What future for Big Data?

Much has already been said about the opportunities and risks presented by 'Big Data' and the use of data analytics. But as the EU law-making institutions proceed to tighten the rules on data protection, will investment in data analytics still be as tempting a prospect? As we approach the final stages of the reform of the EU data protection regime, is there anything we should be learning from other jurisdictions to help inform the debate at EU level?

This article was first published in Privacy Laws & Business UK Report (Issue 81, September 2015).

A BRIEF REMINDER…

The term "Big Data" refers to large amounts of different types of data produced with high velocity from a high number of various types of sources.

The ability to examine and analyse Big Data in order to find unexpected connections and relationships or spot patterns between data is referred to as the science of 'Data Analytics'. Unlike the analysis of more traditional datasets, it is not always necessary to know why the data is linked in a certain way for value to be extracted from those connections and patterns.

HOW ARE THE EU REFORM PROPOSALS LIKELY TO IMPACT BIG DATA AND DATA ANALYTICS?

All three EU institutions involved in finalising the text of the draft General Data Protection Regulation (the 'Regulation') have now published their official versions of the Regulation. Taken as a whole the reform proposals create a number of obstacles to the processing of personal data for the purposes of Big Data analytics.

Profiling

The use of profiling will be more heavily restricted. Profiling is any form of automated processing of personal data intended, among other things, to analyse or predict people's performance at work, economic situation, location, health, personal preferences, reliability or behaviour. The EU Parliament's view is that profiling should be prohibited:

1 Communication from the Commission to the European Parliament, the Council, the European and social committee and the committee of the regions: Towards a thriving data-driven economy, 02 July 2014.

2 Originally proposed by the EU Commission in January 2012: Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (COM 2012)
• where it involves only sensitive personal data or the processing of employee personal data;

• where it results in discrimination; or

• where it is solely or predominantly automated and produces legal effects concerning individuals or similarly significantly affects them (unless such profiling is necessary for the performance of a contract, is authorised by Member State law or is based on the individual’s explicit consent).

Interestingly, the ICO has said it is not convinced that much online profiling has either a legal or a significant effect on the individual.

Perhaps recognising the value of data analytics to the EU economy, the EU Council takes a slightly more pragmatic view than the Parliament and would permit profiling that is not based solely on automated processing of personal data (i.e. where it is only predominantly based on automated processing).

Consent
Profiling is very likely to require the explicit consent of individuals, which may be difficult to obtain in practice in the context of Big Data. How does one explain clearly, succinctly and distinctly from all other processing being carried out how personal data will be used for the purposes of data analytics when the very point of running those algorithms over personal data is to spot previously unknown correlations or patterns of behaviour? There is also a risk that individuals will be presented with too many options to consent to, thus experiencing ‘consent fatigue’ and making consent less meaningful.

In its ‘Summary of feedback’ on Big Data, the Information Commissioner’s Office (ICO) was keen to highlight that, under current law, consent is not the only option to legitimise the processing of personal data for the purposes of data analytics. Organisations can rely on the ‘legitimate interests’ ground instead. This approach emphasises organisational accountability rather than individual responsibility, and is favoured by a number of organisations already. However, despite the ICO recently re-iterating the importance of the legitimate interests ground, it is by no means certain that this option will remain open to businesses carrying out profiling if the reform proposals are implemented as currently drafted.

Finally, mechanisms will need to be designed and implemented to deal with the refusal or withdrawal of consent. Such issues have to be considered at the outset of any Big Data project, in the same way that the project itself will have to be designed to take account of individuals’ rights to erasure and data portability.

Fair processing information
The EU Parliament and the EU Council would both oblige organisations to inform individuals that profiling of their personal data is being carried out and the consequences of such profiling.

The need for greater transparency has also been recognised in the report on ‘Responsible Use of Data’ by the UK’s House of Commons Science and Technology Committee in the context of social media and Big Data opportunities (and risks). However, greater transparency has to be balanced against (i) the issue of complex and lengthy privacy


\[^4\] April 2015


\[^6\] November 2014
policies which users do not read; and (ii) the difficulty in predicting the outcomes of any data analytics carried out. The ICO is currently reviewing its Privacy Notices Code of Practice and it is hoped that some of these issues will be addressed in the revised Code.

