RPC

Size doesn't matter: regulating "big data" in a "small data" world

May 2015

Big data is everywhere. Once the preserve of innovators and technology entrepreneurs, big data analysis is now routinely used by a wide range of public and private sector organisations. It's a tool for planning, resource management and gaining competitive advantage.

For many, the potential benefit from analysing enormous big data datasets is undeniable. However, the view of Europe's most high profile data protection think tank, the Article 29 Data Protection Working Party (the WP29), is that the real value of big data remains to be proven. Either way, big data creates challenges for traditional data regulation regimes.

In its recent restatement about how data protection principles apply to big data[1], the WP29 made it clear that it sees the challenges for players in the big data arena. However, its conclusion remained that, ultimately, big data should be treated no differently to traditional data – "small data" one might say. Naturally, the WP29's views are only "soft law" and national regulators aren't obliged to follow them, but this sentiment did echo a similar view recently expressed by the UK data protection regulator, the Information Commissioner's Office (the ICO)[2].

So should the same rules apply and, assuming that they do, what's the best way to ensure that use of big data fits into a small data world?

Big data, big challenges

The challenges faced by big data players are both practical and regulatory in nature. For a start, part of the trouble with getting consensus on big data issues is that there is no fixed definition of big data. However, both the WP29 and the ICO recognise that, broadly, it will have the following characteristics:

- Volume – big data relates to the use of massive datasets, which are often so large that they cannot be analysed using traditional methods.
- Variety – big data often means pulling together data from a variety of different sources. For example, businesses may combine the data that they hold about their customers internally with datasets bought from third parties and/or data pulled from social media.
- Velocity – big data analytics often requires data to be analysed quickly or even in real time.

Other typical characteristics of big data include the use of algorithms to process

Any comments or queries?

Robert Johnson
Senior Associate
robert.johnson@rpc.co.uk
+44 (0)20 3060 6620

Lara White
Associate
lara.white@rpc.co.uk
+44 (0)20 3060 6418

1. Statement of the WP29 on the impact of the development of big data on the protection of individuals with regard to the processing of their personal data within the EU, adopted on 16 September 2014.
2. https://ico.org.uk/news/latest_news/2014/big-data-rules-28072014

the data; use of "all the data" (rather than just a sample of data); and the repurposing of data. Ultimately, the definition does not matter because all processing of personal data (whether big data analytics or not) must comply with the law.

The uses to which big data is put mean that it has the power to be both incredibly useful and incredibly intrusive. E-commerce and web analytics have been at the forefront of the commercialisation of big data and are swiftly moving on from the already-old world of online behavioural advertising to providing a much deeper customer experience. Mobile technology is leading the way, using big data analysis to predict customer wants and needs, and provide tailored, location-based services and "real time" offers and information at the point of user interaction.

More significant still is the rise of the digital health record. Vast amounts of data can be gathered (and shared) from data generated in devices and wearables. This might come from your phone, your watch, a wristband and even fabrics. Analysis of such enormous amounts of micro-data is not easy, let alone finding a commercial use for it. Nevertheless, it has been harnessed for uses ranging from the stock selection in your local supermarket, to optimizing drug development in clinical trials. The availability of big data also fuels the development of diagnostic software – soon we might face the prospect of algorithm-assisted diagnoses that will have the potential to revolutionize frontline healthcare.

But finding the meaningful needle in the big data haystack is not the only issue. Organisations that process personal data must not only ensure that they have the technical expertise to analyse and distil meaningful information from vast amounts of data but must also reconcile their use of big data with the requirements of European data protection laws. This is no small challenge, not least because these uses were genuinely in the realm of science fiction when current data protection regulation was drafted. The regulations themselves are widely regarded as not having kept up with the digital age.

Of course, many uses of big data analytics do not involve the processing of personal data or use anonymised data that no longer counts as personal data. However, increasingly the challenge posed by big data is that it can sometimes be difficult to establish with absolute certainty that data has been truly anonymised. As ever more detailed personal information enters the public domain, there is every chance that current notions of anonymisation will have to be revisited. In particular, where our unique health data and our own genetic fingerprints are concerned, at some point there could be so much of our own information in various databases that it will become almost impossible to ensure that it is effectively anonymised in the hands of any major player in the big data market.

