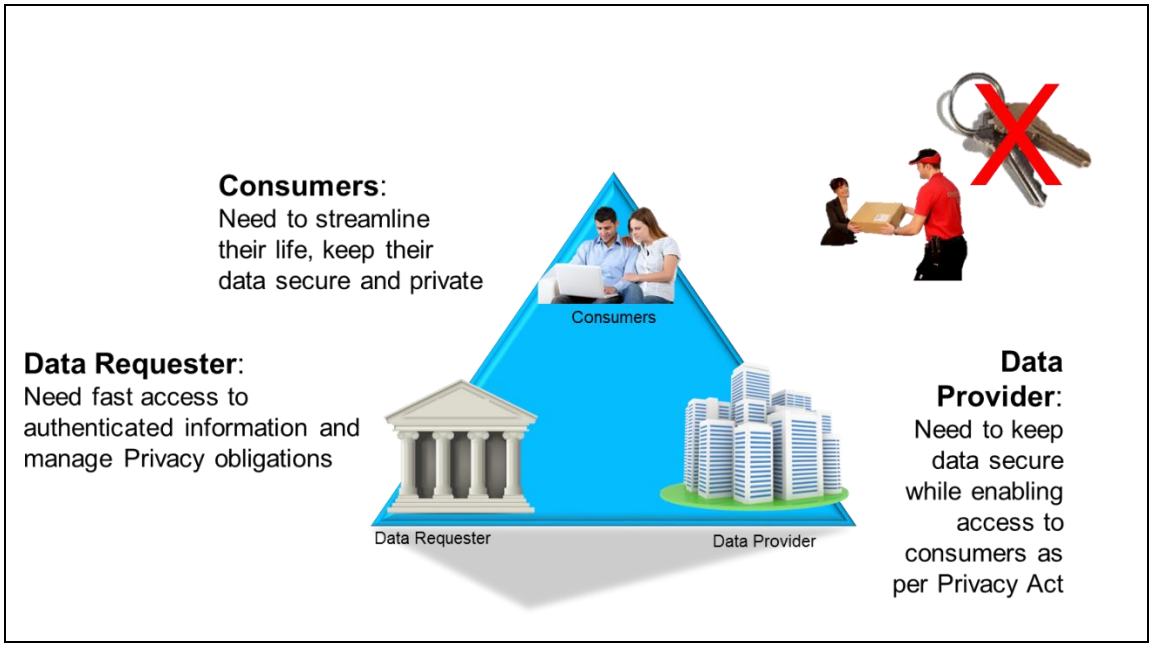

Figure 1 - The three parties to a data exchange (excluding any third parties that may assist them in effecting data sharing)

V. Data Sharing Philosophies

The Hunt for Solutions - Looking for Data Sharing Principles to Guide Us

The challenge we face is how to manage the market for credit data, well summarised as follows[vi] - the bottom line being data needs to be viewed in context of the market and the situation for which it is directed:

> *What is the allocation of surplus gained from the usage of individuals' personal data? How should that surplus be allocated — based on market forces, treating privacy as another economic good, or based on regulation, treating privacy as a fundamental right? And should an allocation favor the data subject as the owner of the data, or the data holder who invested in collecting and analyzing the information?*

When new data is shared it changes traditional information asymmetries which in turn advantages some participants over others:

> *In choosing the balance between sharing or hiding personal information both individuals and organizations face complex, often ambiguous, and sometimes intangible trade-offs*.

The principles for data sharing presented by UK Government[vii] provide a useful foundation for considering how credit data should be shared:

- *For data sharing to be useful to users, it should be simple, low friction and scalable;*
- *Users should provide fully informed consent before their personal data is shared and should remain in control of how it is used;*
- *To create optimal conditions for innovation, datasets that do not contain personal or commercially sensitive information should be made as accessible as possible;*

However, these principles need to consider the unique aspects of credit data, particularly recognising its distinct purpose including its role as a reputational mechanism and need for prudential oversight to prevent adverse financial system outcomes. The cost of maintaining the veracity of the data must also be considered, particularly incentive for the custodians of the data to continue to invest in its development.

