IT+ Our insight and thought leadership service | OUR WHITE PAPERS

Big Data and Data Protection

Richard Kemp

November 2014

IT+

IT law & regulation
Information & IP law
Internet & Telecoms

blog, alerts
white papers
webinars

KEMP IT LAW

kempitlaw.com

21 Napier Avenue London SW6 3PS Tel 020 3011 1667 Web www.kempitlaw.com

BIG DATA AND DATA PROTECTION

TABLE OF CONTENTS

BIG DATA AND DATA PROTECTION[1]

A. CONTEXT

1. **Introduction**. Big data (the harnessing, processing and analysis of digital data in huge and ever increasing volume, variety and velocity) has quickly risen up the corporate agenda as organisations appreciate that they can gain advantage through valuable insights about their customers and users through the techniques that are rapidly developing in the big data world.

Much big data, for example, climate and weather data, is not personal data, that is data that relates to an identifiable living individual. But for big data that is or could be personal data, data protection law and big data analytics collide in ways that the drafters of the EU Data Protection Directive 95/46[2] (**Data Protection Directive**) and even the draft EU General Data Protection Regulation[3] of January 2012 (**draft Data Protection Regulation**) could barely have foreseen.

Unsurprisingly, data protection authorities have taken the view that, like any other form of data processing, big data falls within the scope of data protection law and so must comply with the data protection principles. This has created some tension, which is likely to become more pronounced as big data techniques develop and use becomes ubiquitous at the same time as the EU undergoes a substantive revision of its own data protection framework.

This note will provide an overview of the data protection law issues that arise in the world of big data. For an analysis of the legal aspects of Big Data and Big Data management more generally (which reviews the intellectual property and contractual aspects and considers how Big Data is used and can be managed within the organisation), please see our White Paper at http://www.kempitlaw.com/category/white-papers/.

2. **What is big data?** Standard 2382-1[4] of the International Organization for Standardization/the international Electrotechnical Commission (ISO/IEC) defines:

> "**information** (in information processing) [as] knowledge concerning objects, such as facts, events, things, processes, or ideas, including concepts, that within a certain context has a particular meaning; [and]

[1] Reproduced from Practical Law with the permission of the publishers (www.practicallaw.com).

[2] Directive 95/46 of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and the free movement of such data, OJ L281 - http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:HTML

[3] 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 the free movement of such data (General Data Protection Regulation), COM(2012)11 Final and associated draft Directive, COM(2012) 10 final, 25 January 2012, http://ec.europa.eu/justice/newsroom/data-protection/news/120125_en.htm

[4] See https://www.iso.org/obp/ui/#iso:std:iso-iec:2382:-1:ed-3:v1:en. Information and data are used interchangeably in this paper.

> **data** [as] a reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing [which] can be processed by humans or by automatic means".

This ties in with the definition included in section 1(1) of the DPA, which defines data as information which is being processed by means of equipment that operates automatically in response to instructions given for that purpose, or is recorded with the intention that it should be processed by means of such equipment.

The term "big data" normally refers to high-volume, high-velocity and high-variety assets that demand cost effective, innovative forms of information processing for enhanced insight and decision-making[5]. Attributes of big data therefore include next generation data mining techniques that use more data, faster processing and new software:

- **Volume:** big data uses massive, diverse, complex, longitudinal, and/or distributed datasets that are generated by, or collected from, a variety of different devices, sensors and transactions. Examples include internet searches, credit and debit card purchases, social media postings, mobile phone traffic and location data, data collected by smart meters, vehicle sensors, or wearable computing devices. This results in large datasets that cannot be analysed using traditional data mining methods.
- **Variety:** big data often brings together data from different sources including both structured and unstructured data. For example, big data may combine information received from a social media feed and information gained from online behavioural tracking with point-of-sale data relating to online purchases to produce a richer picture of personal preferences and alignments for the purpose of direct marketing. Big data also facilitates the combination of the data controller's own information with externally sourced data.
- **Velocity:** big data is generally available for processing more quickly (often in real time).

It is therefore characterised by a need for new tools and methods like powerful processors, software and algorithms to handle those massive, highly variable and real-time datasets, that go beyond traditional data mining tools designed to handle mainly low-variety, small scale and static datasets.

3. **New tools and methods**. Big data relies on:

- **Aggregation**:
  - size: vast volumes of digital data;
  - shape: large variety of formats (text, image, video, sound, etc.);
  - structure: a combination of unstructured (typically, 80%) as well as structured (typically, 20%) data; and
  - speed: data collected, generated and processed at a faster velocity.

