BIG DATA ANALYTICS: EVOLVING BUSINESS MODELS AND GLOBAL PRIVACY REGULATION

Peter Leonard, Partner, Gilbert + Tobin Lawyers, Australia and Director, iappANZ

pleonard@gtlaw.com.au

13 August 2013

At the heart of the current global debate as to how privacy regulation should address big data lie three questions:

- Can national privacy laws and regulation facilitate socially beneficial uses and applications of big data while also precluding 'Big Brother', 'spooky, 'creepy' or otherwise socially or culturally unacceptable big data practices?

- Can diverse national privacy laws and regulation, including markedly different constructs as to what is personally identifying information and sensitive information, be applied or adapted so as to accommodate socially beneficial uses and applications of big data, or is a more fundamental overhaul of law and regulation required?

- If fundamental design precepts of privacy regulation require adaptation or supplementation to address big data, can those changes be made without threatening broader consistency and integrity of privacy protections for individuals? Can any adaptation or changes be made quickly enough to address growing citizen concerns about unacceptable or hidden big data practices?

From the summer of 2012 media and policy attention in the United States as to privacy and big data focussed on data analytics conducted by offline ('bricks and mortar') businesses in relation to their customers and on the nature and range of analytics services offered by third party providers collectively labelled 'data brokers'. Media reportage reinforced unease and a perception of many people that business data analytics principally involves hidden and deliberatively secretive identification and targeting of individual consumers for tailoring of 'one to one' marketing material directed to them, including targeting by marketers with whom the individual has no prior customer relationship. The fact that this has been a U.S. led debate is of itself is not surprising, for at least two reasons. First, in contrast to the European Union and other advanced privacy regulating jurisdictions such as Canada, Australia and Hong Kong, the U.S.A. has not had economy wide collection and notification requirements in relation to PII or as to notification to the data subject as to collection and processing of PII collected about that data subject other than directly from the data subject. Second, the U.S. Do Not Track debate has focussed consumer attention upon online behavioural advertising and probably reinforced perceptions that the dominant reason for offline retailers implementing big data projects is for 'one to one' targeting and marketing.

The European big data debate since early 2012 has been quite differently focussed. The debate has included discussion of the long standing, particularly European concern as to decisions made by automated data processing without significant human judgement – so called 'automated individual decisions', or 'profiling'. The European profiling debate has a philosophical core: is the personal dignity and integrity of individuals compromised by decisions made by automated processes, as contrasted to individual decision making by humans constrained both by laws against discrimination and also, perhaps, human empathy? The profiling debate in the United Kingdom has also included a pragmatic, economic dimension. In response to consumer advocate concerns as to differential pricing online, the U.K. Office of Fair Trading examined possibilities for geo-location based and 'personalised pricing': that is, "the possibility that businesses may use information that is observed, volunteered, inferred, or collected about individuals' conduct or characteristics, such as information about a particular user's browsing or purchasing history or the device the user uses, to set different prices to different consumers (whether on an individual or group basis) based on what the business thinks they are willing to pay."

The commonality of concerns around overly intrusive or 'bad' big data practices has been partially obscured by regional and national differences in privacy regulation and in the detail of technical legal analysis as to the interpretation of privacy law. There is an engaged and continuing global debate as to how fundamental privacy concepts of notice and consent should be adapted to apply in a fully networked world of individuals and of interworking devices (the so called 'internet of things'). There has also been an active debate as to the continuing differences in national regulatory approaches to PII and particularly sensitive information such as health data and how these differences may affect implementation of now common transnational services such as global or regional data centres and software applications delivered as cloud services. Although the debate as to privacy regulation of big data has usefully focussed upon how the business practices of big data analytics can be appropriately risk managed through adaption of regulation and application of privacy by design principles, the discussion has often failed to give due credence to the depth of community concerns as to analytics about individuals conducted by third parties that do not have a direct business or other relationship with the individual and analytics that feel 'spooky' or 'creepy'.