The Unique Consumer Credit Context – Data Sharing Motivates Sustainable Consumer Behaviour

Credit data sharing, as mentioned earlier, has a strong reputation all effect which is not always the case when discussing data in the economy. As such, it is well worth reflecting on this role and a good reference is a paper authored by Daniel Klein in 1992[viii]. While illness, divorce and unemployment – primary causes of financial hardship – are the kind of life events no one is immune from or can necessarily predict – keeping debt manageable to handle the ups and downs of life as best one can is the primary obligation of the consumer. And, the credit report is one way that that obligation is reinforced.

Drawing On Consumer Agency Thinking from Technologists

When data sharing is discussed, it is generally considered as data sharing **about** an individual **between** two third parties. That is certainly the model of the credit bureaus, the primary data sharing mechanism in consumer credit.

At the same time, however, technologists are looking at the problem differently - starting from the position that it should be possible to share data without central repositories, just as the internet functions without a command/control approach. Interestingly, that philosophy matches with technical capabilities we hear about every day - the internet of things, distributed computing and the flexibility of common application programming interface standards (API protocols). So, it's the technologists that are forcing us to examine the philosophical underpinnings of the choices we make and the trade-offs we accept. Privacy v. efficiency would appear to be a false dichotomy if you dig into the work of the groups highlighted below.

Below are brief summaries of four work streams occurring globally that are working towards realising the vision of a privacy oriented data sharing economy. The list is not exhaustive, but rather, I am seeking to highlight some of the work that might not get picked in a financial services oriented literature review. These lines of though do point to the ability for consumer opt-in as a practical way for getting consumer credit data to flow more effectively – which is a new paradigm in how we think about data sharing.

1. The privacy by design philosophy

   This approach takes the position that privacy can be embedded in design of new systems/business models in such a way that it creates a win between functionality and privacy. Tools are built this way across many sectors. A very good paper that illustrates the potential was co-authored by the Privacy Commissioner of Ontario of the time - Ann Cavoukian [ix].

2. The openPDS/Safe-Answers project – at MIT

   This approach is advocated by a number of the most influential data scientists operating globally, and is designed to enable rich data analysis while preserving full anonymity. In their model, the individual controls their raw data (or an agent does on their behalf). Researchers can share algorithms with individuals via their personal data stores, which are designed to use the full richness of the data in the personal data store to create the answer to the research question and retrieve it without any identifying information. Because the answer is to a specific question, the dimensionality of the information being retrieved from the personal data store is so reduced, it makes it almost impossible for hackers to guess who the person was from the "safe answer". This method was designed in response to the problem identified by these same researchers via their repeated ability to re-identify even completely anonymised data sets[x].

3. The Project VRM Working Group(s)

   This is a project centred on the idea of reversing the paradigm of CRM – Customer Relationship Management. It starts from the idea of the consumer selecting vendors, with the consumer as the active, controlling agent. Central to this thinking is the concept (and reality) of individual controlled data and what the world would look like if consumers, via their intentions, drove commerce and enterprises had to respond on their terms. A small example – imagine if I could update my address in my personal data store and every enterprise I wanted to could access that updated data. In a world with consumer agency that is possible.

   Project VRM emerged from the work of David "Doc" Searls particularly the book published

originally on the internet in 1999 (with several collaborators) entitled the Cluetrain Manifesto. Project VRM now involves many researchers, technologists and organisations. Project VRM was originally incubated at Harvard's Berkman Center for Internet and Society. The key to this work is its practical development emphasis, which involves technologists working on how to realise a consumer centric, permission based economy.

For instance, a related but separate is the work on OpenID and the considerable efforts going on globally to develop decentralised identity models to streamline online activity. Specifically, see below for discussion of the Financial API working group which has been recently launched to address the issues around the impersonation method raised earlier in this paper.

4. OpenID and within that the FAPI working group which was founded in March 2016 (FAPI)

The charter of the FAPI (Financial API) working group, founded in March of this year, is worth referencing because it speaks to the heart of the first step suggested in this submission – the principle of secure courier as the required method for access to an individual's data.

