THE FUTURE OF DATA

BFSLA Conference 2016

Dr Andrew Butler[1]

INTRODUCTION

1. I'm the third member of the trio presenting this morning. I've been asked to cover privacy topics from the point of view of a private practitioner who regularly advises businesses on these matters. The topics I thought I could usefully cover are:

   (a) Some key "big themes" about technology and its privacy implications that all advisors need to have at the front of their minds.

   (b) How some of the Information Privacy Principles play out in the big data field, particularly in relation to:[2]

      (i) informed consent;

      (ii) anonymisation; and

      (iii) disclosure of information to third parties.

   (c) Specific issues relating to use of "The Cloud" and "Big Data".

KEY THEMES

Trust

2. From the outset, I want to affirm what John has said regarding the importance of trust.

3. Data is a currency that will only remain usable so long as society tolerates that use. In that sense, it is subject to a "social licence". The core to maintaining that "social licence" is trust.

4. From a practical perspective, it is important to note that legal exposure may not be co-extensive with "trust exposure".

   (a) Thinking of the examples John has spoken on, activities such as analytics, algorithms etc that are, at least strictly speaking, legal and not in breach of the Information Privacy Principles, may nonetheless erode the trust of clients and customers.

   (b) Another reason why trust is hugely important here is that privacy policies are often lumped together with, and treated the same by customers, as the proverbial "terms and conditions", which are not famed for how widely they are read. This is particularly an issue with apps on smartphones, all of which have

---

1 Partner, Russell McVeagh; Convenor of the New Zealand Law Society Human Rights and Privacy Committee; editor, Human Rights Reports of New Zealand. Thanks to Tim Morrison and Andrew Pullar for their assistance preparing this paper. The opinions expressed in this paper are my own. This paper is a draft and, as such, is not fully referenced. The paper should not be cited without the author's permission until it is finalised.

2 All references in this paper are to New Zealand law, specifically the Information Privacy Principles under the Privacy Act 1993, unless otherwise specified.

different privacy policies, all of which individuals are meant to read on small screens.

5. The ideal is for individuals to be able to make informed decisions about how their personal information is collected, used and disclosed. **The best way to gain and maintain trust is by ensuring meaningful informed consent, and being targeted in the choice of information that is collected, how long it is held for, and how and when it is used and disclosed. But I acknowledge there is a tension here because some within the business will want to future proof the use of data in innovative and unanticipated ways; confining consent to today's uses can cut across the use of that data in new ways in the future.**

6. Another perspective worth considering is that privacy, and trust, is not just a burden for businesses: it can be a **competitive advantage**. Security of data storage, favourable terms and conditions, strong privacy protections and practices can set an organisation apart as more desirable or deserving of customers' confidence in the security of their personal information.

Bargain / exchange

7. Another point worth noting is that the quick uptake of technology, notwithstanding the enhanced ability for transactions, locations, personal connections and the like to be surveilled, has been facilitated in large part by customers accepting implicitly, if not explicitly, that they are engaged in an exchange or bargain. Put another way, they are compensated for giving up the privacy to which they are otherwise entitled. But if there is no movement of value in their direction customers will not be so happy to give up their privacy entitlement.

Privacy and Technology

8. The other theme I need to touch on is how privacy "manifests" in business. How do we interact with privacy issues?

9. To that end, it is important to recall that different technologies:

   (a) give potential for access to different kinds of information; and

   (b) raise different legal issues and engage different Information Privacy Principles.

10. Stepping through some examples:

   (a) **Eftpos / sales:** Potential to gather data on the nature of purchases, where and when purchases are made, how much is spent on particular kinds of products at those times and in those places, etc.

   (b) **Apps:** Potential to gather data on a user's location through "geotracking", data from user analytics, for example the time of day particular apps are accessed and for how long, device identifiers and IP addresses of devices and users.

