GILBERT +TOBIN

LEGAL IMPLICATIONS OF BIG DATA

NOTES BY PETER LEONARD

Australian National University Public Law Weekend-2016
law.anu.edu.au/event/conference/public-law-weekend-2016
Session 4, Friday 28 October 2016

Session 4, Friday 28 October 2016

How is data disruption changing businesses and creating new legal issues?

In the sixteen years of this century we have already seen four phases of data disruption of business models. It is reasonable to suggest that the impact of each successive phase has been greater than the phase that preceded it. The phases overlap and some earlier phases continue to work out: some commentators suggest that the most radical restructuring of the media industry is only now in play.

The **first phase** of data disruption of business models through online services was **substitution out of offline sources of supply of products or services in favour of supply from new online sources**. The best recognised examples are:

- **disruption of newspapers by online content sources** (which in turn enabled advertising to escape to other online media and to other formats such as outdoor).
- **search engine** 'organic search' and 'advertiser sponsored links' capabilities enabling disruption of traditional business and trading relationships and customary offline means of finding of products and services.
- **disruption through disintermediation**: for example, suppliers of replacement parts being able to take orders directly from users without using local distributors or maintaining local stocks; Amazon selling directly to consumers (firstly books and then a broad range of consumer products).
- **disruption of proprietary or locked down distribution systems** through offer of broadband and mobile alternatives – e.g. of broadcast television channels and of satellite and cable TV systems by over-the-top (OTT) internet fixed and mobile broadband.

These disruptions fundamentally changed business models. These disruptions also created a key data driven change: the new businesses created new sources of data driven value because data becomes available for collection and use as an incident of service delivery, which in turn could be monetised through better segmentation of the customer base for marketing and re-targeting. 'Network effects' of scale were magnified by availability of more granular, and therefore more valuable, data.

New data value led to rapid growth in the importance of unstructured search and created a new industry of 'below the line' advertising including online behavioural advertising and associated ad exchanges. This also promoted specialist sites addressing particular needs - cars, real estate, jobs, restaurant, hotel, travel etc. – amplifying challenges for traditional media's principal sources of revenue.

The **second phase** of data disruption was the **rise of social networks** (Facebook, QZone, Twitter, VKontakte, etc.), closely followed by **rapid take-up of smartphones**.

Rapid consumer take-up of both social networks and smartphones enhanced the depth and range of data available to some service providers about individuals and their interactions and locations. This enabled new value capture and lock-in by the large social network providers, created new revenue sources for telcos and comms carriage providers, and enabled entirely new disruptive businesses such as Uber and AirBnb.

This collection and use of data is now pervasive, closely related to identifiable individuals, readily captured and retained over enduring periods. Often this data is not consciously contributed by an individual. As one illustration, look at any smartphone setup and see which of the apps accepted by clicking 'I agree' is collecting data unrelated to the stated functionality of the app and not readily explicable as improving the user experience.

Much of the valuable information that is collected is so-called 'metadata' - information *about a communication* (the information often being a human to machine transaction, and increasingly machine–to-machine interaction (the Internet of Things). Contrast human to human communication and capture of *the content of the communication* (transaction) itself. Many individuals will either not be conscious of the collection of the information, or if once aware, will not retain that awareness or recognise its significance. The value of this information is demonstrated both by alacrity of law enforcement agencies in seeking access to it and the range of applications fuelled by data analysis of information about communications.

This phase created **fundamentally new legal challenges**.

Firstly, pervasive devices and communication challenged traditional ways of thinking about data protection and privacy. Most obviously, smartphone convenience has created bad habits of smartphone users clicking straight through user terms and privacy policies and hoping for the best. These habits run counter to privacy regulation that relies upon **transparency through disclosure** to moderate excessive collections and privacy invasive uses and disclosures of personal information. Accountability of data collectors and data controllers based upon notice or consent to terms will fail if most users pay no attention to the terms. This is particularly a problem in our current world where regulators and independent consumer advocates that monitor supplier terms are challenged by limited resources and sometimes limited understanding (not helped by differences in terminology and complex service delivery eco-systems). Many of the relevant services are global services, bringing associated cross-jurisdictional challenges.