The WP29 also states that it is important to ensure that big data is not used in an anti-competitive way. This is an issue in markets where companies have built up effective monopolies that are sustained by feeding off and exploiting the data that they collected at a time when regulators were less vigilant and customers were less savvy, and when most people did not know the value of the information that they gave away. The extent of the data held becomes an effective barrier to entry to new suppliers, who cannot offer the same level of customer experience without the benefits that big data provides.

Both the ICO and the WP29 have been clear in their views that, despite the specific challenges for controllers of big data, the rules do not change. Both the ICO and the WP29 emphasise that the practical difficulties are no excuse for non-compliance with data protection laws. Big data is, in the words of the ICO, "not a game to be played by different rules".

The WP29 certainly dips a toe in the water here by suggesting that big data might need

some "innovative thinking" on how some of the data protection principles should be applied in practice but is predictably woolly when it comes to the details of how this might happen.

Innovative thinking may therefore be hard to come by when in fact we are dealing with the same regime, but there are small adjustments that can be made to comply with data protection laws.

Big data, small adjustments

Fair processing and purpose limitation

A fundamental requirement of UK data protection law is that the processing of personal data must be fair and lawful. A key element of fairness is the need to be transparent with data subjects about how, and the purpose for which, data will be collected and used. This can be challenging, especially where data used for big data analytics is being used for a different purpose than the purpose for which it was originally collected.

The ICO and the WP29 offer some potentially useful guidance on this issue. Predictably, they both note that data protection law requires that personal data should be collected only for specified explicit and legitimate purposes. However, they also note that although this "purpose limitation" principle prohibits processing for any other purpose "incompatible" with the original purpose, it is not an absolute prohibition (the words used are "incompatible with", not "different from"). In the ICO's opinion, a key factor in deciding whether a new purpose is incompatible with the original purpose is whether the use of the data is "fair". "Fairness" is clearly a very subjective test, and links into other concepts such as ensuring that the data is adequate, relevant and not excessive. However, those who would argue for a more flexible approach to the use of big data believe that it might be better if the law only focussed on the "use" of the personal data, and somehow link what is "incompatible" with the original purpose to the level of risk or harm to the individual.

Naturally, this approach has its own risks. Just imagine how many organisations would happily justify extending the scope of their big data processing on the basis that "there's no real harm done". And who would police this? There are understandable concerns about private organisations making decisions on the use of information data in a big data context based on their own assessment of what could be harmful to the data subject. We have already seen the uneasiness of many commentators in relation to the take-down procedures under which Google is complying with the so-called "right to be forgotten" ruling in the Google Spain case[3]. Similar unease would doubtless follow a risk-based approach to the use of big data.

The fact remains that the best way to avoid the "purpose limitation" issue is to face it squarely and simply ensure, to the extent possible, that any consent received from the data subject is wide enough to cover the required processing without resorting to subjective judgements of "compatibility".

Keeping data relevant and not excessive

Of particulate relevance to organisations dealing with big data is the challenge of complying with the principle that the amount of data that they collect and process is "relevant and not excessive"[4]. This is fundamentally at odds with big data analytics, which tends to involve collecting and analysing as much data as possible from a variety of different sources. To assist in compliance with this principle, organisations need to articulate at the outset why they need to collect and process particular datasets. They also need to be clear about what they expect to learn or be able to do by processing that data and satisfy themselves that the data is relevant and not excessive in relation to that aim.

This DPA principle requires personal data not to be kept longer than is necessary for the purpose for which it is being processed. However, commercial organisations routinely keep data for longer than is necessary, in

3. *Case C-131/12 Google Spain SL, Google Inc. v Agencia Española de Protección de Datos, Mario Costeja González.*
4. Data Protection Act 1998, Schedule 1, Part 1, Principle 3.

order to maintain a competitive advantage and, if old data is valuable to the company, then why throw it away? However, no matter the value, it is important that organisations bear this principle in mind by having a proper data retention policy in place, or by properly anonymising any historic data that is being stored.