   (c) **Cloud:** Raises issues not so much on what data is collected or how data is collected, so much as how it is used and disclosed. As a business, using "Cloud" providers for data storage may engage jurisdictional issues, if the provider company's servers are outside New Zealand. Other issues include whether the provider company uses the personal information stored for its own purposes, and if affected individuals are informed of this use; and how businesses can best

ensure compliance with their Privacy Act obligations once the information is stored with another party.

(d) **Big data sets:** Again, "Big Data" sets raise issues of how data is used and disclosed more than as to how it is collected. Issues engaged by use of "Big Data" sets include:

(i) How many other users of the big data set are there? Have the individuals whose information is stored with the big data set consented to this kind of disclosure?

(ii) For what purposes is the data used? Are the relevant individuals informed of this kind of use?

(iii) Is the data anonymised? If so, how well? John has spoken about the risk of re-identification, to which I will return.

(iv) Again, how are businesses best able to ensure compliance with obligations to clients whose personal information you hold?

(e) **Electronic messages:** These engage issues relating to the Unsolicited Electronic Messages Act 2007. This legislation provides:

(i) Section 9: unsolicited commercial electronic messages must not be sent.

(ii) Section 15: exposes third parties to liability, where they have aided, abetted, etc a breach of the Act.

Issues engaged by this legislation include, have recipients of electronic messages consented to being sent such messages? Does the business provide an effective unsubscribe facility? The Act creates exposure to damages where direct or consequential loss or damage can be made out from a breach.

(f) **Website:** Websites engage issues on how personal information is collected: aside from information actively shared by users, is your website collecting information that is passively shared, by the fact of visiting the website? This information could include cookies, analytics, device identifiers (IP addresses), browser history, bookmarks and social media information. Are individuals informed of this and of the uses to which this information will be put when they visit the website?

11. Questions businesses need to be asking when using any form of technology include: what personal information is the business collecting? Do individuals know this information is being collected when they interact with the business? Do they consent to this collection? Is collection of this nature reasonable and necessary for the operation of the business? How is the information stored? Can individuals' information be retrieved and provided? Are steps taken to ensure the information remains accurate? How is the information used and disclosed, and do individuals consent to this use and disclosure?

INFORMATION PRIVACY PRINCIPLES

Informed consent

12. The key to maintaining trust, as well as effective compliance with the Information Privacy Principles, is ensuring individuals give informed consent to the collection, use and

disclosure of their personal information. At the core of the Information Privacy Principles is that control of personal information is a human right; and therefore the emphasis is on the ability of everyone to control how information about them is collected, generated, and used. It is, in short, theirs to control.

13. The Information Privacy Principles require (subject to exceptions) that individuals be informed of the collection of their personal information, including what information is being collected, the purposes for which it is collected, intended recipients of the information, and who will hold the information.

    (a) Principle 2: Provides that where an agency collects personal information, the agency shall (generally) collect the information directly from the individual concerned.

    (b) Principle 3: Provides that where an agency collects personal information from an individual, the agency shall take reasonable steps to ensure the individual is informed of the fact the information is being collected, the intended purposes for which the information is collected, third parties to whom information may be disclosed, the agency that collects the information and the agency that will hold the information.

    (c) Principle 4: Provides that personal information shall not be collected by an agency by unlawful means, or means which are unfair or unreasonably intrude on the personal affairs of the individual concerned.

14. A key aspect of Principle 3 is that when disclosing personal information, including to cloud providers or to big data sets, the individuals whose information is disclosed must be informed of the disclosure, the purpose of disclosure, and any use the third party may make of their information.

15. The need for meaningful informed consent is a result of the fact that information is meant to be collected from the individual (Principle 2), and the fact that exceptions to that requirement include whether an individual has authorised particular use or disclosure of their information. I acknowledged that Principle 4 prohibits collection of information by unlawful means, or means which are, in the circumstances, unfair or unreasonably intrude on the personal affairs of the individual. Between this principle, and the exceptions in Principle 2, there may not be an *absolute* requirement to obtain informed consent. However, considering the extent of the obligations to keep individuals fully informed, and the aforementioned risk of "trust exposure" beyond legal exposure, it may not be prudent not to do so.