[5] See for example http://www.gartner.com/it-glossary/big-data

- **Analysis:**
  - aggregated datasets are analysed on a real-time rather than batch basis;
  - analysis is performed by quantitative analysis software using artificial intelligence, machine learning, neural networks, robotics and algorithmic computation; and
  - analysis enables a shift from retrospective to predictive insight.
- **Increasing value:**
  - facilitates small but constant, fast and incremental business change; and
  - enhances competitiveness efficiency and innovation and the value of the data so used.

4. **What is data in legal terms?** Unlike real estate for example, information and data as expression and communication are limitless and it would be reasonable to suppose that subjecting information to legal rules about ownership and use would be incompatible with its nature as without boundary or limit. Yet digital information is only available because of investment in IT, just as music, books and films require investment in creative effort.

In legal terms, this means that while there are no rights or obligations in data, extensive rights and obligations arise in relation to data. In the UK, it was held in the case of *Oxford v Moss*[6] that confidential information in an exam question was not "intangible property" within the meaning of Section 4(1) of the Theft Act 1968 and so could not be stolen. However, the rights and duties that arise in relation to data are both valuable and potentially onerous. As an area of law they are also developing rapidly at the moment as big data techniques become more prevalent.

These rights and duties arise through intellectual property rights (IPR), contract and regulation. From the perspective of a commercial actor, they are important as:

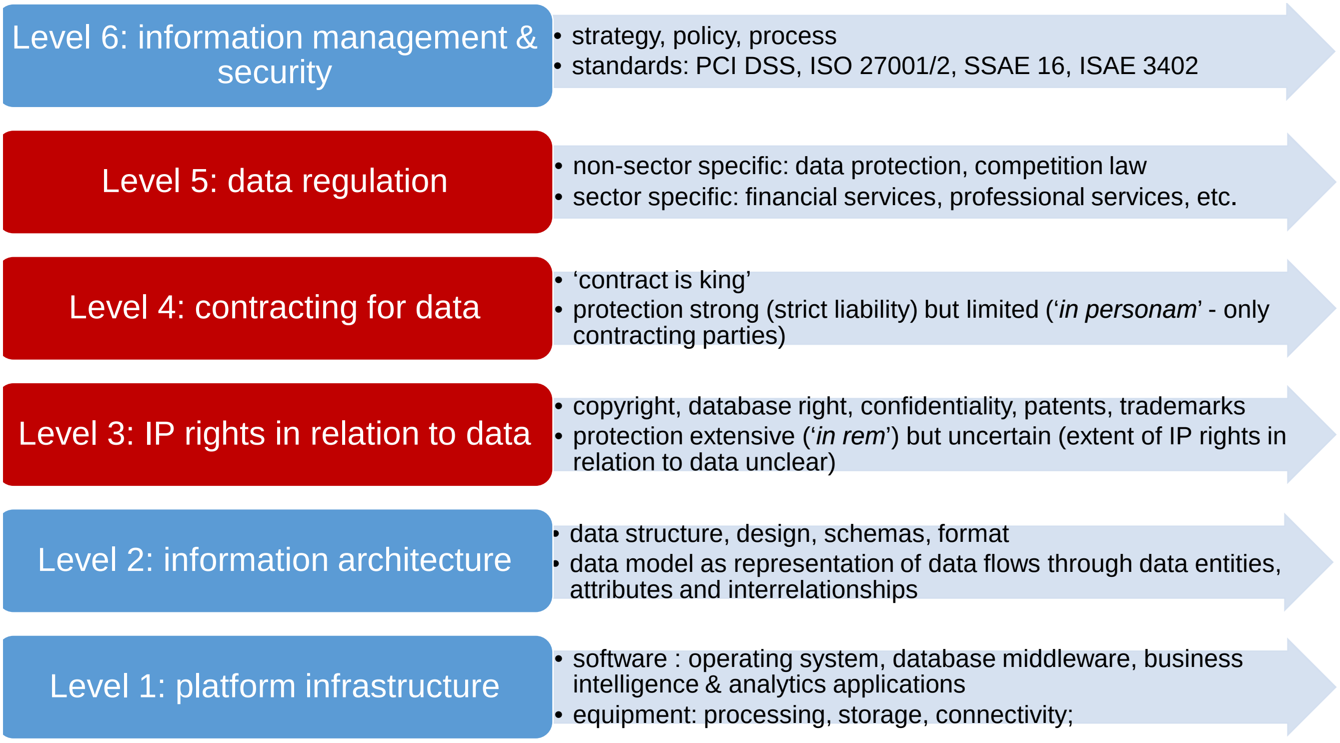
- They can increasingly be monetised (in the case of IPR and contract).
- A breach can give rise to extensive damages and other remedies (for IPR infringement and breach of contract), fines and other sanctions (breach of regulatory duty).

