

## **Privacy in the digital world**

**Privacy Commissioner John Edwards' presentation to**

**Supertech**

**– the Employers and Manufacturers Association EMA's Inaugural**

**Digital Technology Conference**

**17 March 2016, Auckland**

You've hear a lot already today about disruptive technologies.

They displace an established technology and shake up the industry or create a completely new industry.

History is a progression of how one disruptive technology displaces an existing one.

The personal computer displaced the typewriter.

Email disrupted snail mail.

Smart phones replaced pocket cameras, video and voice recorders, MP3 players, calculators and GPS devices.

Cloud computing displaced data hosting in organisations.

And so it goes - the Internet of Things, drone deliveries, Artificial Intelligence, robotics, driverless cars, virtual reality, augmented reality, facial recognition technology.

These are just a few of the technological trends that are changing the way we currently live and work.

Existing business models need to adapt as new businesses arrive to take advantage of emerging opportunities.

We are living Moore's Law.

The view from my office is that disruptive technologies are forces for good because they encourage and stimulate innovation.

### **Price versus value**

But we believe that in a world where the cheapest smartphone is being sold in India for four dollars, consumers need to know the hidden price they have to pay.

What's the value of the personal information they are trading away and the cost to their privacy?

The organisation that has collected the information will have a price for it but how does that price compare to the value to the individual?

In 2014, a Dutch student Shawn Buckles decided to illustrate the market for data by selling his personal information at an auction.

After a few weeks and 53 bids, his information sold for 350 euros - or about 570 New Zealand dollars.

Shawn Buckles' data bundle included all sorts of private information - every thing from online browsing data to email conversations.

He told a media organisation that he had read that a person's data could be bought for under a dollar. That's because organisations bought data in bundles and that made it cheap to do so.

He said he had added lots of value to his data, but it included his most intimate information - and there was no fair price for that.

Buckles hoped his auction would help people understand that the issue was about people.

His stunt was designed to illustrate what data was being collected and how private this data was.

"Our digital data says more about us than our living rooms. A question to you: do you have curtains?" he asked.

### **Trading health information**

Take this recent example out of the United States where employers can employ 'wellness firms' to monitor their employees' health.

Imagine the wellness firm and insurers working together with the employer to mine the health data of the employees.

Should, for example, the employer know what prescription medicines you are taking?

While employers don't get access to which individuals are flagged by this data mining, they do receive aggregated data on the number of workers at risk for particular health risks or conditions.

This presents a real risk of re-identifying individuals from the anonymised data - but more on that later.

In the meantime, while wellness firms, insurers, and employers all extract benefit from the data, what's in it for the employees whose information is being mined?

Who benefits the most?

### **Trouble with Uber**

Uber is a topical example.

It's a disruptive business model that challenges the traditional taxi industry.

Since it was founded in 2009, Uber has grown spectacularly but as a brand it has been clouded by a number of privacy controversies.

Through its app on a customer's phone, Uber is able to track a customer's movements - a feature that was known within the company as 'God's view'.

God's view shows an aerial view of all customers and drivers in a particular area.

One Uber executive bragged the company - if it wanted to - could use its tracking ability, to dig up dirt on the personal lives of journalists who were critical of the company's business practices.

In one incident, the company showed a journalist how it tracked her on her way to interview another top Uber executive.

There was also the notorious incident in which a well known businessman was told by an acquaintance that his movements in an Uber taxi in New York were being shown in real time on a big screen at an Uber launch party in Chicago.

One Uber job candidate revealed he was given unrestricted access to Uber's customer tracking function during the job interview, access he was able to keep for hours after the interview ended.

And in February last year, Uber admitted to a data breach in which the names and licence plate information of 50,000 of its drivers was disclosed.

Perplexingly, Uber waited more than five months before it notified the drivers affected and publicly revealed the error.

The name of the businessman who suffered the privacy breach at the Uber party in Chicago is Peter Sims. On his blog, he asked 'Can we trust Uber?'

