Presentation by the Privacy Commissioner, John Edwards to the Banking and Financial Services Law Association (BFSLA) on 29 August 2016

Customer data, customer trust

Financial organisations have two important assets that generally don't show up on their balance sheets: customer data and customer trust.

You can get a significant amount of commercial value from customer data.

You can deepen relationships with customers by providing targeted services.

You can reduce business risk by developing a better picture of a customer's ability to repay a loan.

In giving you this data, customers are making a significant investment. They are trusting you with their sensitive data. The way you use it will determine whether you maintain, increase or lose this trust.

Maintaining customer trust is just good business.

Trustworthy practices on an ongoing basis give you more credibility, or "capital" in the event of a breach or any other negative data incident.

By contrast, customers who don't trust a business will take their business elsewhere.

Today I want to talk about how financial organisations can extract value from large data sets while also maintaining, and perhaps even enhancing, customer trust.

Who owns the data?

First, I want to address a perennial issue in this space: who owns the data?

It's the wrong question. Sure, someone can own the medium. The hard drive, the paper stock, the USB, the ink, or the intellectual property, but this ownership does not determine or affect the rights and responsibilities of the subject and the holder of the information.

The Privacy Act gives people *control* over their information, and it gives agencies *responsibilities* when it comes to handling peoples' information.

For example, principle 6 gives people the right to see information agencies hold about them. In this context, it doesn't matter if they own it, the agencies own it or someone else owns it. The point is that they get to see it either way.

Big data

Now that we've dispensed with data ownership, we can move to data itself.

Organisations are trying to find value in big datasets. There has been some success, but it is very much a developing field.

It is certainly not as far along as some would have you believe.

I think of big data as a form of modern-day alchemy.

Data scientists are the alchemists, promising gold if we can just get bigger, better data sets.

Along with this mindset, we have a lot of confidence in data. It can become easy to rely on information from a dataset "because the data says so."

Cory Doctorow and the hypothetical algorithm

Canadian blogger Cory Doctorow spoke about algorithms on Radio NZ last month.

He argued that data is not as objective as we would like to believe. The models we build into our algorithms are built by people, and they rely on human assumptions.

He gave the example that you could write a program that estimates people's heights based on their weight.

You would build it by inputting a number of different people's heights and weights. The programme could then use that information to take an educated guess at someone's height for a given weight.

If the original sample was large enough and diverse enough, the programme would probably be reasonably accurate.

But what if the original sample was composed entirely of eight-year olds? The programme would quite confidently insist that everyone it meets is four feet tall.

Doctorow's hypothetical example shows that big data solutions are only as good as their inputs.

Here's a real-world example:

Boston Street Bump

In Boston, the local council used an app called Street Bump to track potholes and send repair crews to fix them.

Anyone could download the app to their smartphone, and it used sensors to detect pot holes.

The implicit assumption here was that smartphones were evenly distributed throughout the city.

That assumption was not correct. There are far fewer smartphones per capita in poorer communities.

This meant that the app detected more pot holes in middle-class and wealthy communities.

The data appeared to be directing crews to where they were most needed, but in fact it was entrenching existing social divisions and under-serving the poorer communities[1] .

[1] (http://www.reuters.com/article/us-usa-obama-privacy-idUSBREA3Q00M20140427

The point of this story is that the programme worked exactly as it was told to work. The problem stemmed from the human assumptions behind it.

Good data is good business

But why does this matter? If the assumptions behind your algorithms are giving you slightly skewed results, isn't this irrelevant, particularly if your competitors are also getting skewed results?

It is, in fact, very relevant.

For example, if you're using an algorithm to determine credit risk, and the algorithm's assumptions tell you someone is a greater risk than they actually are, you are missing an opportunity by not lending to them. The converse is also true.

The same applies to anything else you may use algorithms for.

Flawed assumptions create incorrect results, and incorrect results drive you away from making the best decisions.

Data Futures Partnership

So how can the commercial sector manage these risks?

They can start by looking to the Data Futures Partnership.

You may have heard of it before.

The Data Futures Partnership is a cross-sector group, established by the government, to help public and private sector organisations navigate data issues.

The DFP created a series of principles that should underpin any big data initiative. I support these principles. They are:

Value: data should drive economic and social value. In other words, any data initiative should have a clear purpose rather than a "fishing expedition."

Inclusion: the Partnership should support all New Zealanders. This includes those who work in financial institutions and their customers.

Trust: data management should build trust through transparency and openness. Privacy is a key component of this principle. You earn customer trust by being open about how you use their data, as well as being held to account when you get it wrong.

Control: individuals should know how their data is being used, and be able to opt out

I'll work through how some of these principles can work in practice.