Current developments in each of these areas mean that data law is emerging as a new area in its own right around these three constituents of IPR, contract and regulation. (For further detail, please see our White Paper referred to at paragraph 1 above).

5. **Towards a common legal analytical framework for data**. Big data is regulated through a set of architectural, organisational and legal constraints that determine the way in which it can be used. The legal constraints of IPR, contract and regulation must therefore be viewed in conjunction with other constraints like platform infrastructure, information architecture, information management and security. The data protection framework is the most important component of data regulation. Please see the Figure below for a representation of this common analytical framework.

[6] [1979] Crim LR 119, where it was held that confidential information in an exam question was not ‘intangible property’ within the meaning of Section 4(1) of the Theft Act 1968 and so could not be stolen

Figure: Towards a common legal framework for Big Data



Level 6: information management & security
- strategy, policy, process
- standards: PCI DSS, ISO 27001/2, SSAE 16, ISAE 3402

Level 5: data regulation
- non-sector specific: data protection, competition law
- sector specific: financial services, professional services, etc.

Level 4: contracting for data
- 'contract is king'
- protection strong (strict liability) but limited ('*in personam*' - only contracting parties)

Level 3: IP rights in relation to data
- copyright, database right, confidentiality, patents, trademarks
- protection extensive ('*in rem*') but uncertain (extent of IP rights in relation to data unclear)

Level 2: information architecture
- data structure, design, schemas, format
- data model as representation of data flows through data entities, attributes and interrelationships

Level 1: platform infrastructure
- software : operating system, database middleware, business intelligence & analytics applications
- equipment: processing, storage, connectivity;

B. BIG DATA AND DATA PROTECTION IN VERTICAL SECTORS

The insights gained from the analysis of big data can be used to bring about small, incremental changes in behaviour that are nonetheless commercially valuable.

6. **The healthcare sector**. Healthcare is the sector both where adoption and use of Big Data is likely to have the greatest impact on people's daily lives and where information is about identifiable living individuals. In its January 2013 report '*The 'big data' revolution in healthcare*'[7], consultants McKinsey & Co pointed to four changes that were creating a tipping point for innovation in healthcare around Big Data:

- **Demand-side pressures.** Users of health data of all kinds (including medical professionals, administrators, pharmaceutical companies and IT providers) call for better data as cost pressures intensify, structural reforms continue and early movers and adopters demonstrate advantage;
- **Increased availability**. On the supply side, national collections of patient, clinical and treatment outcome data are starting to become available in particular areas (for example, the aggregated national collection relating to the outcomes of cardiac heart treatment and procedures in the UK);

[7] http://www.mckinsey.com/insights/health_systems_and_services/the_big-data_revolution_in_us_health_care.

- **Increased investment**. investment is gathering pace in technical developments for aggregating and anonymising data from individual hospitals and treatment centres and in the BIA software tools that generate insights from them; and
- **Data sharing between public and private entities.** Governments are catalysing market change by their continuing commitment to making data held by public bodies available for commercial and certain public interest purposes (for example, the UK government's open data and care.data schemes). The creation of new interoperability standards also encourages private sector participation.

Although the McKinsey report focused on the USA, these change agents are even more powerful in the UK through the NHS which produces significant amounts of big data.

7. **The insurance sector**. Big data enables the insurance sector to assess risk much more precisely than in the past. Subject to some statutory constraints, more specific data about the insured and the risk insured enable individualised pricing and the refusal of policies to individuals and businesses considered high risk.

As well as the traditional "top down" statistical and actuarial techniques of risk calibration and pricing, insurers can now rely on actual data relating to the insured concerned. For example, in vehicle insurance, location data from the driver's mobile can track the insured's movements, and telematics data from on-board IT can show how safely they were driving. Similarly, smart domestic sensors can help improve responsiveness to the risk of fire, flooding or theft at home. Crime maps can help to calculate the risk of burglary down to the level of individual streets or buildings, and health apps and "wearable technologies" (body-borne small electronic devices) can provide information relevant to health and life insurance.