Territorial application and scope

It will become much harder to avoid the EU data protection regime biting on Big Data analytics. The main reasons for this are:

- The Regulation envisages a wider definition of ‘personal data’ (which may include IP addresses and cookies for example) and has an ambitious extra-territorial effect. It is interesting to note that the UK courts have already given consideration to the possibility that browser-generated information (i.e. information about one’s Internet usage) could amount to personal data. In addition, the EU Parliament and the Article 29 Working Party are both in favour of adding the concept of being ‘singled out’ to the definition of personal data. This all means that more businesses and commercial activities are likely to fall under the ambit of the European data protection rules than before.

- The lack of clarity surrounding the meaning of, and framework for, ‘pseudonymous’ data and ‘anonymous’ data may mean that companies will end up processing personal data when they hadn’t originally thought they would. Depending on the definition one relies on, it may be impossible to use ‘anonymised’ information without risking re-identification. This echoes the views of Hong Kong’s Former Privacy Commissioner, Allan Chiang, who warned that the assertion that privacy is no longer an issue for anonymisation is a “fallacy”.

Purpose limitation and data minimisation

The current principles of purpose limitation and data minimisation are set to remain in the Regulation. However, these principles sit awkwardly alongside Big Data analytics. The Regulation provides no further clarity on how the use of data analytics over vast quantities of data for the purposes of spotting previously unknown correlations can co-exist comfortably with the principle that only the minimum amount of data necessary must be processed for a given purpose. In its ‘Big data and data protection’ report, the ICO attempted to reconcile the two mindsets by reminding us that:

- The Data Protection Act 1998 does not require a new purpose to be the same or even compatible with the original purpose, rather that it must not be incompatible with it. Key factors to determine incompatibility include whether the original purpose is fair, whether the privacy of individuals is affected and whether the new processing was within their reasonable expectations.

- Big Data "raises questions not only about whether the data is excessive, but also whether it is relevant. […] Finding the correlation does not retrospectively justify obtaining the data in the first place." The ICO’s answer here is that "Organisations therefore need to be able to articulate at the outset why they need to collect and process particular datasets. The challenge is therefore to define the purposes of the processing and establish what data will be relevant to them." There is no further information on how to overcome such a challenge...

See Google v Vidal-Hall [2015] EWCA Civ 311
Sanctions and remedies
Although of more general concern, it is clear that the potential for maximum fines of up to 5% of annual worldwide turnover (or 100m Euros, whichever is the greatest) for non-compliance with the Regulation will contribute to making data protection compliance in the field of data analytics a key concern.

WHAT IS THE LATEST ON TIMING FOR THE REFORM?

The overhaul of EU data protection rules is nearing its final stages. The EU Parliament, the EU Council and the EU Commission are negotiating to reach agreement on a final text (the 'trilogue' process). The Parliament and the Commission have both clearly expressed their view that the reform of EU data protection laws is a 'package' deal, meaning that the proposed Directive on Law Enforcement also needs to be agreed. All eyes are therefore turned to the Council to see whether it can agree its position on that instrument in the autumn.

The Commission has emphasised that “We must conclude the ongoing negotiations on the data protection reform before the end of this year”\(^8\) and that the law-making institutions were "on track" to do so in 2015\(^9\). In his blog of 26 August 2015, David Smith, Deputy Commissioner of the ICO, stated that "If all goes according to […] plan, then we'll know pretty much what's going to be in the Regulation by the end of this year", although he anticipates that the final text would not be adopted before early 2016. This was also the view of a number of other data protection regulators and EU bodies present at the Privacy Laws & Business conference in Cambridge on 6-8 July this year.

Once adopted, the Regulation would not require any further implementing legislation, but would apply directly in Member States two years following adoption.

WHAT ARE OTHER JURISDICTIONS SAYING ABOUT THE USE OF BIG DATA?

In the US, policymakers are also focussing on Big Data. In May 2014, the White House published a report on Big Data\(^10\), acknowledging that “Big data tools offer astonishing and powerful opportunities to unlock previously inaccessible insights from new and existing data sets”. However, it also pointed out that “these developments challenge longstanding notions of privacy and raise questions about the "notice and consent" framework, by which a user gives initial permission for their data to be collected”. The report's recommendations include progressing the Consumer Bill of Rights (see below) and also, interestingly, advising the federal government to build the technical expertise to be able to identify practices and outcomes facilitated by big data analytics that have a discriminatory impact on protected classes.

