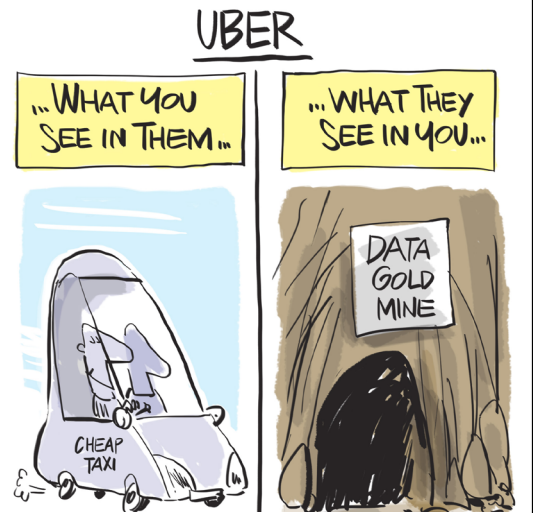


## Big Data: The Legal Issues

June 2016

This paper summarises Big Data issues presented at the New Zealand Law Society Cyber Law Legal Conference held in early 2016. If you are interested in further content from the conference see the [NZLS Website](#).

Read the full article below.



Wigley+Company

PO Box 10842  
Level 6/23 Waring Taylor Street, Wellington  
T +64(4) 472 3023 E [info@wigleylaw.com](mailto:info@wigleylaw.com)

and in Auckland  
T +64(9) 307 5957  
[www.wigleylaw.com](http://www.wigleylaw.com)

*We welcome your feedback on this article and any enquiries in relation to its contents. This article is intended to provide a summary of the material covered and does not constitute legal advice. We can provide specialist legal advice on the full range of matters contained in this article.*

---

**BIG DATA: ARE LAWYERS SCREWED (AND WHAT TO DO ABOUT BIG DATA UNTIL THEN)?**

---



# **BIG DATA: ARE LAWYERS SCREWED (AND WHAT TO DO ABOUT BIG DATA UNTIL THEN)?**

**Michael Wigley**  
Wigley and Company  
Wellington

---

## **Overview**

Big Data is taking away our legal jobs. Why, and what do we do about it in the meantime?

This paper complements Katrine Evans' paper on privacy aspects of big data and Judge Harvey's paper on discovery, Internet of Things and big data. They have described what Big Data is. We'll build off that to outline the impact in business and government, including on law firms and in-house lawyers. This builds from the examples that the earlier two papers give.

We'll overview the closely related area of disruptive technologies, as big data is integral to many of those new services.

We'll use real life examples to illustrate the impacts of big data and the legal issues.

Then we will outline why Big Data is relevant to lawyers advising companies and public sector entities, and what to do about it. We end with a checklist for lawyers dealing with Big Data.

## **Disruptive technologies and big data**

Recapping, big data involves drawing conclusions from large and complex data and using that, in a business context, to drive profit and efficiencies (and in a public sector context to driver better outcomes and efficiencies).<sup>1</sup>

Big Data generally involves using statistical techniques (particularly, in this context, predictive analytics) to derive value. The 5 V's described in the earlier papers are useful indicators.

Few businesses are immune from the threat and opportunity of Big Data and the evolution of overlapping technologies such as artificial intelligence (e.g. the IBM computer playing chess against grandmasters, and computers doing the "thinking" instead of lawyers)<sup>2</sup> and the Internet of Things.<sup>3</sup> It's also often a key part of so called disruptive technologies, the

---

<sup>1</sup> There's snake oil on what is and isn't Big Data and in the end it's not a particular closely defined model. Leading ICT consulting firm, Gartner, have one of the most accepted definitions of Big Data.

"Big data is high volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization."

Some add another two V's: Veracity (how accurate is the data and the outputs) and Value (what value does it deliver).