These examples (data sourced remotely from telematics, location services, home sensors, publicly available environmental data and wearables) are early illustrations of big data (and also the "Internet of Things") in consumer insurance. They will over time have a material impact on the pricing of vehicle, home, life and health policies.

The use of big data by the insurance sector therefore especially illustrates the ethical questions that arise and the tension that exist between the commercial benefits of big data and the privacy of the individual whose personal data makes those benefits possible. The availability of big data to business and the state and the constraints on the use of that data pose regulatory challenges that become greater the more sensitive the data in question is (for example, patient data or data about genetic pre-dispositions) and the more the data is likely to be used for commercial decision-making processes that could be seen as discriminatory.

8. **The air transport industry**. The air transport industry (ATI) has grown up with computerisation and standardisation as key components in getting passengers (three billion globally in 2012) and their baggage to the airport of departure, on to the plane, and to and from the airport of arrival. Airlines and other ATI companies therefore generate and hold vast amounts of personal data about passengers' preferences during all stages of their journey. These enable ATI players to develop insights about their customers that might give them a competitive advantage.

In particular, passenger data emerges as both a key enabler and regulatory issue for big data in the ATI. This has triggered a number of policy initiatives around the standardisation of passenger consent to use of their personal information in and around the airport and their journey more generally. The increased use of mobile phones as data source, data store and processing point also contribute to this development. For the airline passenger, the mobile wallet facilitates paperless ticketing and boarding passes and its NFC (near field communication) feature enables mobile check-in. These improve efficiency and reduce time and costs at the point of sale and in the airport. In the future, use of this data is likely to segue to m-commerce and shopping at the airport (by enabling improved mobile marketing opportunities) and beyond to smart cities. This means that much more information not just about the passenger's movements, but also about his retail behaviour could now be available for commercial exploitation.

At the same time, the use of passenger name record data to combat terrorism and serious crime constitutes another regulatory concern. The long-term retention, sharing and use of this data for those purposes without sufficient safeguards has been controversial and has attracted criticism from privacy groups and some courts.

9. **The public sector**. As with all developed countries, UK government departments have huge and growing databases about UK citizens. Those departments increasingly master their own digital data while central government starts to move towards increased data sharing. The UK government's "data estate" is therefore becoming a valuable national asset.

Managing the UK's data estate raises complex policy questions about:

- Security, growth, maintenance and monetisation of the data.
- Reconciliation of all the competing interests, including:
  - the protection of privacy and other individual liberties;
  - public and national security;
  - crime and fraud prevention;
  - commercial interests;
  - safeguards against state overreaching; and
  - maximising the benefits of technological progress for citizens.

In 2014, the publication of the Cabinet Office's Data Sharing Policy Team's data sharing discussion document once again raised public interest in the government's agenda in this area[8].

The discussion document advocates an open policy-making approach, which balances the delivery of better public services with citizens' concerns about their privacy. The document proposes to remove barriers to sharing or linking different datasets, while taking steps to safeguard individuals' rights. Among other things, the government plans to develop further proposals of the Administrative Data Taskforce, a collaborative initiative between the Economic and Social Research Council, the Medial Research Council and Wellcome Trust, for two models: the Trusted Third Party and the

[8] http://datasharing.org.uk/current-proposals/

Firewall Single Centre[9]. Both models would allow data sharing for cross-linked research on de-identified data while restricting access to and use of identity data to the extent needed to cross-link the datasets concerned. The report also proposes the following structural safeguards:

- Accreditation and registration of projects and individuals having access to de-identified data.
- A formal process to be carried out by the UK Statistics Authority to accredit the four Administrative Data Research (ADR) Centres that form part of the ADR Network the UK government set up as a vehicle for public sector big data.
- Compliance with the Data Sharing[10] and Anonymisation[11] Codes of Practice published by the Information Commissioner's Office (**ICO**).

C. BIG DATA AND DATA PROTECTION: RECENT DEVELOPMENTS

10. **The US National Intelligence Council's December 2012 Report**[12]. The report signposts big data's direction of travel and articulates a focus on data solutions and big data as a key IT driver over the next two decades. It highlights the fact that:

- processing powers and storage are becoming almost free;
- networks and the cloud provide global access to data; and
- pervasive services, social media and cyber security will open up large new markets.

The report also emphasises the continuing trade-offs that individuals and organisations will need to make between utility and privacy of personal data.