The Obama administration published its draft Consumer Rights Bill in March 2015, with reference to Big Data included as one of the first items. The Bill has attracted a certain amount of criticism but as Brendan Lynch, Chief Privacy Officer for Microsoft, comments: "Not all will agree with every aspect of the proposal […] but it's a good place to start the conversation.”\(^11\)

\(^{9}\) Press release, Brussels 24 June 2015.
\(^{10}\) Big Data: Seizing Opportunities, Preserving Values (May 2014)
\(^{11}\) Microsoft blog, 27 February 2015.
Following the above, on 6 April 2015, the National Institute of Standards and Technology Big Data Working Group published a draft ‘Big Data Interoperability Framework’ to find consensus on fundamental questions that it believes are holding back progress and confusing users (such as defining the attributes of Big Data solutions and identifying the central scientific, technological and standardization challenges that need to be addressed). The report delves into the detailed technology and terminology whilst still addressing the privacy concerns that are relevant in the US. The Framework is particularly interesting in that it appears to bridge a gap between legislators and the technology involved.

The UK government has gradually been following suit. It is developing some of the infrastructure required to support the growing field of ‘Big and Open Data’, including through investments in a number of projects, such as the £42 million earmarked in 2014 for the ‘Alan Turing Institute for Data Science’ to research new ways of collecting, storing and analysing huge data sets. More recently, the House of Commons Science and Technology Committee, following on from its 2014 report referred to above, has launched a public inquiry on the opportunities and risks of Big Data, the support provided by the Government in this field, the skills required to make use of opportunities and public awareness of Big Data and informed consent.

Other jurisdictions have also been exploring the possibility of encouraging the use of data analytics whilst offering adequate levels of protection to individuals’ privacy. For example, Japan adopted an amendment to the Act on the Protection of Personal Information (Act No. 57 of 2003) on 3 September which may make it easier for Japanese businesses and investors to carry out data analytics. Although many of the proposed changes are likely to tighten the current rules on how personal data can be processed, the Japanese government hopes that by permitting the processing of Big Data without always obtaining the individuals’ consent, Japan will take full advantage of the opportunities presented by Big Data.

WHAT IS THE WAY FORWARD?

Big Data analytics are a huge opportunity, for the UK and EU economies as a whole, for businesses, but also for individuals.

"Overall, by 2020 big and open data can improve the European GDP by 1.9% (or 200bn Euros) - an equivalent of one full year of economic growth in the EU".12

The UK and the EU should ensure they are at the forefront of Big Data innovation rather than allowing other jurisdictions such as the US and Japan to take the lead. Businesses have much to gain by using data analytics, whether through tailored advertising, increased agility in their supply chains, or even boosting cyber security defences. As for individuals, they can benefit from better designed products and services, tailored to meet specific needs and with better availability. And that ignores potential benefits to society as a whole, such as medical research.

If one truly wishes to take advantage of all opportunities presented by Big Data, a prescriptive approach is not the answer as it will create more obstacles without necessarily providing a corresponding benefit to the individual. There is much to be said for the US approach, which focuses on the risk of harm to individuals, whether through discrimination or otherwise. For example, if an organisation has carried out a privacy impact or risk assessment

and is truly transparent about the data analytics it carries out, what is the real added value of obtaining the individual’s consent?

There is no clear and easy way forward for Big Data in the EU. However, whether one favours the US or EU approach, it is clear that the need to plan ahead and carry out privacy risk assessments will be unavoidable. Businesses must have a clear understanding of the sources of data they plan to use and how their use of analytics over personal data may impact individuals. This is part of the wider concept of ‘privacy by design’, which requires that privacy risks be considered at the outset of any project.

Privacy by design appears to be the main solution presented by law-makers and regulators to deal with the issues raised by Big Data now and in the future. Whilst a very useful concept, some challenges remain with its use in respect of Big Data. In particular, how can organisations ensure they have sufficient visibility as to the impact of data analytics on individuals when manipulating a technology that is inherently opaque? Is this truly a stumbling block for Big Data? It is vital here that the right people are involved in the debate. Lawmakers need to call on the relevant industries (such as Fintech, financial technology companies or software developers) to generate practical technical solutions and creative thinking. A dialogue between the two may be the only way to design a realistic regime that adequately protects individuals’ rights without hampering Big Data innovation. The US is already doing this and some EU countries are following suit (as mentioned above). This may be too little, too late. If the momentum to finalise the Regulation this year continues throughout the autumn, there is a serious risk that time will run out and we will be left with an imperfect, impractical and limiting solution for Big Data in the EU.