Second, the privacy 'trade-off' underlying a consumer's decision as to use of a service has become more complicated and nuanced. That trade-off is as to how much personal information a consumer is willing to give up in return for the benefits afforded by the service, such as access to particular features or provision of the service without charge. Uber needs to know where I am (and ideally also what I look like) to improve my user experience and to afford personal safety, but collection and use of that information compromises (what I regard as my right of) freedom of unobserved movement. The detail as to what Uber can and can't do is deep in online pages expressed in terms quite unfamiliar to many individuals and, indeed, to many privacy lawyers. Finance economists has demonstrated in recent years that many individuals closely associate risk with lack of trust and accountability: if there is a deficit of trust, perception of risk is substantially enhanced. Any fair evaluation by a consumer as to the privacy trade-off (and of any associated value exchange) requires reasonable transparency of how relevant information is being collected and used, and perceived accountability of the service provider in the event of misuse or other misbehaviour. Industry norms are difficult to find, partly because business models rapidly change, which limits development of consensus as to what is fair and what is not. So there is a substantial information asymmetry and likely deficit of trust, even if consumers elect to actively 'negotiate' the privacy trade-off.

Third, the very concept of personal information as the basis for determining whether to apply regulation is increasingly problematic. Most sophisticated privacy jurisdictions draw a distinction between personal

information and other information that is not identifying of a particular individual having regard to both direct identifiers and indirect identifiers (e.g. address and available information that might be 'looked-up' to re-identify an individual) and in combination with other information reasonably available to a party that has access to this information. In other words, purportedly non-identifying information may in fact be 'personal information' or 'personally identifying information' (PII)) because re-identification remains possible (taking into account both particular information and other available information). There is currently a vigorous debate in many privacy jurisdictions as to the level of re-identification risk at which indirect identifiers that may be available and associated with information that is not directly identifying of an individual may be discounted and the relevant purportedly non-identifying information treated as de-identified or anonymised.

Fourth, the balance of incentives today is arguably distorted because of overreliance upon the notice and consent framework as the basis for accountability of data collectors and data controllers. Unless the disincentives for collecting and using personal information are sufficiently strong, data collectors and data controllers may elect to use and analyse personally identifying information. If the regulatory incentives are right, service providers will properly de-identify information and conduct data analytics in a safeguarded environment. De-identified (or anonymised) information should be managed through implementation of verifiably reliable processes and practices such that the risk of re-identification of affected individuals is sufficiently remote. If data collectors and data controllers are not appropriately incentivised to conduct analytics upon properly deidentified information in properly managed environments, privacy affecting data analytics is more likely to occur. Risk of inadvertent privacy breaches are also increased.

The **third phase** of data disruption is **merging of offline and online data sets**, and of **many offline data sets from multiple players**, through 'big data' platforms and data analytics and data sharing between multiple players. We are in the midst of this phase. This third phase is more disruptive and transformative because many data custodians could be challenged by new players deploying advanced data analytics and agile business models without incumbent encumbrances of legacy systems, legacy thinking and traditional sources of revenue.

This third phase is also now working out at the same time as a **fourth, data driven phase** of business disruption. This fourth phase has two parallel, often interrelated streams. The first is **IoT sensor driven disruption.** The second is what might loosely be described as **autonomous systems**, be they artificial intelligence or machine learning applications, driverless cars or other applications which can operate - and sometimes self-diagnose, repair and improve – without human intervention. Associated with this fourth phase is a new set of difficult legal issues. These include:

- allocation of responsibilities for **information security vulnerabilities**,
- transparency as to **data uses and data 'ownership**' and
- how to adapt traditional concepts of **negligence** in tort law, and consumer protection statutes, to a world where overlooked or unanticipated fact scenarios are not addressed in code, or errors made in coding, that lead to damage, injury or death, and
- humans making decisions in **reliance** upon technology and devices that may be faulty for entirely unanticipated reasons.