11. **UK government 2013 strategy paper**[13]. The strategy paper presents a positive view of the UK's ability to seize the data opportunity. It stresses the government's determination to position the UK at the forefront of the new "data revolution". It plans to address privacy and data protection issues through a clear and pragmatic policy that ensures public trust in the confidentiality of their data, while increasing the availability of data to maximise its economic and social value. The paper also acknowledges the conflicting views around the review of the EU data protection regime. It makes it clear that while the UK supports the need to bring data protection rules in line with the reality of the 21st century, the UK government does not believe that the European Commission's proposals for a General Data Protection Regulation strike the right balance between privacy and innovation. In

[9] Report of the Administrative Data Taskforce (a collaborative initiative between the Economic and Social Research Council, the Medial Research Council and Wellcome Trust) on Improving Access for Research and Policy - http://www.esrc.ac.uk/_images/ADT-Improving-Access-for-Research-and-Policy_tcm8-24462.pdf.

[10] http://ico.org.uk/for_organisations/data_protection/topic_guides/data_sharing.

[11] http://ico.org.uk/for_organisations/data_protection/topic_guides/anonymisation. See below

[12] 'Global Trends 2030: Alternative Worlds', http://globaltrends2030.files.wordpress.com/2012/11/global-trends-2030-november2012.pdf

[13] 'Seizing the data opportunity – A strategy for UK data capability', 30 October 2013, https://www.gov.uk/government/publications/uk-data-capability-strategy

particular, it opposes overly prescriptive regulation "that increases red tape and costs for businesses, the public sector, and for regulators themselves".

12. **The Executive Office of the US President's May 2014 report - Big data: Seizing Opportunities, Preserving Value**[14]. The report focuses on the way in which big data will transform everyday life and how it will alter the relationships between government, citizens, business and consumers. It considers big data and privacy both in the public and the private sector and concludes that the existing US "notice and consent" approach to data privacy may have to be reviewed in the light of big data. It highlights that in "a technological context of structural over-collection, in which re-identification is becoming more powerful than de-identification, focusing on controlling the collection and retention of personal data, while important, may no longer be sufficient to protect personal privacy".

13. **The European Commission's 2014 Communication: Towards a Thriving Data-Driven Economy**[15]. The Communication sets out a number of activities it considers necessary for the EU to be able to seize big data opportunities and compete globally in the data economy. They include:

    - A data-friendly legal framework and policies. Policies on issues relevant to big data like interoperability, data protection, security and IPR should lead to more regulatory certainty for business and create consumer trust in data-technologies.
    - Concluding the reform of the EU legal frameworks for data protection and network and information security.
    - Supporting the exchange and co-operation between the relevant enforcement authorities for data protection, consumer protection and network security.

14. **The UK ICO's 2014 report: big data and data protection**[16]. The report applies the relevant principles of the Data Protection Act 1998 (**DPA**) to the different aspects of big data and provides useful practical pointers on how to address them.

15. **The Article 29 Working Party's September 2014 Statement**[17]. The statement acknowledges that the challenges of big data might require innovative thinking on how some of the key data protection principles are applied in practice. However, the working party confirms its view that the EU data protection principles, as they are currently enshrined in the Data Protection Directive, remain valid and appropriate for the development of big data, subject to further improvements to make them more

[14] 'Big Data: Seizing Opportunities, Preserving Value', May 1, 2014, http://www.whitehouse.gov/issues/technology/big-data-review

[15] 'Towards A Thriving Data-Driven Economy', COM(2014) 442 Final, 2 July 2014, https://ec.europa.eu/digital-agenda/en/news/communication-data-driven-economy

[16] 'Big data and data protection', 28 July 2014, http://ico.org.uk/for_organisations/data_protection/topic_guides/big_data

[17] 'Statement of the Article 29 Working Party on the impact of the development of big data on the protection of individuals with regard to the processing of their personal data in the EU, 16 September 2014', http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/index_en.htm. The Article 29 Working Party also considered big data at Annex 2 (pages 45 to 48) of its Opinion 03/2013 of 2 April 2013 on purpose limitation (see previous hyperlink).

effective in practice. The working party also makes it clear that the data protection rules and principles are applicable to all processing operations, starting with collection in order to ensure a high level of data protection. This rebuts the suggestion often made by the data industry that regulation of big data should place limits on the use rather than the collection of data. The working party stresses the importance of creating and keeping consumer trust. It believes that compliance with the EU data protection framework and investment in privacy-friendly solutions is essential not only to protect individual rights but also to ensure fair and effective competition between economic players on the relevant markets. The statement particularly highlights compliance with the purpose limitation principle as one factor to ensure that companies which have built monopolies or dominant positions before the development of big data technologies hold no undue advantage over newcomers to these markets.