He observed: "A great, long-lived brand begins and ends with trust."

Uber has since been addressing its privacy failings to placate US regulators.

As part of a planned settlement with the New York Attorney General, Uber agreed to make changes to its privacy and security practices.

These include password-protection and encryption for the location data of passengers and drivers, limiting employee access to the data, and adding more security tools to protect personal information.

Uber has also agreed to notify the Attorney General's office if it started to collect GPS data from customers' smartphones when they were not using the app - something the company claims it doesn't do.

### **Trust and confidence**

For the private sector, collecting and using personal information can give you a competitive edge.

But we've also seen what happens when large numbers of people lose trust and confidence in, for example, the banking system.

In a very recent example, I was asked to comment on one of the unfortunate outcomes of the collapse of the Dick Smith Electronics retail chain.

The liquidators were planning to put the Dick Smith customer data base up for sale and our office received a number of enquiries from concerned customers.

They were worried about the possibility that information about them would pass from a retail chain that they had entrusted their information to an unknown third party they did not know or trust.

I put it to you that respect for personal information is necessary for the maintenance of your customers' trust and confidence.

Here's another thing - disruptive technologies have created new ways of easily collecting huge amounts of customer information.

But just because you can collect massive quantities of data, does it mean you should?

The answer is no - absolutely not.

The Privacy Act actually prohibits this. You're only allowed to collect what you need, not what you might or might not find a use for in the future.

There's a huge temptation to collect everything you can - the more data points you can string together, the more useful conclusions you are likely to draw.

But when you collect information from somebody in New Zealand, you need to be able to tell them why you're doing it.

"Keep it because one day we'll figure out a use for it" as a justification doesn't cut it.

If you cannot explain how a piece of information is related to the service or product you're providing then you shouldn't collect it.

That's why it's important to be clear on what you want to do before you decide what you need to know.

The Privacy Act sets out the obligation to inform people about the information you are collecting, but you need to make sure you're working towards informing users, not just as an exercise in legal compliance.

When people try and comply with the Privacy Act, there's a tendency to go heavy on the legalese.

We don't think the people should have to have a law degree to read the terms of service.

We believe you have the responsibility to speak to people in language that they understand. To ensure that when they do consent, it's informed.

Privacy is part of the landscape, and your customers and clients are getting more and more privacy literate.

If you treat them with respect, they'll be more likely to trust you.

### **Public opinion**

Our office commissions UMR to undertake a public opinion survey for us every two years.

UMR is currently undertaking the latest survey - the results of which will be out in early May during Privacy Week.

The last one is still pertinent because it maps a long term trend.

In that survey, half of all New Zealanders polled said they were becoming 'more concerned' about privacy issues.

That was the highest level yet recorded for us in that tracking survey - up from 40 percent in the 2012 survey.

81 percent of respondents said they were concerned about businesses sharing information with other businesses without permission

67 percent said they were concerned about government agencies sharing information with other agencies without permission

More importantly for the big data context - 37 percent of the survey respondents said they didn't feel in control of the way business use their information.

### **Government and big data**

In 2013, the government introduced its Integrated Data Infrastructure, a project led by Statistics New Zealand.

The project combines information from a range of government organisations to provide the insights into how the government can improve social and economic outcomes for New Zealanders.

With all personal information removed, integrated data gives a safe view across government.

The project is particularly useful to policy makers and social researchers working on addressing complex social issues such as crime and vulnerable children.

For example, last month, the government released a data visualisation tool based on the Integrated Data Infrastructure.

By using it, you can see how many vulnerable children there are in your region or in your neighbourhood.

I'm on record as describing the Integrated Data Infrastructure as a 'grand bargain' in terms of how privacy and confidentiality are not compromised.

The IDI promises that personal information is never seen by researchers. This is done by anonymising the information and removing identifying information like names, addresses, and dates of birth.

The research findings are then grouped in a way which makes it impossible to identify individual people.

### **Big data overseas**

In 2012, the Obama administration in the US announced something similar - its Big Data Research and Development Initiative.