Anonymisation

16. An exception to the Information Privacy Principles relating to informed consent for collection, use and disclosure of information if the information will not be used in a form in which the individual concerned is identified.

17. Effective anonymisation of personal information can mean that the information is no longer "personal information". In that case, the Privacy Act and Information Privacy Principles do not apply. This exception comes from:

    (a) the definition of "personal information" as "information about an identifiable individual";

    (b) the exception in several of the Information Privacy Principles that an agency need not comply with the principle where the agency believes, on reasonable

grounds, that the information will not be used in a form in which the individual concerned is identified. This exception is an exception to the following principles:

(i) Principle 2, an agency is to collect information directly from the individual concerned;

(ii) Principle 3, an agency is to take reasonable steps to ensure individuals are aware of the fact their information is being collected, the purpose of collection, intended disclosure etc;

(iii) Principle 10, an agency that holds personal information obtained in connection with a purpose shall not use the information for another purpose; and

(iv) Principle 11, an agency that holds personal information shall not disclose the information to another person or agency.

18. However, extreme care is needed when dealing with anonymised information to ensure the anonymisation is effective:

(a) John has already discussed the risks of re-identification and the damage this can do to public trust.

(b) Information is still personal information if the individual can be identified when that data is considered in combination with other data.

(i) The exceptions to the Information Privacy Principles are phrased: the agency must do [x] unless the agency believes, on reasonable grounds, that the information "is to be used in a form in which the individual concerned is not identified".

(ii) If the individual can be identified from some of their personal information in the context of other information, this is not met.

(iii) Anonymisation needs to go further than just removing names: for example, the 16 year old male resident at 4 Privet Drive is clearly identifiable, as is the female resident at 10 Downing Street!

(c) Australian Information Commissioner Tim Pilgrim has cautioned the business community about not under coding de-identification. While he is open to considering de-identification data outside the strict protection of the Commonwealth Privacy Act, he will "only [do so] when the ID stripping process meets the highest standards".[3]

19. Again, a reminder that both legal and trust exposure are at stake, if apparently "anonymised" information, is "deanonymised"; particularly if individuals have only provided their personal information on the basis that they will remain anonymous.

Third parties

20. As outlined previously, when disclosing personal information to third parties, businesses need to keep the clients or customers, whose information is disclosed, fully informed, including as to the purpose or use third parties will make of their information. That is required by Principle 3.

---

3 P Cowan, "Pilgrim warns data de-identification is 'rocket science' ", IT News 20 April 2016.

21. Security of the information disclosed is a further concern. Principle 5 requires agencies, if it is necessary for personal information to be disclosed to a third party, to do everything reasonably within their power to ensure there will be no unauthorised use or disclosure of the information by that third party.

22. A further question is who is responsible for the information? Section 3(4) of the Privacy Act provides that where an agency holds information solely as an agent, for the purposes of safe custody or for the sole purpose of processing information on behalf of another agency, and does not use or disclose the information for its own purposes, the information shall be deemed to be held by the agency on whose behalf the information is held. This could apply to users of cloud services or big data sets, so that they remain responsible for the information stored.

23. Disclosure of information to third parties may also involve jurisdictional issues if third parties, to whom information is disclosed, operate outside New Zealand; particularly so if the applicable privacy laws are not the same as those operating on the agency here. This is a particular issue for use of cloud storage providers, discussed further later.

24. Section 10 of the Privacy Act provides that a number of the Information Privacy Principles apply to information held outside New Zealand, including Principle 5 (storage and security of information), Principle 6 (right of access to information), Principle 7 (right of correction), Principle 8 (agency to check accuracy of information), Principle 9 (agencies to keep information no longer than necessary), Principle 10 (limits on use of information) and Principle 11 (limits on disclosure of information).