"In many cases, Fintech services such as aggregation services uses screen scraping and stores user passwords. This model is both brittle and insecure. To cope with the brittleness, the new OpenID Foundation Work Group invites developers, architects and technologists to contribute to an open standard approach using an API model with structured data and to cope with insecurity, it should utilize a token model such as OAuth [RFC6749, RFC6750].

The OpenID Foundation Financial API (FAPI) Working Group aims to rectify the situation by developing a REST/JSON model protected by OAuth. Specifically, the FAPI Working Group aims to provide JSON data schemas, security and privacy recommendations and protocols to:

- enable applications to utilize the data stored in the financial account,
- enable applications to interact with the financial account, and
- enable users to control the security and privacy settings.

Both commercial and investment banking accounts as well as insurance, and credit card accounts are to be considered[xi]".

Leaving aside the technical dimension to this charter, the point is that Australian financial services regulation will not assist the FinTech community if it tolerates data sharing models that impersonate the individual and are unsustainable in the long run (even if they request it[iv]).

What These Approaches Have in Common – Consumers in The Driver's Seat

The core premise of the work streams listed above is that the individual, in this age of distributed computing, can truly be an active agent in the data economy.

Lack of voluntary reporting of comprehensive credit data and the problem of the impersonation method being used to access income data, are symptoms of the lack of this consumer agency. And, as a result they can be solved by putting the consumer at the centre of the financial services data

ecosystem, creating an opt in overlay to existing data regimes (credit reporting) and creating a new model for accessing data.

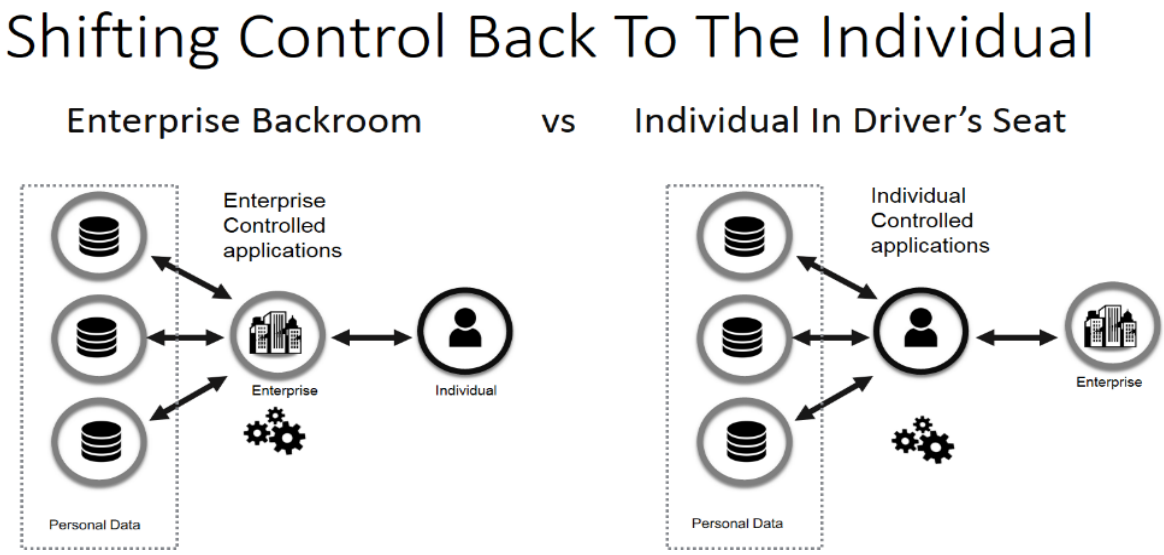

Figure 2 Consumer agency in practice

VI. Solution – Consumer Opt-In Creating a Dynamic Data Sharing Environment

In this section, I articulate the five recommended steps to achieving the consumer credit data sharing outcomes and future proofing that data sharing for a constantly shifting data environment.

1. Consumer Opt-In to Address Credit Capacity (The NCCP Requirement)

Consumers, evidence suggests, are just as irritated as lenders about providing the data regarding **capacity** to pay. Lenders and consumers are looking for online solutions.