In advanced privacy law jurisdictions privacy interests of individuals are often given effect through privacy regulation and legal sanctions and remedies (at least where these are available and affordable) attaching to breach of collection notices, privacy statements and customer terms. However, citizen concerns are also given practical effect through the significant reputational damage, and in particular adverse media coverage, suffered by governments and businesses that misjudge consumer sentiments and tolerance of perceived privacy invasive practices, regardless of whether those practices contravene laws. Lack of transparency as to activities that may conform to present law can create significant citizen concern, as most recently illustrated in the debates as to acceptable limits to alleged online metadata mining conducted by US intelligence agencies in the PRISM program and as to uses by journalists employed by Bloomberg News of their privileged access to information relating to Bloomberg customers use of Bloomberg Finance services and terminals. Sentiments expressed as dislike of 'creepiness' or 'spookiness' often reflect citizen concerns about lack of transparency and lack of control or accountability of businesses dealing with personal information about them. These concerns are often not expressed in terms of these basic privacy principles and often do not map to existing laws. There is a growing deficit of trust of many citizens in relation to digital participation, as demonstrated by pressure for expansion in profiling restrictions under European privacy law, for 'just in time' notices as to use of cookies, enactment of Do Not Track laws and laws restricting geo-tracking and employers access to social media. That deficit of trust threatens to spill-over to offline data applications and by so doing endanger socially beneficial applications of big data by businesses and by government. The perception of citizen unease has pushed some businesses to be less transparent about their data analytics projects, which has reinforced the sense of a growing climate of business and government colluding in secrecy.

The counter-view is that a growing sector of the public comfortably live their digital lives reflecting the oft-quoted aphorism that 'privacy is dead' and may therefore be expected to become more accepting of privacy affecting big data analytics as time goes by. However, there is already compelling evidence that many individuals presented with privacy choice will display a more nuanced and contextual evaluation as to what personal information they particularly value or regard as sensitive, as to particular entities with whom they will entrust their personal information and as to the trades that they are willing to make for use of that information. As individuals come to understand the economic value that increasingly accrues around personal information, it is reasonable to expect that these contextual judgements will become even more nuanced and conditional. It may be that the deficit of trust in digital participation is growing and not merely a relic of inter-generational differences.

Application of today's privacy regulation to map a course through big data implementation may miss the mark of sufficiently addressing this deficit of trust. Not infrequently, business customer analytics projects stall at a point where a chief marketing officer has successfully addressed the concerns of the chief information officer, the chief privacy officer and the general counsel, but the chief executive or a consumer advocate within a corporation is then persuasive with her or his view that customers will not trust the business with the proposed implementation. Moreover, the trust deficit can be highly contextual to a particular transaction type, a particular vendor-client relationship, a distinct geography, or a particular culture. Many consumers understand that enabling geo-location on mobile devices for

a particular app enables the provider of that app to target content of offers to them based upon that location. Many consumers understand that they derive a benefit from a loyalty card in a value exchange with a vendor who will use that loyalty card data for customer analytics to target offers to that consumer. A direct and proximate vendor-client relationship promotes accountability: consumers may vote with their trade if the vendor betrays the customer's expectations, whether those expectations are based on legal rights or not. A direct and proximate relationship also leads to accountability: many consumers will draw no distinction between a vendor and the vendor's sub-contractors, such as external data analytics providers, in relation to breaches of security or uses or abuses of personal information given to that vendor. By contrast, the term 'data broker' of itself conjures the sense of lack of accountability and lack of transparency, in addition to there being no value exchange between the broker and the affected individual.

Engendering trust requires more than good privacy compliance. Compliance is, of course, a necessary component of responsible business governance for using data about individuals for marketing purposes, but it is only one component. Responsible governance of data analytics affecting citizens, whether by businesses or government, requires a new dialogue to be facilitated to build community understanding as to appropriate transparency and fair ethical boundaries to uses of data. This requires both businesses and government to acknowledge that there is both good big data and bad big data and that transparency as to data analytics practices is necessary for this dialogue and community understanding.

Fundamental failings of many data analytics projects today include unnecessary use of personally identifying information in many applications where anonymised or de-identified transaction information would suffice and omission of technical, operational and contractual safeguards to ensure that risk of re-identification is appropriately risk managed. Both good privacy compliance and sound customer relations requires planning of operational processes to embed, in particular, safeguards against re-identification of anonymised information, in how an organisation conducts its business, manages its contractors, offers its products and services and engages with customers. Privacy by design and security by design is sometimes implemented through a binary characterisation of data as personal and therefore regulated, or not personally identifying and therefore unregulated. The developing privacy theory adopts a more nuanced, graduated approach. This graduated approach puts re-identification into a continuum between certainty of complete anonymisation and manifestly identifying information and then seeks to answer four implementation questions:

- Can this graduated or 'differential' approach be made to work within diverse national current regulatory regimes and varying definitions of personal information and PII and requirements as to notice and consent, data minimisation and limits to data retention?