25. Intellectual property and ownership of data is a further issue. It is important for business to determine, by contract, issues such as which party owns the data supplied, which party owns any information created and derived from it, and the purposes, if any, for which the party receiving the data is able to use it.

26. Third parties are also subject to privacy obligations under the Information Privacy Principles. In respect of cloud providers and big data sets, principles on collection of information would not apply insofar as another agency uploads the personal information it has collected onto the third party's system. However, these entities would be subject to other Information Privacy Principles by virtue of the fact they hold personal information. This would include:

    (a) Principle 5, relating to storage and security of information, applies to agencies that hold personal information, which would capture cloud providers / data sets. That means they are to ensure the information they hold is protected by such security safeguards as it is reasonable in the circumstances to take.

    (b) Again because they hold information, such entities are subject to Principles 6 and 7 relating to access and correction of personal information.

    (c) Principles on the use of information may also apply to entities that hold information, including:

        (i) Principle 8, which provides that an agency must ensure the personal information held is accurate before use;

        (ii) Principle 9, which provides that an agency is not to keep the information for longer than necessary;

        (iii) Principle 10, which prohibits use of personal information for purposes other than the purpose in connection with which the information was

obtained (though note exclusions including where information is anonymised);

(iv) Principle 11, which prohibits disclosure of personal information (exceptions include where disclosure is one of the purposes in connection with which the information was obtained or is directly related to those purposes, and where the information is anonymised).

27. Third parties such as cloud providers and data sets may therefore seek **contractual indemnities** from agencies that store information with them or use their services, against the risk of breaching these Principles. This is because third parties will often be relying on the agency that supplies the information to warrant that they collected the information in compliance with the Information Privacy Principles, such that the information is accurate, that the relevant individuals were notified of the disclosure to the third party and of the third party's intended uses of the information.

CASE LAW

28. So how do these Information Privacy Principles play out in practice?

***Vidal-Hall v Google* [2015] EWCA Civ 311, [2015] 3 WLR 409: Anonymisation**

29. Google placed a cookie on devices that run the Safari browser, allowing Google to collect browser-generated information that was then fed to a service which delivered advertising to consumers based on their online behaviour.

30. Google had stated it would not collect information from Safari users without their express consent. Various individuals asserted the delivery of the ads had caused them distress and three filed proceedings against Google in the UK courts under the Data Protection Act 1998 (UK).

31. Strict offshore service requirements meant the claimants had to show their claim had a serious chance of succeeding. Here, to make out a breach of the Data Protection Act, they needed to be able to prove the information in question is "personal data".

32. Considering that question, the Court said identifiability hinges on whether a person can be distinguished from other members of a group, notwithstanding that their name is not attached to the information. Here, it was arguable that the browser-generated information was "personal information" because:

(a) The cookie ascribed a unique ID code to the device used, allowing tracking of websites visited, time visited, time spent on the site etc.

(b) Devices such as smartphones and tablets are generally used by single users, so identifying a device effectively identifies an individual user (and therefore tracks their activities, not merely the activities of the device).

(c) Google's own business model is based on the ability to individuate users and target advertising to them based on their personal interests.

(d) The fact the data controller may not use the information to identify an individual is immaterial: the issue is whether a data controller can do so if it wishes (for example, by linking an IP address to a Google account). The answer here was that it is possible.

(e) Targeted advertising itself revealed personal information on the user's browser history. A third party using the device could discover information about the user by being exposed to that advertising.

33. Google appealed the decision to the UK Supreme Court, but the appeal was withdrawn following agreement between the parties to it.[4]

34. This decision illustrates how careful agencies need to be on whether data is effectively anonymised. In particular, it highlights that identifying a device may be equivalent to identifying an individual.