D. BIG DATA AND THE DATA PROTECTION PRINCIPLES

The application of the existing data protection principles to big data technologies raises a number of compliance issues.

16. **Fair Processing**. Under the first data protection principle (paragraph 1, Part 1, Schedule 1, DPA), personal data must be processed fairly and lawfully in circumstances where one of the conditions in Schedule 2 of the DPA is met.

To determine whether personal data is processed fairly, it must be established:

- How the personal data was obtained.
- Whether the data subject was misled or deceived about the purpose of the processing.
- Whether the data controller informed the data subject at the time of collection or as soon as possible thereafter about his identity, the purpose for which he intends to process the data and any further information which is necessary under the specific circumstances to enable the processing to be fair.

In the ICO big data report, the ICO emphasises the importance of fairness, transparency and meeting the data subject's reasonable expectations in big data processing. It states that transparency about how the data is used will be an important element when assessing compliance. It also highlights the need to consider the effect of the processing on the individuals concerned.

The central rationale of big data is to discover hitherto unobserved correlations between different datasets and to provide actionable insights from these (by definition) "unexpected" results. Reconciling the broad concept of fairness in data protection terms with this central feature of big data will therefore lie at the heart of an organisation's data protection compliant big data use.

The ICO big data report illustrates how unexpected big data correlations may be by quoting the well-known example of US retailer Target whose marketing department's analytics found a pattern between the purchase dates of certain products by expectant women and their due date. When Target sent marketing literature for baby-related products to a High School student in Minneapolis, her father complained to Target about her receiving this kind of marketing. The father was not then aware of his daughter's pregnancy and subsequently apologised to the store.

17. **Consent**. Consent is only one of the legal grounds included in Schedule 2 of the DPA that ensures that personal data is processed fairly and lawfully. However, in the context of big data is it the ground most often relied upon. The bar for consent has risen markedly in recent years following detailed guidance issued by both the Article 29 Working Party and the ICO after the introduction of the new cookie regime as part of the revision of the E-Privacy Directive (2002/58/EC)[18] in 2010. That guidance is also expected to inform the relevant provisions that will be adopted as part of the new Data Protection Regulation. However, for present purposes consent must be freely given, specific and informed.

    The ICO big data report makes it clear that just because big data is complex, this is not an excuse for failing to obtain consent where it is required. In particular, organisations must "find the point at which to explain the benefits of the analytics and present users with a meaningful choice - and then respect that choice when they are processing their personal data".

    Where an organisation is relying on consent in the big data context, people must be able to understand how the organisation will use their data and there must be a clear indication that they consent to that use. If an organisation has collected personal data for one purpose and then decides to start analysing it for completely different purposes (or to make it available for others to do so) then it needs to make its users aware of this and, where necessary, obtain their further consent.

    Consent must be appropriate as well as reasonable, meaning that while it may be reasonable for organisations to use consent as a condition for processing in a big data context, it may not be the most appropriate legal ground. In this context, organisations should also remember that consent may be withdrawn. They should therefore consider whether reliance on one of the other grounds contained in Schedule 2 of the DPA is a more suitable long-term strategy to ensure the data protection compliance of their big data activities.

18. **Purpose limitation**. The purposes limitation principles provide that personal data must be collected for specified, explicit and legitimate purposes, and not be further processed in a way incompatible with those purposes (paragraph 2, Part 1, Schedule 1, DPA).

    Issues may arise where personal data obtained in relation to providing a particular service is then used for another purpose that does not necessarily fit with the original purpose or that is not "intrinsic" to the provision of that service. In the ICO big data report, the ICO suggests that while, for example, a retailer's use of loyalty card data for market research would be permissible, a social media provider's use of user profile data for the same purpose would not be. The challenge of determining which purposes are compatible with the purpose for which the data was originally collected is likely to increase with big data, and the nexus between compatible and incompatible purposes will the subject of much discussion.

    The ICO big data report makes it clear that the purpose limitation principle does not bar the repurposing of personal data as a matter of principle, or prescribe that the new purpose and original purpose must be identical. Instead, it reminds data controllers that the new and original purposes must not be incompatible. Where the new purpose is incompatible with the purpose for which the

[18] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32002L0058:EN:HTML

data was originally collected, then unless data subjects are informed of this and asked to give their consent, processing is unlikely to be fair or compliant with the purpose limitation principle. Where the new purpose would be otherwise unexpected, and it involves making decisions about data subjects as individuals, then in most cases the organisation concerned will need to seek specific consent, in addition to establishing whether the new purpose is incompatible with the original reason for processing the data.