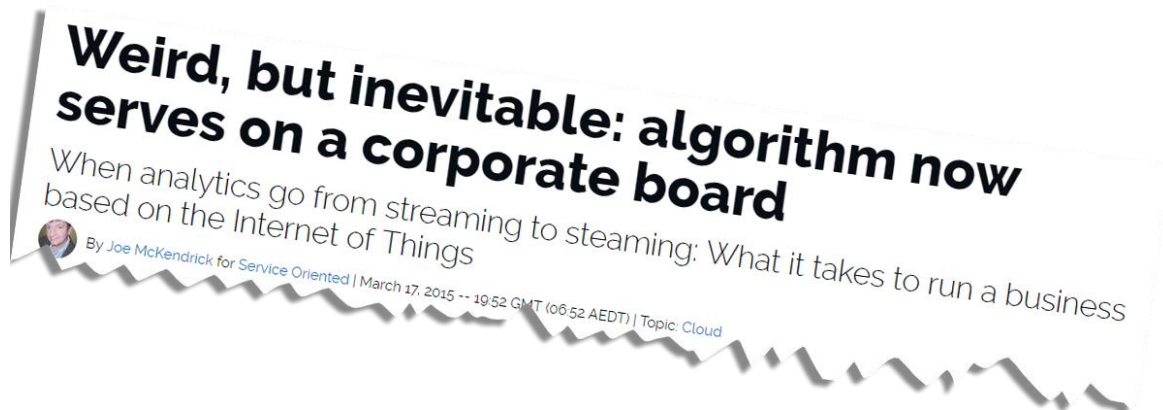
<sup>2</sup> Although computers technically don't "think" the same way as humans so "artificial intelligence": can be misleading: they produce results by a different path.

<sup>3</sup> For more detail see our articles, The Internet of Things – ramping up privacy and security considerations

new services that up-end traditional businesses. Classic examples are Uber and taxis, and Amazon and bookshops, where the disruptors analyse and use vast amounts of information to provide a service tailored to individual customers. Amazon trawls millions of pieces of data to be able ultimately to recommend to the reader the best books to read.

As Katrine points out, those are B2C examples, when there are countless B2B, internal business applications and Government applications as well.

Check this news headline for example.



The headline refers to an English company appointing a big data software application (algorithm) to its board to provide input towards investment decisions. The company takes investments in health companies, based on its analysis of the big data in the space. Of course the truth is that computers can't yet be directors, but where is this all heading?

### **Even the safest businesses are not immune**

We only have to look at what is happening to the seemingly inviolable businesses of electricity utilities and banks, to see what's ahead for many industry sectors, as we outline in our article, *It could happen to you*.<sup>4</sup> If it can happen to such AAA investment-grade businesses, it can happen to any business.

For example, due in large part to Big Data techniques, new entrants can steal much of the banks' business lines. Apple Pay is an example that might usurp the credit, debit and EFTPOS card world. Or a bank can make the right crap shot now and develop systems that are good enough to keep the newbies out. But in a world where it's not clear where the market will go, this truly is a Bet-the-Bank scenario where the choice to go down Path A may prove wrong. From what we've seen from working in this area, it also wouldn't be too surprising to us if, as in other businesses, some bank boards and managers don't recognise the gravity of the situation.

Now to examples of practical application of Big Data, which will also help describe the legal issues.

---

<http://tinyurl.com/mm3989d> and A Telco regulator's take (and our take) on the Internet of Things

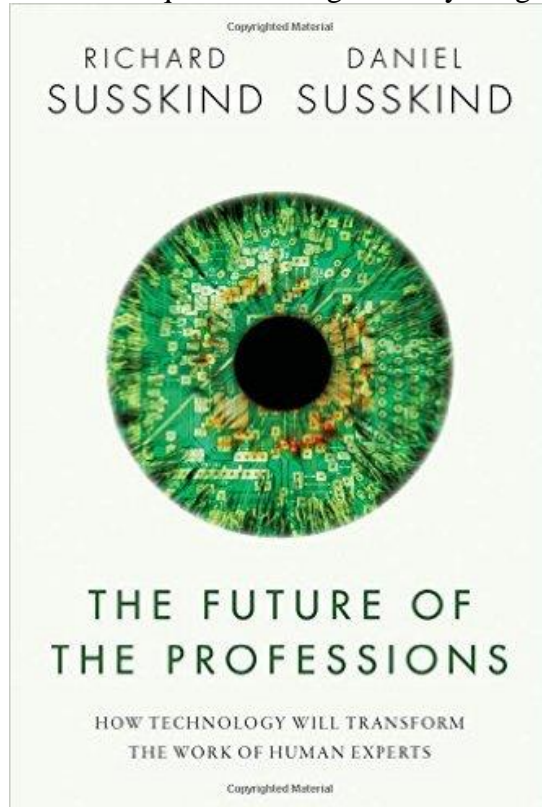
<http://tinyurl.com/kl9zfob>

<sup>4</sup> <http://tinyurl.com/z9hxox9> -

## What of lawyers' jobs in the future?

Where lawyers' work – inhouse and external – is heading is a good example of big data in action, particularly of the "Variety" V.