***Grubb v Telstra Corp Ltd* [2015] AICmr 35 (Australian Information Commissioner) and [2015] AATA 991 (Administrative Appeals Tribunal): Anonymisation**

35. Grubb is a technology journalist who asked his mobile phone provider, Telstra, for access to his telecommunications metadata. Mr Grubb's request was expressed as follows:

> I'd like to request all of the metadata information Telstra has stored about my mobile phone service. ... The metadata would likely include which cell tower I'm connected to at any given time, the mobile phone number of a text I have received and the time it was received, who is calling and who I've called and so on. I assume longitude and latitude provisions would be stored too. This is the type of data I would like to receive.
>
> Handing over RAW data would probably be easiest but if its in a CSV format that'd be great.

36. Telstra agreed to provide outbound mobile call details and the length of his data usage sessions via online billing, but refused to provide information regarding details and location of the numbers that called and sent messages to his mobile phone. Telstra advised Mr Grubb he would need a subpoena for the other information he had requested.

37. Mr Grubb complained to the Australian Information Commissioner. The evidence showed that Telstra was able to retrieve data, link it to an individual, and provide the resulting information to law enforcement agencies on request; and did so approximately 85,000 times in the year 1 July 2013 to 30 June 2014.

38. The Commissioner made a formal determination that Mr Grubb was entitled to have access to much of the information requested. The Commissioner:

(a) did not accept arguments that Mr Grubb's identity was not reasonably ascertainable from the metadata;

(b) did not accept that the information was about the network, not him; and

(c) accepted that it was possible to cross-match to link network data to an individual, so that his identity could be "reasonably ascertained", noting that Telstra did this routinely for police.

39. The Administrative Appeals Tribunal overturned the Commissioner's determination, finding the information was about the network and how it operated, not about Mr Grubb. The Tribunal also considered that Mr Grubb's IP address was not information about him.

40. The Commissioner has now appealed to the Federal Court. The appeal was heard by the Full Bench of that Court on Monday 23 August this year.

---

4 www.carter-ruck.com/blog/read/vidal-hall-v-google-goes-to-the-supreme-court (update 1 July 2016).

41. These upcoming appeals have significant implications. If both appeals follow the decision appealed from, New Zealand will have two divergent lines of authority from which it will need to choose.

Case C-362/14 *Schrems v Data Protection Commissioner* EU:C:2015:627, [2016] QB 527 (CJEU): Offshore disclosure

42. Austrian citizen, Mr Schrems, was a user of Facebook. Some or all of the data provided by Mr Schrems to Facebook was transferred from Facebook's Irish subsidiary to servers located in the US. Mr Schrems lodged a complaint with the Irish Data Protection Commissioner on the basis that, in light of the Snowden revelations as to the activities of US intelligence agencies, the law and practice of the US did not offer sufficient protection against surveillance of the data transferred to that country such that transfer of the data was in breach of the European Union's Data Protection Directive.

43. The EU Data Protection Directive provides the transfer of personal data to a third country may only take place if that third country ensures an adequate level of protection of the data.

44. The European Commission had decided in 2000 that the US ensures an adequate level of protection of personal data transferred: the "Safe Harbour Decision". This allowed US companies to self-certify that they would comply with EU data protection standards, in order to allow for transfer of European data to the US.

45. The CJEU declared the Safe Harbour Decision invalid. This was for several reasons:

    (a) it did not prevent interference by US public authorities with privacy (US public authorities were not subject to the Safe Harbour Decision);

    (b) legislation permitting public authorities to have access on a generalised basis to the content of electronic communications must be regarded as compromising the essence of the fundamental right to respect for private life (because it was not limited to uses strictly necessary); and

    (c) legislation not providing for any possibility for individuals to pursue legal remedies in order to access personal data relating to himself or herself, or to obtain rectification or erasure of such data, compromises the essence of the fundamental right to effective judicial protection.