The purpose was to improve government's ability to extract insights from various sources of data and make better decisions to support national security objectives, scientific research or drive economic growth.

This type of analysis can save taxpayers money, strengthen public trust and increase efficiency by providing timely and needed interventions.

In the private sector, information mined from transactions can be used to analyse population demographics and calculate consumer purchasing habits, credit risks and predict consumer trends.

In both the commercial and government arenas, the field of data analytics is relatively new and has room for growth.

As the world's data production doubles every two years, the ability to store, analyse and share data is a key to research and development investment.

In Britain, the government identified in its House of Commons Science and Technology Committee report - The Big Data Dilemma - that big data was one of the 'eight great technologies'.

The country's Centre for Economic and Business Research estimated in 2012 that big data could create 58,000 new jobs over the following five years and contribute 2.3 percent of GDP.

While the lure of big data is seemingly irresistible, the same House of Commons report also observed the benefits of big data have to be weighed against a potential loss of privacy and the risks of our data being lost or misused.

"Given the scale and pace of data gathering and sharing, distrust arising from concerns about privacy and security are often well founded and must be resolved by industry and government if the full value of big data is to be realised," the report said.

### **Re-identification**

One of biggest concerns about big data is techniques that can re-identify individuals when previously anonymised data are combined with other data sets.

The Privacy Act says organisations should collect personal information for a stated lawful purpose.

Organisations can't use personal information collected for one purpose for another purpose.

The way around this is by anonymising or de-identifying the information so that it is stripped of information about identifiable individuals.

In theory, anonymised or de-identified data is no longer personal information and therefore not subject to the Privacy Act.

But does de-identifying data as a fail safe to protecting personal information work?

It depends on what you mean.

One British NGO observed that as data sets become more sophisticated, the technical possibility of undertaking 'jigsaw' re-identification of individuals increases.

Jigsaw re-identification is the ability to identify people using two or more different pieces of information from two or more sources of information.

A famous example of the dangers of anonymising personal information was demonstrated by an American researcher, Dr Latanya Sweeney.

In the mid 1990s, the Massachusetts Group Insurance Commission decided to release anonymised health data on state employees.

The aim was to improve healthcare by giving researchers access to the data.

In an effort to retain anonymity, name, address, social security number and other obvious identifiers were removed from the data.

The public were assured that these steps were sufficient to protect their privacy by the Governor of Massachusetts at the time, William Weld.

Latanya Sweeney was a graduate student at MIT, interested in anonymity and studying computer science.

She requested a copy of the data, and got to work.

She knew that Governor Weld lived in Cambridge, Massachusetts, a city of 54 thousand residents and seven postal codes.

By combining the dataset with publicly available information she was able to identify which records related to Governor Weld, and she arranged for details of his health records, including diagnoses, prescriptions and details of hospital visits, to be delivered to his office.

There are many other examples. The Care.Data initiative in the UK was derailed last year because of concerns about the robustness of the deidentification. Another study showed that in an anonymised dataset of location data over a year of 1.5 million telecommunications company subscribers, researchers could identify individual records with 85% accuracy, if they knew where they had been just four times in than year.

These cases reveal that our intuitive beliefs about anonymity and identity are often misplaced.

The distinction between personal and non-personal information is becoming increasingly blurred.

Rapid advancements in technical capability exacerbate that blurriness even more and raise ethical issues in relation to informed consent, trust and what security of information means and how best to secure it.

These ethical issues are currently being hotly debated among lead commentators in the area of re-identification.

Even privacy advocates are divided on the issue. Dr Ann Cavoukian - the former Ontario Privacy Commissioner recently published a paper suggesting that re-identification isn't as much a concern as Latanya Sweeney led others to believe.

She argued that concerns about re-identification are hindering progress and said re-identification 'myths' are putting data custodians off the sharing of data sets, and that some aren't even bothering to de-identify in the first place.

What we do know, and is widely accepted is that there will always be a risk that some of the records can be re-identified.