Algorithmic transparency

I encourage organisations to be transparent with their algorithms so that people can see why decisions were made – and challenge those decisions if they think they're wrong.

I think that if you're going to make a judgement call about someone based on a data set, then that person has the right to see why you made that call.

If you don't allow this transparency, and instead rely on the data without questioning it, then you could miss opportunities, or end up breaking laws against discrimination.

This issue was flagged in the White House Report on Big Data, which was released in May of this year.

The report discussed big data in the context of providing credit. On one hand, it's a compelling premise: by collecting more data about potential borrowers, you can ostensibly make a more informed choice about someone's ability to pay you back.

This has the potential to open up access to credit to people who, while creditworthy, would not have appeared creditworthy under systems used in the past.

A hypothetical example was a programme that analyses patterns in someone's social media connections to create a credit score.

It's not hard to imagine this programme immediately writing off entire neighbourhoods or races as high credit risks, even though none of their behaviour or character traits are actually connected to their ability to repay loans.

This means an algorithm could result in a bank missing opportunities to lend to people who are perfectly sound credit risks.

Worse, it could put you on the wrong side of human rights law for discriminating based on race or ethnicity.

This is where algorithmic transparency can be helpful. If someone asks why they were denied a loan, they should be able to see the process that led you to make that decision.

If their complaint reveals that the algorithm isn't working, then you can tweak it.

Algorithmic transparency isn't enough on its own, though. You also need to proactively assess your algorithms and make sure they aren't jumping to erroneous conclusions.

If you don't do this, you could end up discriminating without realising it – as people aren't going to complain about your process if you never engage with them in the first place.

Credit reporting

Credit reporters are a good example of organisations that have built trust through transparency.

These organisations send regular assurance reports to my office, outlining a number of aspects of their processes. My office publishes these reports on our website.

We recently undertook a "mystery shopping" exercise, sending contractors to request their own credit reports and tell us how long they took.

We then published our findings.

The credit reporters took advantage of this opportunity, and issued media releases about improvements they had made to their processes. Their willingness to be transparent gained them positive media coverage. This coverage likely improved their reputation.

FATCA and transparency reporting

The credit reporters were transparent about how responsive they were to access requests for information. Businesses can also benefit from being transparent about who they share information with.

One example was the FATCA legislation that took effect in 2014. As you'll probably be aware, this legislation essentially forced banks to share information about the US person customers with US tax authorities.

Some customers were not pleased with this, but since banks were transparent about FATCA, they were able to reduce some of the fallout.

Imagine the converse – if banks were caught sharing this information without having informed their customers ahead of time. I expect the public reaction would have been much more vitriolic.

Indeed, you do not need to imagine the converse. A similar situation happened late last year, when Westpac was found to have shared Nicky Hager's data with police.

For those of you who are not familiar with this story, here is what happened.

In 2014, journalist Nicky Hager wrote a book called *Dirty Politics*. The source material came from a hacker, who got private Facebook messages, emails and other documents from a prominent blogger[2]. The material showed that the blogger had been working behind the scenes with lobbyists and PR professionals.[3]

After the book was published, police investigated the hack.

As part of their investigation, they asked a variety of organisations to share information they had about Nicky Hager[4].

The Privacy Act says you can disclose information when it's necessary to avoid prejudice to the maintenance of the law.

There has been discussion in the courts as to the extent to which that provides a power for Police to ask for information.

Most of the organisations did not comply with the Police request for information – with the exception of Westpac, who shared 10 months worth of transaction information.

In Westpac's defence, their terms and conditions are very clear about the fact that they will disclose customer information to police if Westpac believe that doing so will help it to comply with the law.[5]

[2] http://www.stuff.co.nz/business/industries/10417726/The-hacker-revealed
[3] http://www.stuff.co.nz/business/industries/10417726/The-hacker-revealed
[4] http://www.stuff.co.nz/national/73125717/nicky-hager-seeking-full-and-frank-disclosure-from-westpac-over-data-release

These terms and conditions are probably similar across all banks.

Further in their defence, they changed their processes after this incident to internally clarify how to deal with information requests like this[6].

Being transparent about how often you share information, and with whom, can help customers make an informed decision about who to trust.

I encourage businesses to be transparent about how many requests like these they receive, and how they respond to them.

This gives customers the ability to make an informed decision about who to do business with.

In order to encourage this, my office facilitated a transparency reporting pilot. We worked with 10 companies from a variety of industries between August and October last year.

Our goal was to produce a report on how businesses generally responded to requests for information.

The results: **11,799** requests for personal information, of which **11,349** were complied with and **449** declined.