46. The CJEU required the Irish Data Protection Commissioner to examine Mr Schrems' complaint and make a decision whether, pursuant to the Directive, transfer of the data of Facebook's European subscribers to the US should be suspended on the ground the US does not afford an adequate level of protection.

47. The decision is based on particular requirements of EU Data Protection Directive, but nonetheless raises interesting concerns:

    (a) Even a developed state like the US may not have the most outstanding data protection in the world.

    (b) Inadequacy of data and privacy protection can be a disadvantage: the flipside is that strong privacy protection can be a competitive advantage. Users of cloud services, for example, can "shop around" for the best terms in the best jurisdictions, and use a relationship with a reputably secure cloud provider as a competitive advantage in turn.

(c) Potential implications for New Zealand: As a member of the Five Eyes, if a similar ruling is made in Europe about New Zealand's data protection this could have significant implications for New Zealand businesses doing business with European entities.

Hammond v Credit Union Baywide **[2015] NZHRRT 6: Disclosure of personal information and damages**

48. In 2012, Ms Hammond posted on her private Facebook page a photo of a birthday cake she had made for a friend. Both were former employees of Credit Union Baywide (NZCU Baywide). The cake had been iced with the words "NZCU F**K YOU". Baywide "unreasonably pressured" a junior employee, who was a Facebook friend of Ms Hammond's, to log into her account so that Baywide could obtain a screenshot of the photo. Baywide then disseminated the photo to its staff, recruitment agencies, and others, warning them not to employ Ms Hammond. Baywide also contacted Ms Hammond's employer, encouraging them to fire her and saying that it would not process referrals from Ms Hammond, which would have financially crippled that company.

49. Ms Hammond alleged that Baywide had breached Principles 1 to 4 on collection of personal information, and Principle 11 regarding disclosure of information, and that these breaches resulted in actionable harm including significant humiliation, loss of dignity, and injury to feelings (under s 66 of the Privacy Act).

50. The Human Rights Review Tribunal did not determine whether Principles 1 to 4 had been breached, because such breaches would not have resulted in actionable harm under s 66, but concluded that Baywide had breached Principle 11 through its dissemination of Ms Hammond's personal information.

51. Baywide apologised, but the Tribunal considered this was "mechanical" and insincere. The Tribunal held the breaches caused harm under s 66, because Baywide intended to ensure Ms Hammond's employment would be terminated and that she would be unable to find employment in the region, such that Ms Hammond suffered economic loss, as well as significant humiliation, loss of dignity and injury to her feelings.

52. Ms Hammond was awarded a total of $168,000 in damages. Notably, $98,000 of that was for humiliation, loss of dignity and injury to feelings. The Tribunal noted compensation for that type of harm could be placed in three bands, the most serious of which would lead to compensation over $50,000 (in which Ms Hammond's claim fell). The Tribunal also expressly disavowed that punishment or disapproval was incorporated into the award. This set a new standard for damages under the Privacy Act.

THE CLOUD

53. Use of a cloud computing platform involves disclosing personal information to a third party (the cloud provider). Businesses should consider contractual terms to ensure:

(a) you have your clients' consent for placing their information on the cloud, and for any use the cloud provider may make of their information;

(b) you are making all reasonable efforts to ensure the security of the information when held by the cloud provider, so that hackers and cyber criminals cannot access client data when placed on the cloud;

(c) your data is not retained on the provider's servers once you have ceased using their services.

54. As to jurisdictional issues, as *Schrems* illustrates, some quite detailed considerations can arise. The key driver is the need to ensure data is not kept in a location that may lead to the privacy of clients' information being compromised:

    (a) Is the privacy law of the jurisdiction where the information is held similar to that of New Zealand? *Schrems* illustrates that EU privacy protection is relatively strong, while that in the US, comparatively speaking, is weaker and subject to US security and intelligence legislation. What about the situation in other jurisdictions? How do you make that assessment initially? How quickly can you terminate if regulators' views of that jurisdiction turn sour?