There are plenty of obvious areas for lawyers where large amounts of data can be used to impact outcomes, such as legal document creation and estimating costs for particular jobs. But that's just touching the sides. Big Data works on both structured and unstructured data. For structured data, think spreadsheets, timesheets, and the like: it's neatly set out like soldiers in a row. That's relatively manageable across multiple databases when it comes to deriving value. For unstructured data, think emails, videos, etc. It's much harder to take millions of emails, etc., and derive usable conclusions from that. However, that is what Big Data techniques are doing and they are getting better at it increasingly quickly.



Take the unstructured data that are legal judgments. At the end of last year, the leading thinker on the future of the law profession, Richard Susskind, produced his latest book, with his son, Daniel, *The Future of the Professions*.<sup>5</sup> They concluded that, based on analysing judgments:<sup>6</sup>

Big Data techniques are underpinning systems that are better than expert litigators in predicting the results of court decisions.

Predictions based on the foibles of the particular judge where relevant can happen too. Add other material such as submissions and there's plenty of opportunity.

Note that this can be done today, so that's just the start.

No wonder Lexis-Nexis is buying up Big Data companies.

And before anyone starts to think that computers will be stumped by emotional foibles when it runs the odds over predictive analytics, don't forget that human prediction and human decision making, including by top QCs and by top judges, comes with a bunch of biases and foibles. Biases that change over time and over circumstances. The Susskinds give the example of judges in Israel hearing applications for early parole release by prisoners. Research shows that those dealt with in the morning are more likely to get parole than those dealt with in the afternoon. All of which can better be predicted by computers (and the biases adjusted by computers too). Maybe having a computer doing the judgments will be more reliable when it comes to deciding issues as serious as loss of liberty? Let's be clear: you are more likely to stay in prison if your case is heard in the afternoon. Now that is a serious human rights failing. While there are big issues of course in computerising, it certainly doesn't follow that the human system is necessarily optimal.

<sup>5</sup> Oxford University Press

<sup>6</sup> At Page 69

## Target and the teenage pregnant daughter – second example

For a good example of big data in play, there's the New York Times article, *How companies learn your secret*.<sup>7</sup> The article focusses on the discount retail chain, Target, a chain similar to the Warehouse, which sells products from lawn mowers to baby clothes. Target analyses much of its data, such as purchase histories derived from sales on customer loyalty cards.

The article reports an angry man demanding to see a Target store manager about his school-aged daughter getting Target ads in the mail for baby clothes and cribs. The manager rang a few days later to apologise. Says the article:

On the phone, though, the father was somewhat abashed. "I had a talk with my daughter", he said. "It turns out there's been some activities in my house I haven't been completely aware of. She's due in August. I owe you an apology."



### Target's use of big data

Maybe that story is an urban myth despite being treated as news. But, what is clear is that, using predictive analytics, Target (or The Warehouse here) are able to work out whether a woman is pregnant early on in the pregnancy, and can even predict the likely birthdate. It's normally very hard for retailers to break buying habits (such as changing a customer from one retailer to another or from one brand to another). But that's more likely during a major life event. If the retailer can target the mother as early as the second trimester, it can get in before others.



---

<sup>7</sup> <http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html>

A Target employee gave an example: if in March a woman, hypothetically Jenny Ward, buys cocoa-butter lotion, a purse big enough to double as a diaper bag, magnesium supplements and a blue rug, there's an 87% chance she is pregnant and due in August.

Then, how do the marketers use that data? Sending out a flier saying, "Congratulations, Jenny, on your first child due later this year", will meet a hostile audience. She'd think that Target is stalking her.

Even a mailed catalogue devoted only to baby products can spook the mother. As the article notes:

...for pregnant women, Target's goal was selling them baby items they didn't even know they needed yet.

"With the pregnancy products, though, we learned that some women react badly," the [Target] executive said. "Then we started mixing in all these ads for things we knew pregnant women would never buy, so the baby ads looked random. We'd put an ad for a lawn mower next to diapers. We'd put a coupon for wineglasses next to infant clothes. That way, it looked like all the products were chosen by chance.

"And we found out that as long as a pregnant woman thinks she hasn't been spied on, she'll use the coupons. She just assumes that everyone else on her block got the same mailer for diapers and cribs. As long as we don't spook her, it works."

So this gets across the line by disguising the real agenda. Same use of data: just disguised.

But of course this happens all the time: it just looks like Big Data will expand it exponentially.

### **Wider implications**

Expand this story out to a wider range of information, sharing of data between companies, and so on, and it can be seen that organisations face legal and reputational risk on a grand scale.

Security experts are pointing to a new set of greater and different security issues raised by big data, not easily handled by traditional security measures. Add to that the shortage of expertise in the big data area and there are risks for organisations. See Katrine's paper for a more extensive overview of this aspect.