The principles to enable the data sharing at scale required leverage existing practices and legislation:

Step One: access to personal data by third parties must occur by representing the identity of the third party. In other words, a decision that the secure method of gaining access to data be used as opposed to the impersonation method.

With this approach, consumers do not have to trade off the convenience of straight through processing and their information security when transacting with Australian lenders.

Step Two: Clarify the individual's right of access to their data in the Australian Privacy Principles to incorporate the following:
-Can the data be **used at scale** (i.e. electronically – via application programming interface)
-Is the data delivered **securely** (including the issue of not impersonating the individual in the process as per Step One)
-Is the data **authenticated** as part of the delivery so that other parties can rely on its veracity when the consumer shares it with them?

The consumer's right to review their data and challenge the data held on them remain as per the current Australian Privacy Principles.

At the point these tests are met, instead of the right of access being focused on individuals checking their records are accurate, the right of access would confer on them the ability to share their data with other organisations in a way that those other organisations can meaningfully use that data in their business process.

A consumer in the driver's seat could use their access right to source proof of income data (and expense data), which would fix the problem created by the sensible principles of the NCCP legislation which, as has been discussed, created a need for data which could not be supplied at scale.

2. Willingness to Pay (CCR) – Using Consumer Opt-In to Tilt the Balance

Once the two steps above are in place, the ability exists to rethink the comprehensive credit reporting dynamic. I would suggest that by leveraging consumer opt-in, additional fields beyond those considered in the CCR legislation might be made available over time. However, and this is key, only with the individual's explicit consent. And, in the shorter run, consumers sharing comprehensive information about themselves, would reduce the disincentive for large banks to share the data they have on customers (because the consumers have shared it already).

The Options – Mandating CCR or Making Voluntary Contributions Compelling

Fundamentally, there are two high level options for getting full comprehensive reporting to happen. Either, mandate it in some way or manage the nature of Voluntary CCR so that lenders find it compelling to contribute. I would reiterate the earlier comments that there is a very strong empirical case to extend beyond the fields currently provided for in the CCR regime in Australia – I have called that "Mandatory CCR+" below. For instance current outstanding balances is a field that is noticeably missing and highly predictive of risk outcomes. A good resource on global best practice and why it would be worth increasing beyond the current five fields in CCR – the World Bank's 2011 report on best practice in credit reporting[xii].

1. **Make it mandatory, one way or another**
   - At the current number of fields (five) – **CCR**
   - Extending the fields and making all of them mandatory – **Mandatory CCR+**
   - Make **use of the data** mandatory – which would create an innate demand for which the current CCR scheme is already set up (this obviously feels like making CCR mandatory – but with different legislative impacts) – maybe via clarifying the need to use CCR in ASIC's Regulatory Guide 209 to Serviceability (under the NCCP legislation).
2. **Make Voluntary CCR compelling for large lenders so that they start contributing**
   - **Include telecommunication and utilities data**
     - In the paper in Appendix A, numerous comparisons to the New Zealand case are made. The specific option that seems the most likely to succeed in tilting the balance towards lender participation would be to invite telecommunication and utility provider participation.
   - **Harness consumers via consumer opt-in methods to tilt the dynamics in favour of voluntary sharing**
     - As will be discussed in the next section, consumers who opt in are likely to be lower risk customers in search of a better deal. These are the exact customers that are problematic for large lenders looking at voluntary CCR – the benefits in lower risk they gain are offset by sharing data about their best customers. The more consumers share, the less there is a disincentive to sharing.

Consumer Opt-In – Creating A Tipping Point

I would suggest that consumers have a strong incentive to share their comprehensive data in this market, on the basis that doing so will lead to better outcomes for them. Either in more access to credit, or on better terms. This presumes a sufficiently competitive environment to foster product innovation.

Quite rightly, only good customers will want to share additional data. This means that over time, the lenders will get much more nuanced information about the lowest risk borrowers in the market.

The group that will remain largely unknown are those individuals who do not yet have an adverse credit history but who are in a cycle that is potentially headed that way.

If good customers have already shared via consumer opt-in then the banks no longer have the same game theory disincentive to share their comprehensive data because the net impact to them will be positive as they avoid risks they cannot yet see. So, by allowing consumers to drive sharing, the result might be more voluntary sharing by lenders.