Security

The collection and analysis of big data potentially gives rise to increased security risks, in particular in relation to data breaches and leakages of information. However, some of these risks can be mitigated when organisations apply their existing risk management policies and procedures to big data analytics (including, where relevant, in relation to the use of cloud providers).

Small steps to compliance

The ICO makes some recommendations about the tools organisations can use to comply with data protection principles and ensure that people's privacy rights are respected. These recommendations are not unique to big data and are cited as good practice more generally. However, in practical terms, they may be more difficult to implement in the context of big data analytics in particular where the collection and storage of some of the data to be used in the big data analytics has already started.

Privacy impact assessment: it is important to assess the extent to which the processing of big data is likely to affect the individuals concerned before processing begins. The ICO recommends that organisations do this by undertaking privacy impact assessments and building privacy controls from the start. The ICO also notes that it is important that a number of people involved in big data projects (eg the organisation's data protection officer and other staff involved in the processing of data) understand privacy impact assessments and their use.

Privacy by design: the idea behind this is that if you are developing data analytics projects from scratch and building in privacy controls from the very start, then you can identify privacy risks and find creative technical solutions for dealing with such risks in a way that can deliver the real benefits to the project while protecting privacy. The solutions might include putting in place data minimisation, data segregation and purpose limitation controls. Anonymisation is an effective control if done properly, but organisations using anonymised data should be able to demonstrate that they have carried out a robust assessment of the risk of re-identification and have adopted solutions proportionate to the risk.

Conclusion

The key message from the regulators and the think tank is that there are no special rules for big data.

Given the challenges mentioned above, transparency between the data collector and the data subject will be the key to achieving compliance.

In particular, successful exponents of big data are likely to move towards a much more customer-centric model where data consents are not a mere footnote, but something the data subject actively seeks out and buys into. The key will be to demonstrate to data subjects the value of sharing their data and allowing them to be part of a big data analysis. There are plenty of people who are already happy to do this and, for every risk-averse lawyer who guards his privacy, there is a willing "millennial' who's happy to share details of their shopping habits in exchange for a cup of tea.

The EU Article 29 Working Party's Views on Regulating "Big Data" in a "Small Data" World was first published in World Data Protection Report Volume 15, Number 1 in January 2015.

About RPC

RPC is a modern, progressive and commercially focused City law firm. We have 77 partners and 560 employees based in London, Hong Kong, Singapore and Bristol.

"... the client-centred modern City legal services business."

At RPC we put our clients and our people at the heart of what we do:

- **Best Legal Adviser status every year since 2009**
- **Best Legal Employer status every year since 2009**
- **Shortlisted for Law Firm of the Year for two consecutive years**
- **Top 30 Most Innovative Law Firms in Europe**

We have also been shortlisted and won a number of industry awards, including:

- **Winner – Law Firm of the Year – The Lawyer Awards 2014**
- **Winner – Law Firm of the Year – Halsbury Legal Awards 2014**
- **Winner – Commercial Team of the Year – The British Legal Awards 2014**
- **Winner – Competition Team of the Year – Legal Business Awards 2014**
- **Winner – Best Corporate Social Responsibility Initiative – British Insurance Awards 2014**
- **Highly commended – Law Firm of the Year at The Legal Business Awards 2013**
- **Highly commended – Law firm of the Year at the Lawyer Awards 2013**
- **Highly commended – Real Estate Team of the Year at the Legal Business Awards 2013**

Areas of expertise

- Banking
- Commercial
- Commercial Litigation
- Competition
- Construction
- Corporate
- Employment
- Insurance
- Intellectual Property
- Media
- Outsourcing
- Pensions
- Private Equity
- Real Estate
- Regulatory
- Reinsurance
- Tax
- Technology

14668

Tower Bridge House
St Katharine's Way
London E1W 1AA
T +44 (0)20 3060 6000

Temple Circus
Temple Way
Bristol BS1 6LW
T +44 (0)20 3060 6000

11/F Three Exchange Square
8 Connaught Place
Central Hong Kong
T +852 2216 7000

8 Marina View #34-02A
Asia Square Tower 1
Singapore 018960
T +65 6818 5695