We are currently doing another transparency pilot with different organisations – keep an eye on our website to see the results.

Transparency reporting like this has been endorsed by the International Conference of Data Protection and Privacy Commissioners in a 2015 resolution.

Some companies release their own transparency reports. Trademe, for example, released theirs last month. I encourage all businesses to be open about who they share customer information with.

Reidentification

I'll shift now to another issue that emerges from big data initiatives –reidentification.

Reidentification ties in with the DFP principle of "control."

Big data is generally predicated on the assumption that it's anonymous; people are identified only by behaviours or by their relation to others in aggregate.

However, it's often not difficult to reidentify purportedly anonymous datasets by cross-referencing them against other datasets.

As a crude example, if you have a list of people's names and addresses, and an "anonymous" list of addresses with house prices, it would be fairly easy to match the two datasets against one another and find out how much any one individual paid for his or her house.

---

[5] https://www.westpac.co.nz/assets/Who-we-are/About-Westpac-NZ/General-Terms-Conditions.pdf

[6] http://www.stuff.co.nz/national/73125717/nicky-hager-seeking-full-and-frank-disclosure-from-westpac-over-data-release

When you create a situation where people can be reidentified, you are potentially exposing them in ways that are incompatible with the Privacy Act – and risking a privacy breach for your organisation, in spite of your best efforts to anonymise information.

As a more realistic example, you can look to a research project from May of this year.

A researcher scraped data from 70,000 individuals on the dating website OKCupid, and published that information online for other researchers to use.

He didn't publish directly identifiable information such as their names or addresses, but he did publish information such as their age and location, along with more sensitive details like their sexual tastes.

The data was ostensibly "public" beforehand, but putting it in one place made it easier for people to reidentify previously-anonymous people.

It became possible to search the data set for attributes that you knew belonged to someone – say their age and suburb – and unearth all kinds of information they had never intended on sharing.

Reidentification can also come from metadata – that's the data about the data. It's not hard to connect data points such as timestamps and match them against behaviour to identify a person.

In my submission to the Data Futures Partnership, I advocated for a ban on reidentification. This is a matter of putting in proper controls to prevent reidentification.

There are a number of ways to do this.

For example, you can use "differential privacy," which Apple rolled out this year.

This tactic works by adding random "noise" into anonymous data, so you can see overall trends without being able to pinpoint any one person's behaviour or information.

This is a promising approach towards managing reidentification risk.

Naming policy

I'm not just here to offer some advice. There are real consequences for getting privacy wrong, which I will use as part of a regulatory response.

One relatively new consequence is my office's naming policy.

This is a policy we developed in 2014. This policy helps us hold agencies accountable for breaching privacy by publicly naming them.

We name agencies in a number of circumstances, such as when we suspect wider systemic problems or when an agency flagrantly disregards the law.

I don't need to tell you that finance is a competitive industry and the importance of reputation – so it's in your interests to avoid being named.

Referral to the Director of Human Rights Proceedings

When we receive complaints, we have the option of referring them to the Director of Human Rights Proceedings, who may take the case to the Human Rights Review Tribunal.

As with our naming policy, we reserve this option for the particularly notable cases.

The HRRT can award up to $200,000 in damages, and public judgements almost always name the respondent, which carries further reputational harm.

The damages awards from the Tribunal have been trending upwards for a couple years. They set a record in 2015 with $168,000 for the infamous “cake case.” So the stakes of going to the Tribunal are high – these aren’t parking tickets or slaps on the wrist.

Law reform

Parliament is set to reform the Privacy Act to give my office greater enforcement powers.

These include mandatory breach reporting, and the ability to fine agencies up to $10,000 for a variety of offences.

I’ll also be able to issue compliance notices that compel agencies to take an action, or to stop an ongoing action.

What you can do

I’ll finish off by outlining a few practical things you can do to maintain customer trust.

Whenever you implement a new programme that involves personal information, I recommend you undertake a privacy impact assessment. This is the process of flushing out any privacy risks, as well as finding ways to mitigate those risks.

There is guidance and online training on privacy impact assessments on my office’s website.

You can also instil a culture of privacy by making sure everyone in your organisation has a firm base of privacy knowledge. In order to help you with this, we’ve put together free online privacy training modules.

You can undertake this training at your own pace, from anywhere. There’s a Privacy101 module that gives a solid broad-level understanding of the Act.

Finally, you should check out our newest tool – “AskUs”.

These are interactive FAQs on our website. You should be able to find answers to most privacy questions by typing them into Ask s

If you can’t find an answer, there’s a field where you can let us know – please do so! It’s a living product that is constantly evolving.