Big Data and Financial Services

Big data is a hot topic in financial services. Big data can be used to develop, target and price a wide range financial products from life insurance to car insurance through to consumer lending and other financial products. It brings big opportunities but also significant legal issues. Simon Stokes, a Partner in the firm's Financial Technology practice, recently spoke on the evolving big data legal and regulatory framework at the 3rd Winchester Conference on Trust, Risk, Information and the Law sponsored by Blake Morgan. In this article he explores the implications this evolving legal and regulatory framework has for the financial services sector.[i]

What is "Big Data" and why does it matter?

Big Data is a buzzword and means many things to many people – the FCA have defined it to mean:

- The use of new or expanded data sets
- New technologies to generate, collect and store data
- Sophisticated analytical techniques (for example machine learning).

Another way to think of it is a description of the accumulation and use of vast and complex information databases using multiple servers that can't be carried out using just a single server database or PC.

For example in a financial services context the following involve big data:

- A major retail bank using its extensive customer database to profile and target its customers with special offers and products
- A global bank creating a financial model taking data from stock market indices, GDP/economic data, trading data, consumer data (RPI/inflation etc)
- An insurer using a large volume of consumer data (from mobile apps/telematics boxes in cars, information from the "Internet of Things" (objects connected to networks and generating data), claim information for a range of products, price comparison sites, credit reference agencies, social media information, and consumer and civil society organisations) to assess risk and set premiums.

Big data is often characterised by "three V's":[ii]

- Volume – big data involves significant amounts of data
- Variety – in big data data is combined from a variety of sources – weather data, online traffic data, UGC, location/geo data, RFID data, point of sale data and so on
- Velocity – big data is being generated at speed, collected at speed and processed and analysed at speed.

Big data allows organisations using it to create information about data that were not necessarily apparent or intended in the source information. For example take the communications data from millions of phone calls – analyse this and it's possible to discover a huge combination of factors relating to the nature of communications, and user relationships and behaviours.

Commentators see big data having a significant role to play in financial services, for example to: detect patterns of financial or insurance fraud, to combine trader performance data, market data, unstructured news, user data, and general ledger data to gain previously impossible insights. This can enable 'real time' decision making power that makes a difference between winners and losers in the financial markets.[iii]

Big data poses some significant legal challenges and its use is already being scrutinised by the FCA and other regulators including the CMA (competition and consumer law) and ICO (data protection).[iv] There are three major legal issues around big data. First protecting it – that's a matter of intellectual property law. Second fairly and lawfully collecting, processing and using it – that's a matter of data protection/data privacy law. Third how it is used which also falls within competition and consumer law.

Big Data as Intellectual Property

In English law judges have long held that information as such isn't "property". But this doesn't mean data can't be protected as intellectual property (IP). This can be done in six ways:

- By keeping the data (and the proprietary algorithms and code used to mine the data) and any database formats confidential
- Through copyright protecting elements of the data and databases themselves if they satisfy the test for copyright (they need to be original works of authorship) - i.e. they constitute the author's own intellectual creation
- Whether or not copyright protection is available the investment spent in obtaining, verifying and/or presenting the data can potentially be protected by the "sui generis" EU database right under the Database Directive
- Through contract – ensuring that you only grant access on contractual terms which protect and licence your IP.
- Through patents where you have devised a technical innovation which has a technical effect
- Through trade marks and domain names where big data is used as part of a branded service

To benefit from this protection you need to have policies and procedures in place to protect your IP. This means ensuring you address the data lifecycle by answering three broad questions:

- How is the data originated and sourced?
  - How is it created and by whom? Are they your employees or third parties (including contractors or consultants). Are you licensing in data – on what terms? Can you reuse it? Who owns the products of any reuse?
- Who is building your database(s)?
  - What rights do you have? Is any database software used and how is it licensed? Have you kept records of your investment?
- How do you make your data available?
  - Do your standard terms of business protect your IP? Do you have binding licence terms? If you make your data and databases available do you include copyright and database rights notices? Remember that users have certain fair dealing rights under copyright law (e.g. in respect of data mining).

Big Data and Data Privacy – Designing Privacy into Big Data

The challenges of big data to data privacy

EU (and UK) data privacy laws bite on the use of big data to the extent the data in question is "personal data". This means data which contains information so that:

- A living individual (the data subject) is *identified* in this information (e.g. they are named or otherwise identified (e.g. date of birth, sex, address)), or
- If a living individual (the data subject), while not identified, is described in this information in a way which makes it possible to find out who the data subject is by conducting further research – i.e. they are *identifiable*. The test here is whether it is likely that reasonable means for identification will be available and administered by foreseeable users of the information and this includes third party recipients (Recital 26 of the EU Data Protection Directive) – this identification can be direct or indirect. An example would be a car number plate or a telephone number – this data by itself doesn't expressly identify the car owner or phone subscriber but information linking this data to individuals is potentially readily available.

Anonymisation

The second limb – that the data subject is identifiable – poses a big issue for big data users. It is often assumed that if data is "anonymised" then it ceases to be personal data and so data privacy laws can be ignored. That is of course in one sense correct but it is surprisingly difficult to anonymise personal data as all identifying elements have to be eliminated and no element may be left in the information which by exercising reasonable effort would serve to re-identify the person concerned. Also the very existence of big data – combining data from databases of a wide variety of information – can make it more likely that taken together an individual can be identified from the data which the big data user (data controller) has.