- How should a privacy impact assessor or a privacy regulator assess the risk mitigation value of stringent limited access and other administrative, operational and legal safeguards? Are these safeguards only relevant *in addition* to high assurance of technical de-identification?

- Is there a subset of legal obligations that should apply to users of de-identified datasets about individuals to protect against re-identification risk?

- How should citizens be informed about customer data analytics so as to ensure that notices are understandable and user friendly? How can these notices accommodate the dynamic and unpredictable manner in which business insights may be discovered and then given operation in production data analytics?

Privacy theory meets the reality of business and government big data analytics in the way that these questions will be answered in business practices. The last question must be answered sufficiently quickly to build community understanding and engagement as to 'good big data' before concerns by privacy advocates and concerned citizens as to 'bad big data' prompt regulatory over-reach. Although these questions have not been definitively answered by privacy regulators, over the last year regulators in a number of advanced privacy jurisdictions, including the United Kingdom, Australia and Singapore, have published views that usefully and constructively engage the debate.

What is striking from a comparison of these regulatory views is the conceptual similarity between the approach of these regulators in answering the question as to when personal information, or personally identifying information, as diversely defined and interpreted under national laws, should be considered sufficiently de-identified or anonymised as to make re-identification unlikely. The conceptual similarity is of itself is unusual: most areas of national privacy regulation are characterised by marked divergence in national or regional privacy theory and practical application. Each regulatory view requires assessment of the sensitivity of the data, the context and limits of its disclosure and implementation by the data analytics provider of appropriate risk mitigation measures. Once the first assessment has been completed in terms of the possibilities and limits of effective de-identification, the second step of applying additional safeguards will often need to follow. Although the standard for acceptable risk is variously stated, the regulatory views are not dissimilar - 'low', 'remote' or 'trivial'. The possibility of re-identification is contextually assessed, or as the U.K. Information Commissioner puts it, 'in the round'. Risk mitigation measures – being appropriately 'robust' safeguards – are to be implemented before purportedly anonymised data is made available to others. These risk mitigation measures may be a combination of technical, operational and contractual safeguards. The regulatory views also converge in not being prescriptive as to particular safeguards, instead offering a menu board approach for consideration in a privacy and security impact assessment individual to that deployment as to the safeguards appropriate for a particular data analytics deployment.

The menu board of safeguards is relatively long. It includes use of trusted third party arrangements; use of pseudonymisation keys and arrangements for separation and security of decryption keys; contractual limitation of the use of the data to a particular project or projects; contractual purpose limitations, for example, that the data can only be used by the recipient for an agreed purpose or set of purposes; contractual restriction on the disclosure of the data; limiting the copying of, or the number of copies of, the data; required training of staff with access to data, especially on security and data minimisation principles; personnel background checks for those granted access to data; controls over the ability to bring other data into the environment (allowing the risk of re-identification by linkage or association to be managed); contractual prohibition on any attempt at re-identification and measures for the destruction of any accidentally re-identified personal data; arrangements for technical and organisational security, e.g. staff confidentiality agreements; and arrangements for the destruction or return of the data on completion of the project.

While these regulatory views are being developed and refined, the questions that the regulators are tentatively answering are already being addressed through business practices that, if and when done well, deploy technical de-identification and also embed privacy impact assessment, privacy by design and security by design principles into other operational (administrative, security and contractual) safeguards within data analytics service providers, governments and corporations. But because this area is new, there is no common industry practice as to such safeguards, and sub-standard implementations continue and threaten to further erode citizen trust as to big data. If bad practices and bad media further promote other businesses and government to be less transparent about their data analytics projects, public perception of business and government colluding in secrecy will grow, prompting more prescriptive regulation. Big data and the privacy regulatory and compliance response to it will be one of the most important areas for development of operational privacy compliance for the next five years.