    (b) Does the data remain accessible to the client / customer to whom it relates?

    (c) How does the provider respond to requests for information received from government agencies including courts? When will information be disclosed? Will the provider notify you?

    (d) How can your clients complain if their privacy is breached?

    (e) Will the provider notify you if data is lost or stolen?

    (f) How can you ensure that the Information Privacy Principles applying in respect of information held outside New Zealand will be complied with?

55. Some suggestions for businesses to consider when using cloud services:

    (a) Consider classifying the personal information held on a spectrum from high to low risk, and determining whether you store data with a cloud provider based on that classification.

    (b) Consider encryption of data before you transfer it, so it is protected while in transit and at the provider's end.

56. Remember, as outlined previously, you are still responsible if your client's privacy is breached, even where it is held by a third party: s 3(4) of the Privacy Act.

BIG DATA

57. Big data sets have massive potential to be enormously beneficial for society. Big data sets can, for example, be used for:

    (a) Prediction of disease outbreaks in particular areas, based on spikes in internet searches for particular symptoms (Google attempted this with "Google Flu Trends").

    (b) Smart grid electricity use, by tracking which devices and appliances use the most energy and when energy demand is highest, to allow consumers to better manage electricity consumption.

    (c) Traffic management, whereby traffic congestion, and alleviation mechanisms such as variable speed limits and toll pricing systems, can be more accurately managed to suit demand.

58. An example of a business that takes advantage of big data opportunities is Uber, which sets pricing based on demand within the area at the time.

59. A further example is the (not uncontroversial) Predictive Risk Modelling research commissioned by the Government. This is an algorithm that utilises information collected from various sources to generate a risk score predicting the probability of a child's mistreatment.

60. That said, big data sets raise significant legal issues in the privacy space, including:

(a) Whether providing personal information to a big data set would be to use that information for purposes other than those advised to the individuals concerned at time the information was collected:

(i) Principle 3 requires agencies to keep individuals informed of the fact their information is being collected and the uses to which it will be put.

(ii) If individuals were not advised that their information may be disclosed to a big data set at the time their information was collected, the business may need to contact those individuals to inform them and seek their consent. This may be resource intensive considering the extent of data that may be aggregated as part of a big data system.

(iii) Otherwise, businesses run the risk of "trust exposure", if they are using information for a purpose other than that for which it was provided.

(b) Big data sets encourage maximisation of the amount of data collected, rather than minimisation:

(i) The Information Privacy Principles encourage minimal collection of personal information. Principle 1 provides that information shall not be collected unless it is *necessary* for a lawful purpose connected with a function of the agency.

(ii) Big data sets incentivise the collection (and retention) of large quantities of information on the basis that the more data, the more accurate the analysis that can be carried out, and of potential future value of data.

(iii) An additional risk is that outdated information is retained, so results of big data analysis may be inaccurate. This engages Principles 8 and 9, which require that information be kept accurate, and not retained longer than necessary.

(iv) Storage of more extensive quantities of data means greater security risks, both in the sense that the more information held, the more information may be exposed in the event of a security breach, and because storage is logistically more difficult the more information held. This creates a greater strain on the Principle 5 obligation to keep information secure.

(c) Anonymisation of data:

(i) The points made previously about anonymisation are particularly pertinent here. The obligations to fully inform affected individuals of the collection of their data and the purposes for which it is stored and used do not apply where data is properly anonymised.

(ii) However, as we have heard from John and seen from the *Google* and *Telstra* cases raised earlier, anonymisation is not as simple as it sounds.

(iii) Both case examples demonstrate that anonymised big data can and is still used to identify individuals:

(aa) in the *Google* case, for the purposes of tailored advertising;

(bb) in the *Telstra* case, the Australian Privacy Commissioner recognised that individuals' information is identified from telecommunications metadata for law enforcement purposes.