Some of these issues have recently been discussed in the context of technologies that are now early in their development: autonomous vehicles, robots and artificial intelligence systems. But the issues are with us today through take-up of the Internet of Things. Mark Rolston, Cofounder of argodesign, describes the Internet of Things (also known as the Internet of Everything) as being when "everyone, everything, and everywhere will be codified, interactive, and addressable through ubiquitous interfaces scattered throughout our environments."

Why has the IoT era arrived so quickly? First, there has been **a sharp decline in cost of sensor devices** that can interface with remote data analytic capabilities either through special purpose IoT platforms, such as Nest smart home devices, or directly over the internet. Second, there are a number of initiatives to **make data more discoverable and therefore also shareable**, between devices, between services and platforms, between particular services and other service providers, and between service providers. Hypercat and other data

ontology initiatives are fuelling interoperability and stimulating take-up of IoT devices and services.

We are already seeing **rapid growth in IoT devices**. Gartner estimates 6.4 billion IoT devices ('connected things') in use today and 21 billion by 2020.

IoT devices by definition are able to communicate with each other via the internet and without direct human intervention. Typically they are pretty dumb, 'edge' devices, which also means they can be low cost and very low power, with the smarts happening in the centre. Your smart phone or tablet is not typically an IoT device because it usually is collecting data through human intervention, and it can process data inputs under its own power. Similarly, an autonomous car is not really an IoT device, because the smarts are in the car – which is a very good thing when you unknowingly drive out of internet coverage!

Smart phones are however an essential part of many IoT deployments. Typically they deliver insights about remote conditions to a human who can then decide whether to actuate a response, such as by entering a command into the phone that is then relayed back to an actuator device at the distant end. Typical smart phone IoT integrations include smart baby monitors and personal health devices like fitbits or iWatches that sense and deliver personal health data that is remotely analysed with insights presented back to the smart phone. Examples of IoT deployments include:

- **smart homes**, where kettles, curtains, fridges, air-conditioners, motion sensors, sprinklers, pool filters and so on all communicate through a Nest IoT platform device or an internet router with the internet and beyond to supermarkets, energy companies and so on.

- **smart cities** – e.g. major building projects where a developer can create precinct wide systems, bristle with interworked smart city applications, ranging from motion surveillance to lift service control, energy monitoring and so on.

Often sensor devices will also be actuator devices, turning off lifts and lights, ordering up milk, etc., without human intervention.

Eliminating humans eliminates human error, but it also removes human judgement.

Humans are fallible, but on a good day they do think. 'No human hands' creates scope for **error and mischief** – think of the hacked Jeep driving itself into the ditch – and also for **profiling or discrimination by algorithm**, where discrimination is not sensed and controlled through exercise of human discretion.

To summarise the **benefits of IoT**, data analytics coupled with IoT devices will often promote business efficiency and consumer welfare through any (or all) of:

- reduced costs from higher asset utilisation or;

- higher labour productivity,

- more efficient use of assets (just enough) so lower waste,

- improved supply chain logistics,

- businesses gaining new customers from improved product experiences, and

- reducing the time to market for innovations and innovative updates.

There are associated **risks of IoT**:

- compromising consumer trust and invading privacy,

- creating new sources of liability,

- creating confusion as to who is responsible for what, and

- from hacking and other security breaches. Benefits of inter-connectivity bring attendant risk – inter-connectivity clearly carries contagion risk from the weakest or most vulnerable point in the network.

And all these challenges to be addressed while robots and artificial intelligence systems now 'come down the turnpike'. New challenges for ethicists and lawyers.

Peter Leonard
Gilbert + Tobin
+61 2 9263 4003
pleonard@gtlaw.com.au
www.gtlaw.com.au
28 October 2016

SYDNEY
Level 35 International Towers Sydney
200 Barangaroo Avenue
Barangaroo NSW 2000
T +61 2 9263 4000
F +61 2 9263 4111

MELBOURNE
Level 22
101 Collins Street
Melbourne VIC 3000
T +61 3 8656 3300
F +61 3 8656 3400

PERTH
1202 Hay Street
West Perth WA 6005

T +61 8 9413 8400
F +61 8 9413 8444

WWW.GTLAW.COM.AU