Accelerating Opt-In – Mandating Partial CCR

Consumer Opt-In is rendered far less effective if **overall indebtedness** is not known. This is provided by mandating the four fields in CCR that supply this picture – great if this can be extended to getting lenders to share the behaviour on those facilities – but as suggested over time I think consumer opt-in building off a partial CCR platform would tilt the balance towards full, voluntary reporting.

So, as a solution to support voluntary full CCR, I do believe there is a strong case for mandating partial CCR. To be totally clear, this is NOT a plausible alternative to full CCR because partial CCR data is just not predictive enough as it omits repayment behaviour.

However, if the decision was taken to support the current voluntary full CCR regime then mandatory partial CCR would be a way to support consumer opt-in which is a part of the solution. In my opinion, the other part of the solution, as mentioned earlier, would be to extend CCR to utilities and telecommunications companies.

Where Does Mandating Full CCR Leave Consumer Opt-In? An Ongoing Role

To be clear – the most predictive field in the CCR program is the repayment history. To implement CCR without that field being incorporated is setting Australia up for a sub-optimal outcome. If the decision was taken to mandate CCR, then that would solve for the current issue but it would still leave the problem of a fixed set of data fields enshrined in black letter law.

I would argue that even if full CCR was mandated, then consumer opt-in is still a valuable mechanism to enable additional to find its way over time into the credit data sharing regime.

And, if only partial CCR is mandated, then consumer opt-in would enable consumers whose lenders were withholding data to “top up” their data if it suited them and kick start voluntary lender data sharing.

A "Tripod" Model

What we end up in this model is a "tripod" – better credit risk assessment delivered via:

- The existing adverse data (uncontested)
- Full CCR (with telecommunication and utility data)
- Support for consumer opt-in by mandating secure courier options and enhancing the privacy right of access

Step Three: Enable full CCR via either

Option 1 - Mandate CCR; or

Option 2 Adjust Voluntary CRR to make it compelling – the steps are:

A Include telecommunications and utility providers in CCR

B. Mandate Partial CCR to support consumer opt-in

C. Support consumer opt-in with Steps One and Two

There will no doubt be numerous submissions arguing for mandatory full CCR. Fintech Australia has already submitted to the Department of Treasury that it would like mandatory CCR for an expanded number of fields[xiii]. The global evidence (and the reason for this being a World Bank priority) would suggest that would have real domestic economic benefit. Global evidence, however, also points to the difficulty of such changes in the presence of entrenched large incumbents with research showing clearly a drop in adoption in comprehensive data with higher market concentration.

Using consumer opt-in liberates the supply of data from special interest capture. And injects ongoing dynamism regarding what data is shared.

3. Controlling Use of Models – The Role of Data Supervision

While the first three steps focused on freeing up the availability/supply of data, the next steps focus on their effective and responsible use.

The oversight on use of the data, to prevent the kind of systemic risk and overzealous data collection that is feared by privacy and consumer advocates, is handled by a new type of data supervision of lenders operating in the market. In other words, control demand for data.

Step Four: One agency to supervise data use in credit risk models.

That agency works to consider holistically the issue from the view point of prudential supervision, consumer policy and privacy.

The agency either formed or nominated would have clear accountabilities to all regulators in the sector but avoids the fragmentation issues of multiple agencies and the fact that not all organisations using these models are supervised by the same regulators (APRA for banks for instance but not for other lenders). This would ensure consistency and give a point of focus. Since data is so

core to financial services, I would suggest that making credit risk modelling supervision central, not incidental, is crucial to adapting to the data rich world we live in.

4. Anonymous Modelling Sandpits – Enabling Evolution of Credit Models

If consumers start to share more data, then a new flow of data is available for credit risk modelling. For new data fields to be used in credit risk decisions, they need to be modelled in the aggregate, to see how predictive (or not) the new data is.