### **"Uber – the Big Data Company" - third example**

Uber's business is far more than getting people from A to B. It's a disruptive technology that leverages big data on a large scale. Uber gets and uses vast amounts of highly personal information about passengers. Until they were caught out, they even had what they called the God View, and senior managers used it to track Uber's opponents such as journalists, for example.





As Ron Hirson reported in Forbes magazine in Uber: the Big Data Company:<sup>8</sup>

There are only four people/organizations in the world who know my location at all times: my wife (because I tell her), Apple (because Siri), the [electronic spy agency] NSA (because NSA), and now Uber.

Since the service Uber has built is so convenient, and increasingly essential to my life, Uber knows where I live, where I work, where I eat, where I travel, where I stay/visit and when I do all these things.

I am no longer just a passenger or a fare. I am a big data goldmine and, in case you hadn't noticed, Uber just broke out the pickaxes.

This year, we are going to see the transformation of Uber into a big data company cut from the same cloth as Google, Facebook and Visa – using the wealth of information they know about me and you to deliver new services and generate revenue by selling this data to others.

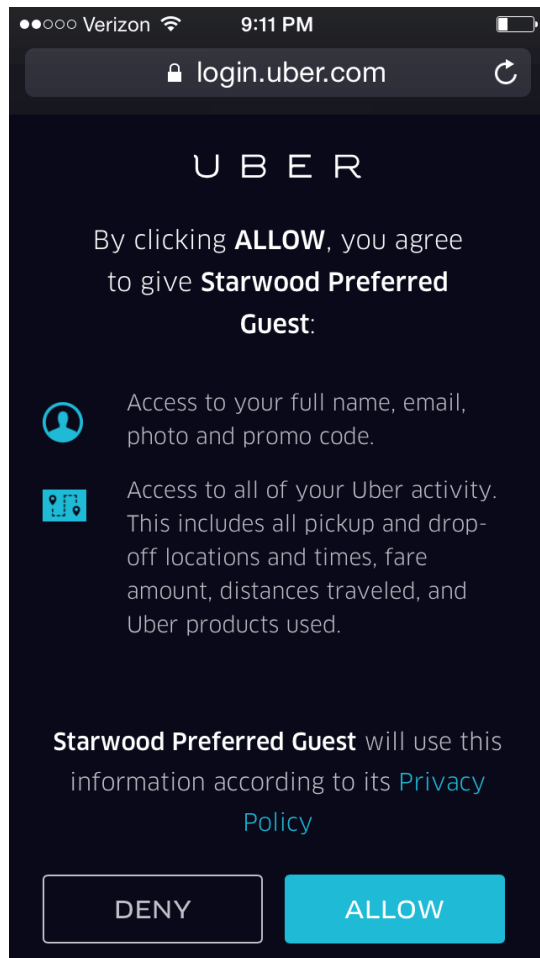
Just like other companies sell the data they collect to others (sometimes anonymised,<sup>9</sup> and sometimes not) so too can Uber sell its data to other companies. As the footprint expands to other companies, that ecosystem can be hugely profitable for Uber.

The Forbes article gives an example of Uber being able to provide all data it collects on a passenger to Starwood, the owner of the Hyatt hotel chain.

---

<sup>8</sup> 23 March 2015

<sup>9</sup> Although increasingly combining databases enables de-anonymisation of so called anonymised data, which is another privacy problem area.



Starwood can track when someone is near one of their hotels and use that to market, etc. Typically, passengers won't fully understand just how much information can be provided. The screen shot shows this can include everything ("all of your Uber activity").

Information exchange will go both ways. For example the Starwood loyalty points scheme can have points added when a scheme member takes an Uber car. Or Uber could be alerted when one of its customers checks out of a Starwood hotel.

There are multiple other businesses doing deals with Uber like this. Plus Uber can use the data for many other services such as Uber couriers, Uber takeaway meal deliveries and so on.

### Some legal issues flowing from the Uber and Starwood situation

The Uber and Starwood situation illustrates big data issues applicable to many other scenarios.

It raises obvious privacy issues (for example, flowing from Katrine Evans' paper, the position under IPP 5(b) in the Privacy Act (and overlapping law internationally) as to customer data shared with third parties). It is also an example, in a privacy and contractual context, of sharing and combining databases. If NZ law applied (and often under foreign law), Uber cannot simply do a Pontius Pilate when giving Starwood access to the data (and vice versa for Starwood giving data to Uber).