### **Regulating open data**

The New Zealand's government's efforts to provide open data have excluded personal information so far, although as we've just discussed - what constitutes personal information is not necessarily well understood.

While our office plays a role in setting the limits on what can be done with personal information, my role does not include a mandate to promote the wider use of data.

That role is not, as yet, part of any organisation's mandate.

There is also an important distinction between privacy and ethics in data use and there is no body charged with providing advice on what constitutes ethical practice - apart from the National Ethics Advisory Council's role in the health sector.

A case which you may have heard about recently involved Facebook admitting to manipulating nearly 700,000 of its users' news feeds to see whether it would affect their emotions.

The social network was roundly condemned by many social scientists for breaching ethical guidelines for informed consent.

### **New Zealand's data future**

We are fortunate in New Zealand that we already have a robust legal framework for managing data with the Statistics Act and the Privacy Act as the pillars of that framework.

The work of the Data Futures Partnership builds on these existing foundations of data management and seeks to unlock the opportunities of large scale data collection while protecting the privacy rights of the individual.

The Partnership's vision identifies four key elements to be maintained as part of the framework, and these are value, inclusion, trust and control.

It is encouraging to see that the Data Futures Partnership recognises that privacy is an integral part of the national conversation about our use of data.

The Privacy Act deals particularly well with the desire to realise value from data from recombination and reuse.

But there are areas where the Act can be strengthened.

### **Strengthening the Privacy Act**

For example, under the Act, there is currently no explicit prohibition on the re-identification of data from which identifying information has been removed.

It's food for thought that a prohibition of this nature could potentially increase public confidence in the safe use of "de-identified" or "anonymised" data.

Similarly, further work could be undertaken on strengthening individual rights to have information about them deleted.

This would raise people's confidence that information provided is not necessarily available forever and able to be combined with yet to be created datasets.

### **In conclusion**

Big data can be tackled with many of the same tools we use for little data.

Our Office helps businesses build their understanding of privacy with tools like our Privacy Impact Assessment guidance and encouraging Privacy by Design.

Privacy Impact Assessments focus on identifying the ways a new proposal or operating system, or changes to an existing process may affect personal privacy.

They help organisations make more informed decisions and better manage privacy risks.

It is important to decide whether to do a Privacy Impact Assessment early in a proposal's life.

When businesses introduce an awareness of privacy to projects early on, many of the major risks can be mitigated. This is what is meant by Privacy by Design - building privacy protections in on the ground floor.

If you fail to identify how your project is likely to affect the individuals whose information you are collecting and using, there are real risks for your organisation and for the success of your project.

You can find our Privacy Impact Assessment guidance on our website.

We also have online privacy training modules which are free for anyone to use. This is a good way to create a better awareness in your workplaces of how to look after your customers' personal information.

And as marketers, you'll know how important it is to communicate directly with people. We developed a tool to help with that – it's an online privacy statement generator, Priv-o-matic.

This is an online gadget that will generate the words you can use on your website that tells people what you do with their information. It takes about five minutes to complete.

These 'principle 3' statements are minimal compliance statements that you need to show people when you collect their personal information.

One of the key goals we set for ourselves was to make privacy easy for business. When we talked to people about our technology strategy, they asked for simple tools and this is something we are keen keeping building on.

Having good privacy protections and practices in place help you avoid dealing with investigators from our office, the cost and emotional drain of litigation, the substantial delays inherent in the court process, adverse and costly Human Rights Review Tribunal decisions and negative news coverage.

Privacy matters because your customers still care about it.

Privacy matters because your reputations depend on it.

The organisations that you represent have an obligation and a responsibility to manage and protect personal information.

Our lives are becoming ever more enumerated and dissected.

If people can't trust the organisations they do business with to look after information that is about them, they will look for competitors they can trust.

Big data, the Internet of Things, Moore's Law - all these developments increase rather than detract from a focus on privacy.

The public demand for restraint will be met both in the marketplace and in regulation.

ENDS