As the use of data in models is opened up because of the supply of data, the issues that are confronted when legislation is written in this area need to be addressed via supervision:

- Is this data really predictive over time (models can be over fitted and lead to lack of resilience in results)?
- Is this data fair to be asked for and used – is the increase in fine grained risk assessment such that it justifies the invasion of privacy
- Is this data fair in the sense that if it drives denial of a loan application – do other social policies (such as anti-discrimination) mean that we will not allow that data to be used in this way.

The overarching principle would be that if an individual elects to share more data via consumer opt-in, they also give permission for the lender to share their anonymised application stage and credit outcome data for the purpose of ongoing research. That enables lenders to fine tune their risk models, and test and learn to see if new classes of data are predictive or not. The regulator/supervisors can also see these models and make decisions about whether to allow use of new types of data (e.g. Facebook or LinkedIn information) be used in credit assessment. This means that new types of data can be experimented with and the community via the regulators gets a say in whether the info is admissible or not.

The data is held should be anonymised at the outset. How is a significant question that requires consideration by a combination of regulators, credit risk practitioners and technologists/data scientists. What would be the exact design to provide both data richness and anonymity is beyond the scope of this submission.

For now, the point here is that it can be done, and there is plenty of work going on around the word to enable analytic richness without compromising the concept of full anonymity. Plenty has been written about the ability of data to be de-identified and the limitations of this. At the best this is always at most a probabilistic solution which needs to be factored in to setting the appropriate balance . The work of Yves-Alexandre de Montjoye and from MIT highlights this commendably[xiv].

There are a spectrum of approaches to anonymisation. The most extreme level would involve no ability to link data about loans held by the same person across multiple lenders. In that case, there would data in the analytics sandpit about ten different loan facilities however the fact that these belonged to the same person would not be known. Lenders would submit their data without any personal information, the record they shared would hold only the information they had when they decided to lend the money and the loan outcome. Alternatively, there is strong precedent for the use of common algorithms to strip personal information from each record before the record leaves the lender's environment but leave a unique identifier that enables linkage in the data pool. These

anonymised records are able to be matched across lenders because the algorithm that each lender uses treats name, for instance, in the same way. Obviously how that is done to prevent re-engineering of identifying information is pivotal. However, if safely adopted, the result would be data stripped of personal information but with data linkages intact – so ten loans related to one person would be identified as belonging to an individual (without any personal identifiers).

Regardless of the exact technical approach, the principle has to be clear that this the data is formed for the purpose of **refinement of credit risk models** rather than for the purposes of assessing specific loan applicants. An analytics sandpit is not a stealth way of creating a larger credit bureau but a new approach that supports consumers having more (not less) control over the data they share about themselves.

Credit bureaus, with their role as credit data repositories and their connectivity to the lenders, are well placed to provide this service but no doubt there are other organisations who could offer these services.

Step Five: A new role in financial services for holders and operators of anonymous credit risk modelling "sandpits".

In Conclusion – A New Model for Credit Data Sharing Using the Power of The Consumer

The intent of suggestions made in this submission are to support a consumer led, permission based overlay to the current credit data sharing framework. Which, in turn, will over time pave the way for more voluntary sharing under the CCR regime and future proof the use of credit risk data to enable innovation and use of new data as it presents itself over time.

The Steps:

1. **Secure Courier** access to consumer data: Access to consumer data must be via secure courier style methods (not impersonation methods) – this is in line with the OpenID Financial API working group goals and the Attorney General's Document Verification Service identification gateway.

2. **Extensions to Part 5 of the Australian Privacy Principles**: The individual's right of access to their data (Principle 12) is clarified to require that data sources provide:
   - access **at scale**
   - access via a **secure courier** methodology
   - **authentication** of the information

   Supplying data directly via API achieves all these three requirements. What would no longer be acceptable would be data sources declining to provide API linkages to data when requested (and offering paper as a substitute).

3. **Enable Full CCR via either**

   Option 1: Mandatory Comprehensive Credit Reporting; or

   Option 2: Support voluntary CCR by the following three steps:

   - **Include telecommunication and utility data in the Voluntary CCR model; and**
   - **Mandate partial CCR to enable consumer opt-in**
   - **Support consumer opt- in via Steps One and Two**

4. **Appoint a data supervisor - use of data in credit model supervision the responsibility of one regulatory agency.** This agency in turn reviews the modelling from prudential, consumer protection and privacy perspectives.