Both Uber and Starwood need to have an agreement to deal with who owns what in the databases, and the co-mingled output from combined use, and what limits there are on Starwood's (and Uber's) use. They would be unwise to rely on the default copyright law position, which varies between jurisdictions anyway, and they should provide specifically by contract as to what is to happen.

There will be payment considerations (what and how will Starwood pay Uber, and vice versa? Should that be auditable?). Who can and should do what marketing?

The parties may wish to provide for responsibility and liability as to the accuracy of the data provided. Plus there will be typical boilerplate considerations such as rights on termination, limitation of liability, etc.

The sheer scale of such personal data, and that it is being shared, heightens the cybersecurity risk and that should receive close focus by the parties in the contracts, and in what they do.

There may be regulatory considerations specific to industries (or specific to the public sector as relevant). A generic one is the emerging competition law implications of sharing large databases, particularly by those with substantial market power. This is a significant

issue already with competition law authorities already considering the anti-competitive effects of acquisitions where data is aggregated. For example, as to commercial objectives and implications:<sup>10</sup>

Google has also been an aggressive acquirer of firms that bring new data to Google. Some prominent acquisitions in recent years by Google do in fact suggest data availability was a major factor in the acquisitions. The purchase of YouTube brought a large and growing amount of video and data on the video watching habits of consumers. Zagat brought extensive and historical data on restaurant reviews. Waze, a traffic and navigation app, brought real-time traffic data and consumer commuting data. Nest brought data on home temperatures and home presence, and DoubleClick brought data on advertisers and the ability to deliver optimal searches for customers and advertisers. In all of these cases, data and its potential for next generation products played a major role in the business extensions.

Plus, increasingly there are multi-jurisdictional issues.

### **But I advise a small company; surely big data isn't relevant?**

SMEs can access big data too, as a recent text points out:<sup>11</sup>

The good news for firms looking to leverage Big Data is that IT solutions for managing and storing such data are becoming dramatically more accessible. This is owed to flexible deployment models, like cloud computing, whereby a set of servers run software for the firm. Given the development of flexible data management tools, accessing data across cloud connections has never been easier. Solutions that allow for leveraging best-in-class analytical resources and hardware are in place, and these solutions do not require buying hardware or managing a team to operate it. Some analytical consulting services like Mu Sigma and EXLservice, among others, can even provide outsourced analytics, whereby the results and insights are delivered and the analytical function is accessed from these firms. The access to analytics and the underlying IT resources have dramatically opened up and will continue to do.

This highlights that there will be big data service supply agreements (cloud based, outsourced, and so on) for lawyers to deal with.

### **Checklist for lawyers**

Big Data analytics raise these issues:

#### **Big Data Legal Checklist**

Privacy, security and confidentiality	<ul style="list-style-type: none"> <li>• • Privacy/data protection legislation</li> <li>• • Confidentiality law</li> <li>• • Best practice security</li> <li>• Where relevant, appropriate stakeholder opt-in to use of data</li> </ul>
---------------------------------------	---

---

<sup>10</sup> Russell Walker, *From Big Data to Big Profits: Success with Data and Analytics*

<sup>11</sup> Russell Walker, *From Big Data to Big Profits: Success with Data and Analytics*

	<ul style="list-style-type: none"> <li>• Contracts with external suppliers and customers</li> </ul>
Copyright	<ul style="list-style-type: none"> <li>• Protect IP rights in databases (e.g. limit use of database by third parties to specific purposes; try to retain IP in databases and information created and derived using the database)</li> <li>• Contracts with users of data such as other companies</li> <li>• Dealing with IP rights the database provider has</li> </ul>
International	<ul style="list-style-type: none"> <li>• For example, international issues from cloud computing.</li> <li>• Off-shore use of data</li> </ul>
Contract	<ul style="list-style-type: none"> <li>• Issues above</li> <li>• Licence scope and limits on use as to supplied databases</li> <li>• Limit liability</li> <li>• Payment and price</li> <li>• Rights in commingled data (i.e. data that has been mixed with other data)</li> <li>• Rights on termination (as to database, commingled data and information prepared from data)</li> </ul>
Regulatory	<ul style="list-style-type: none"> <li>• Industry specific regulation</li> <li>• General regulation e.g. competition law</li> </ul>
Public law (if relevant)	<ul style="list-style-type: none"> <li>• For Big Data use in the public sector there may be specific issues</li> </ul>
Contracts with big data service suppliers	<ul style="list-style-type: none"> <li>• Scope and services</li> <li>• IP rights in data, information and comingled data</li> <li>• Software service terms</li> </ul>