19. **Data minimisation**. The DPA encompasses the principle of data minimisation through the combination of the third and fifth data protection principles. In particular, personal data must be adequate, relevant and not excessive and must not be kept for longer than is necessary for the purposes for which they are processed. Data minimisation therefore fundamentally collides with the concept of big data, which involves collecting as much data as possible.

    The ICO big data report advises that organisations therefore need to be able to articulate at the outset why they need to collect and process particular datasets. Although big data may discover unexpected correlations that may be useful in a commercial context, finding those correlation does not retrospectively justify obtaining the data in the first place. As data storage costs reduce, and the cost of destroying data outweighs that of keeping it, a similar point arises in relation to the length of time for which personal data is kept.

20. **Rights of the data subject**. Individuals are entitled to compensation from data controllers for damage caused by any breach of the DPA (section 13(1) DPA). Compensation is also available, in certain cases where the individual suffers distress as a result of the breach. However, section 13(2)(a) of the DPA makes it clear that data subjects can only be awarded compensation for distress if the can show that they have also suffered financial loss. These provisions have attracted widespread criticism from, among others, Leveson LJ in his report on culture, practices and ethics of the press, who proposed that the DPA should be amended so that compensation can be awarded regardless of pecuniary loss. The ICO also welcomed Leveson’s proposals while highlighting that the European Commission has questioned whether the UK has properly implemented the Data Protection Directive in this regard.

    However, despite this criticism, the need for the existence of either individualised financial harm or distress is not generally questioned. This leads to problems for individuals in a big data context, where those kinds of harm may be difficult to prove and where data subjects may instead suffer more intangible and long-term harms that they are not yet able to perceive or measure. Those harms could include:

    - Price or other types of discrimination.
    - Loss of employment opportunity.
    - An inability to purchase certain insurance products or to secure credit.
    - An overall restriction on the types of goods and services organisations offer to individuals.

    It would therefore have been helpful for the ICO to have expressed its views on both the technical legal questions of quantifying the harm that individuals may suffer and the corresponding questions of compensation and liability that may arise from using non-DPA compliant big data analytics. As big data is all about fast, incremental changes one view would be that any single big data DPA breach may not be material. Such a view would clearly undermine the substantive application of the DPA to big data if its effect were to be that remedies were restricted.

On the other hand, the trend in data protection law is for the rights of data subjects and others to expand. Individual rights introduced or extended by the Data Protection Directive and DPA (for example, the rights in relation to subject access, prevention of processing, compensation and rectification, blocking and destruction of personal data (sections 7-14, DPA) will be expanded further in the Data Protection Regulation. Similarly, regulators' powers are also increasing and will be extended further by the Regulation. It is hoped that the ICO will provide guidance as to how it sees the DPA liability regime operating in the big data world.

E. ADDRESSING DATA PROTECTION ISSUES IN BIG DATA

21. **Getting the right consent to the right processing at the right time**. The ICO big data report advises that data controllers may be able to adopt a process of "graduated consent". This would allow data subjects to give consent (or not) to different uses of their data throughout their relationship with the data controller, rather than having a simple "binary" choice at the beginning. For example, they could give an initial consent to opt in to the system and then separate consent for their data to be shared with other parties. Furthermore, data controllers could consider a value exchange, that is, they could offer data subjects some additional benefit in return for giving their consent.

22. **Promoting transparency and notice**. The ICO big data report reminds data controllers that privacy notices can support transparency at an early stage. Referring to its own privacy notices code of practice[19] the ICO encourages organisations carrying out big data analytics to inform data subjects comprehensively about what they are doing with their data. In particular, data controllers should ensure that they comply with the fair information requirement (paragraphs 2 and 3, Part 2, Schedule 1, DPA). They can do this by stating:

- The identity of the organisation collecting the data.
- The purposes for which they intend to process it.
- Any other information that needs to be given to enable the processing to be fair.