(iv) Anonymisation of data, as a reason for non-compliance with the Information Privacy Principles, can therefore be shaky ground to stand on.

(d) Right to be forgotten?

(i) In 2014, the Court of Justice of the European Union recognised an individual's "right to be forgotten" and ordered Google to delete information relating to the financial affairs of a Spanish man from its search engine.

(ii) The New Zealand Information Privacy Principles include that data should not be kept longer than necessary (Principle 9), should be kept accurate when used (Principle 8) and that an individual is entitled to access and correct that information (Principles 6 and 7), all of which are at least arguably consistent with a "right to be forgotten".

(iii) In any case, the extent of data compiled in big data sets may make it practically very difficult for individuals to be able to rely on these principles, depending how resource-intensive it may be for a big data set to cleanse the set of outdated information, ensure the accuracy of information and give individuals access to that information.

(e) Right not to be discriminated against?

(i) A further risk raised by the use of big data sets is that outcomes of the operation of algorithms may be discriminatory. Assumptions and data inputted into algorithms can create and reflect biases and inequalities, which can, in turn, create discriminatory outcomes.

(ii) John has already spoken on this issue and some examples, including a charge-card operator's algorithm flagging a customer as a credit risk based on where he spent his money. Another example is autocomplete features. These are generally a tally of related searches. Though this may seem a neutral process on its face, experience suggests that it may not be (through no fault of the search engine provider). Type, for example, the phrase "transgenders are" into Google, and the drop-down list of suggested searches includes transgenders are "annoying", "fake", "weird" "stupid", "abominations", "gross". What are we teaching machines?

(iii) This is increasingly becoming an issue in the United States in a number of areas, including:

(aa) hiring practices, where algorithms are relied on to decide which applicants get interviewed, hired or promoted; and

(bb) law enforcement, where algorithms have been used to predict where crimes are likely to occur, based on where people have been arrested previously, and "heat lists" of people most likely to commit violent crime.

(iv) Where these outcomes occur, organisations have tried to shed accountability and "blame it on the algorithm". For example, Chicago Police staff said "this program has absolutely nothing to do with race ... but with multi-variable equations". Algorithm apologists appeal to the fact that these programs deal with data in the abstract. They are quantitative. They are not fed protected attributes, such as race or neighbourhood. Algorithms are math, and math cannot be immoral. However, even such blind use of data is capable of producing discriminatory results where structural inequalities exist, and are embedded in the data used by the algorithm.

(v) Privacy Principle 10, on use of information, provides that information collected for a purpose is not to be used for another purpose. The Principle does not contain any explicit requirement that information must be used fairly, or in a non-discriminatory manner. However, it is highly likely, against the background of New Zealand's human rights legislation, that the Courts will interpret Principle 10 such that any use of personal information in a way that would be discriminatory would fall foul of the law.

(vi) A further issue is that organisations are often not transparent about their algorithms. Publicising an algorithm can not only lead to it being found to be inherently discriminatory, but can also lead to it being used by competitors. That means businesses are unwilling to share, but also that individuals cannot see how the algorithm works, in order to trust.

(vii) Accordingly, this is another circumstance where both "legal exposure" and "trust exposure" are at stake. Businesses using big data algorithms need to be aware that these can be blunt tools, and it is important to ensure outcomes are not discriminatory.

CONCLUSION

61. Emerging technologies and platforms present a range of challenges in the privacy space, but also create opportunities.

62. One such opportunity is to make privacy protection into a competitive advantage. Greater awareness and concern in relation to privacy issues means effective privacy protection can set businesses apart.

63. I will end by reiterating the importance of trust. Compliance with the Information Privacy Principles is one part of that, but agencies can go further to ensure individuals are meaningfully informed and able to give (or decline) consent, so they retain some control, in respect of how their personal information is collected, disclosed and used. Depriving individuals of that control may, in the long run, risk compromising the "social licence" according to which society tolerates the myriad uses to which data is currently put.