5. **Creation of an anonymous credit risk analytics data pool to support ongoing credit model innovation**. This analytics data pool (or multiple, competitive ones) is designed to enable more rapid innovation in credit models and to test the case for admission of new types of data into credit models. If in test cases credit departments are able to request new classes of data - say, for instance, LinkedIn profiles - then they can be tested in the sandpit to see if the meta data is useful in credit assessment process. If the case is made, and the credit model supervision agency judges that the privacy issues are outweighed by the value of less incorrect credit decisions, then lenders can start asking consumers to opt in to sharing that extra data. Bearing in mind that there is a social dynamic to using data for consumer credit assessment and that CCR data is likely to be the most predictive in all cases (and sharing it a necessary reputational mechanism).

[i] Barron JM and Staten M 2004; The Value of Comprehensive Credit Reports: Lessons from the U.S. Experience; May 2004

[ii] World Bank, Rethinking The Role of The State in Finance, 2013, Chapter 5 – The Role of the State in Financial Infrastructure – particularly citing Bruhn, Miriam, Subika Farazi, and Martin Kanz. 2012. "Bank Concentration and Credit Reporting." Policy Research Working Paper, World Bank, Washington, DC

[iii] Bruhn, Miriam, Subika Farazi, and Martin Kanz. 2012. "Bank Concentration and Credit Reporting." Policy Research Working Paper, World Bank, Washington, DC

[iv] Fintech Australia, 24 February, 2016, Priorities for Reform of the Australian Financial Services Industry - submission to the Department of Treasury, specifically the Open Financial Data recommendation-

1. Legitimise the current practices around financial data aggregation and mandate standard open data APIs.

[v] ASIC review of account aggregation in the financial services sector, (31 May 2001) http://asic.gov.au/regulatory-resources/find-a-document/consultation-papers/cp-20-accoun t-aggregation-in-the-financial-services-sector/ AND ASIC report on payday lenders - A more recent review of payday lenders (17 March, 2015) that mentions 3rd party aggregators and also refers back to the original report. http://asic.gov.au/regulatory-resources/find-a-document/reports/rep-426-payday-lenders-and-the-new-small-amount-lending-provisions/

[vi] Acquisti et al 2016; The Economics of Privacy; March 8, 2016; Acquisti, Alessandro and Taylor, Curtis R. and Wagman, Liad,

[vii] Data Sharing and Open Data for Banks – A report for HM Treasury and Cabinet Office; September 2014; Fingleton Associates,

[viii] Klein D. (1992) "Promise Keeping in the Great Society: A Model of Credit Information Sharing", Economics and Politics 4(2):117 - 136

[ix] Cavoukian A. (Privacy Information & Privacy Commissioner Ontario, Canada) and Jonas J. (IBM Fellow, Chief Scientist IBM Entity Analytics), June 8 2012, Privacy by Design in the Age of Big Data, Privacy by Design Canada.

[x] De Montjoye YA, Shmueli E. Wang S., Pentland AS, open PDS: Protecting the Privacy of Metadata through Safe-Answers. PLoS ONE 9(7): e98790. doi: 10.1371/journal.pone.0098790

[xi] http://openid.net/2016/05/23/announcing-the-financial-api-fapi-working-group/

[xii] World Bank (2011), General Principles for Credit Reporting, particularly points 54-60 http://siteresources.worldbank.org/FINANCIALSECTOR/Resources/Credit_Reporting_text.pdf

[xiii] Fintech Australia (the Australian financial services technology industry association), 24 February, 2016, Priorities for Reform of the Australian Financial Services Industry - submission to the Department of Treasury, specifically the LENDING Priority Objective 1: Comprehensive credit reporting to be made mandatory for large credit providers by the end of 2016, and small credit providers by the end of 2017; several additional data fields to be introduced across the board immediately, standard open data APIs.

[xiv] De Montjoye YA, Radaelli L, Singh VK., Pentland AS, Unique in the Shopping Mall: On the Re-identifiability of Credit Card Metadata, Science 347 (6221), 536-539