23. **Anonymising data**[20]. Data that does not identify a living individual is not subject to the provisions of the DPA. Data controllers should therefore make use of anonymisation techniques to minimise their own compliance burden. However, the very nature of big data means that absolute anonymisation may not be possible. Organisations using anonymised data should therefore focus on mitigating the risks of re-identification to the point where the chance is extremely remote. In this context, the ICO big data report also reminds data controllers that they must be able to demonstrate that they have carried out this robust assessment of the risks of re-identification and have adopted solutions proportionate to the risk. This may involve a range and combination of technical measures

[19] http://ico.org.uk/for_organisations/data_protection/topic_guides/~/media/docum

[20] See also http://ico.org.uk/for_organisations/data_protection/topic_guides/anonymisation for ICO/s Anonymisation Code of Practice.

such as data masking, pseudonymisation, aggregation and banding, as well as legal and organisational safeguards.

24. **Carrying out a privacy impact assessments**[21]. Data controllers should anticipate data protection issues arising from their use of big data technologies by carrying out privacy impact assessments and using privacy by design as tools before processing begins. These techniques should be used to assess how big data analytics is likely to affect the individuals whose data is being processed and whether processing is fair.

    The ICO big data report refers data controllers to its code of practice on conducting privacy impact assessments, which gives practical advice on how to do this and links the privacy impact assessment to standard risk management methodologies.

    The ICO big data report highlights that assessing privacy risk involves:

    - **Being clear at the outset about the benefits and aims of the big data project, as well as the impact on individuals' privacy.** In many cases, the benefits are to the organisation that is proposing to process the personal data, but it is important to factor in benefits that may accrue to individuals or to society more broadly. When solutions to mitigate privacy risk have been identified, it is necessary to assess whether the final impact on those individuals, after those solutions have been applied, is proportionate to the aims of the project.

    - **Ensuring that a range of people involved in big data projects understand PIAs.** The organisation's data protection officer may need to co-ordinate the process but other staff, like data scientists, need to understand how to apply PIA techniques to their work. The ICO stresses that for a PIA to be effective in a big data environment those who have the technical expertise in designing and applying algorithms must have an understanding of privacy impact.

25. **Building trust through ethical use of big data technologies**. To foster consumer trust in big data technologies, organisations should place their big data activities in a wider and essentially ethical context. This means that they should not just ask what they can do with certain data sets (both technologically and with regard to regulatory compliance) but whether they should process the data for this purpose ("is it what customers expect, or should expect?").

26. **Observing good information governance**. Organisations should ensure that they put in place good information governance to support a trust-based ethical approach to big data. In practice this means that they should adopt a structured approach to data protection, determining the necessary level of information governance according to data types, data sources and data use and focusing on data security, trust, validation and management efforts.

    Noting a growing emphasis on the issue of data quality and information governance in relation to big data analytics, the ICO big data report refers to a report on big data published by Forrester Consulting in August 2013[22]. The report presented the results of an online survey conducted in summer 2013 of 512 respondents aimed at evaluating their approaches, practices and perceptions around data

[21] http://ico.org.uk/for_organisations/data_protection/topic_guides/privacy_impact_assessment

[22] http://www.ibmbigdatahub.com/whitepaper/big-data-needs-agile-information- and-integration-governance

governance. In particular, the report maintains that understanding the source of data as well as the type allows organisations to classify the data within the contexts of business use and value.

The report introduces the concept of four context-driven information and integration governance data zones:

- Organisations should use **tight governance** when using data in business processes, decision-making, or meeting regulatory requirements.
- **Casual governance** can be sufficient for data coming into the organisation but not used frequently or widely.
- **Validation** can act to ensure a baseline of conformity.
- A **chaotic state** of governance may only be allowed if data is not ready to be incorporated into business use.

F. CONCLUSION: THE FUTURE FOR BIG DATA

27. **The future for big data**. Big data and data protection engage in two principal ways:

- Through the application of established and developing data protection techniques to the rapidly evolving big data world.
- Through the creation of ever larger amounts of “personal data” as big data techniques allow organisations to combine different data sets. This increases the likelihood that the resulting data is capable of identifying living individuals in ways impossible until now. As a result, the capacity to mine and analyse datasets of ever increasing volume, variability and velocity ever more effectively increases and the volume of personal data will increase exponentially.

In areas of law and regulation impacted by innovation, regulators and legislators typically have to respond to technology change through the evolution of the established laws at their disposal. Consequently, the ICO and other data protection authorities are addressing big data by further developing existing tools like notice and consent, anonymisation and privacy impact assessments. While these techniques address the first way in which big data and data protection engage, they do not yet address the bigger question of the massive increase in personal data that big data will produce over time and the regulatory implications of that change.

Richard Kemp,
Kemp IT Law,
London,
November 2014
richard.kemp@kempitlaw.com
Tel: 020 3011 1670