Also anonymisation may destroy the value of the data or indeed it can't be anonymised to be used – the data may need to be linked back to an individual even though those processing the data don't need to know the actual identity of the individual. So whilst some big data may truly be anonymous e.g. statistical data derived from interrogating the data, other data may not be and indeed may need to retain identifiers.

Pseudonymisation and privacy by design and by default

One solution here is to create "pseudonymised data". Personal data typically contains identifiers such as a name, date of birth, sex and address – these identifiers can be replaced by a pseudonym and "pseudonymisation" is achieved by the encryption of these identifiers in personal data. Pseudonymisation is an example of a *privacy enhancing technology* and is an important element in implementing *privacy by design* – which is crucial to making big data work under privacy laws – privacy by design and by default means from the outset data protection is built into big data databases and their data.

Indeed privacy by design and by default is an express requirement of the new EU General Data Protection Regulation (Article 23), in effect from 25 May 2018.

Steps to be data privacy compliant

Those creating collecting and using big data need to pay attention the following questions:

- Is the data either not personal data at all (e.g. stock market data) or is it anonymous? Do not assume the data is anonymous even if on its face it appears to be – you need to apply the identifiability test under data privacy law.
- If the data you have is personal data are you data privacy compliant? – was it fairly collected, are you using it for lawful purposes, is it kept secure and up to date and so on. If you are supplying third parties with the data or exporting it offshore are you compliant with the laws surrounding the transfer and export of personal data? Transparency (making clear to users how their personal data will be used), fair processing and using the data for the purposes for which it was collected ("the purpose limitation principle"), and consent underpin data privacy law.
- Will you be using the data in "profiling" – this will be expressly regulated by the General Data Protection Regulation
- Are you building privacy by design and by default into your databases and data collection and use strategies?

Answering these deceptively simple questions will often not be easy. But it will be much easier if the financial services company concerned has implemented information governance and management policies and procedures and is abreast of current EU developments such as the General Data Protection Regulation which will come into force on 25 May 2018.

Big data and Competition, Regulatory and Consumer law

The FCA and the CMA are taking an increasing interest in how the use of big data affects or may affect consumer outcomes and competition – for example looking at retail general insurance products big data could have positive implications in terms of tailoring products but it might also affect consumer behaviour (if they became suspicious about the data collected or its uses) and it could also enable insurers to better segment risk – this might reduce premiums for some but increase or deny coverage to others. In November 2015 the FCA issued a Call for Inputs in relation to big data in retail general insurance, raising these and other issues, and the Competition & Markets Authority (CMA) has also recently carried out a fact-finding project to understand how consumer data is being collected and used commercially.[v] The CMA in particular has highlighted the complex legal regime that applies to the collection and processing of consumer data which in the UK includes:

- Data Protection Act 1998 (DPA)
- The Privacy and Electronic Communications (EC Directive) Regulations 2003 (as amended) (PECR) – these deal among other things with direct marketing, cold calling and the use of cookies as well as the security of public electronic communications services and the privacy of customers using communications networks or services
- A range of consumer laws relating to unfair terms in consumer contracts, unfair commercial practices, laws on distance selling and the Consumer Rights Act 2015. These laws mean that practices and contract terms used in relation to consumer data collection and use are subject to additional regulation beyond the DPA and PECR.[1]

We can expect the collection and use of big data to have increased regulatory scrutiny by regulators more generally, and not just ICO (data protection). Also the FTC in the USA has recently issued a report highlighting the exclusionary risks of big data, for example in unfair discrimination as regards consumer lending.[vi]

Conclusion

Big data raises significant legal issues but the use of big data also potentially brings great rewards in designing and pricing products and in targeting and best serving customers. Financial services businesses need to take the protection of their IP and how they collect and manage their information, including data privacy and regulatory compliance, seriously if the opportunities big data presents are to be fully exploited.

Author

Simon Stokes
PARTNER
T: 0207 814 5482
E: simon.stokes@blakemorgan.co.uk

[i] A previous version of this article is included in Compliance Office Bulletin (Issue 134 – March 2016, Thomson Reuters 2016)
[ii] Gartner IT glossary, Big Data http://www.gartner.com/it-glossary/big-data (accessed 30 January 2016)
[iii] Comments of Experian to the House of Commons – see *The big data dilemma* (House of Commons 2016)
[iv] See discussion in the FCA CfI (note 44 above) pp7-8. See also the Opinion of the EDPS (European Data Protection Supervisor) Opinion 7/2015 Meeting the challenges of big data https://secure.edps.europa.eu/EDPSWEB/webdav/site/mySite/shared/Documents/Consultation/Opinions/2015/15-11-19_Big_Data_EN.pdf, Article 29 Working Party Statement on Big Data http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp221_en.pdf
and ICO guidance https://ico.org.uk/for-organisations/guide-to-data-protection/big-data/
(accessed 1 February 2016).
[v] The commercial use of consumer data, Report on the CMA's call for information, CMA38, June 2015
[vi] Big Data: A Tool for Inclusion or Exclusion? (FTC, January 2016) https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf (accessed 1 February 2016)