The Information Commissioner’s response to the Financial Conduct Authority’s call for inputs on big data in retail general insurance

1. The Information Commissioner has responsibility for promoting and enforcing the Data Protection Act 1998 (DPA), the Freedom of Information Act 2000, the Environmental Information Regulations 2004 and the Privacy and Electronic Communications Regulations 2003 (PECR). He is independent from government and upholds information rights in the public interest, promoting openness by public bodies and data privacy for individuals. The Commissioner does this by providing guidance to individuals and organisations, solving problems where he can, and taking appropriate action where the law is broken.

2. The Information Commissioner welcomes the opportunity to respond to the Financial Conduct Authority’s call for inputs on big data in retail general insurance. The Information Commissioner recognises the potential benefits to consumers that the use of big data can bring. In the context of the consultation, this may include, amongst other things, more granular pricing of insurance premiums and increased innovation in the retail insurance market. However, there are several important data protection and privacy points to consider.

3. On 25 January 2012 the European Commission proposed a comprehensive reform of data protection rules in the EU, the General Data Protection Regulation (GDPR). A political agreement on the new rules that will be put in place across the EU was reached in December 2015. The final text of the Regulation has not yet been published in the Official Journal, but we have tried to give an indication of those areas we feel may impact on processing personal information in the big data context. As a result, our response to this call for inputs is framed in general terms.

4. As was noted in the call for inputs (CfI), in 2014 the Information Commissioner produced a paper looking at big data and data
This paper sets out the Commissioner’s position on big data when it involves the processing of personal data. The paper highlights, in general terms, the data protection and privacy compliance areas that organisations employing big data analytics must consider. In this response we have considered the retail general insurance market context.

5. Of particular interest to the ICO is how data is collected and shared, where the data is sourced from and how well informed individuals are of the processing. For example, the CfI mentions social media data being used in the context of big data analytics. The use of such data is referred to in our big data paper and in our Personal Information Online Code of Practice. We note such data is referred to as “publically available” in the CfI, but the collection and analysis of an individual’s social media data may not align with their expectations, particularly if the data is used to help make inferences and decisions which impact upon them.

6. Given the ICO’s remit, we have not provided answers to all of the questions posed in the CfI. Rather, we take the opportunity given in paragraph 6.4 of the CfI to provide information on the relationship between big data and the DPA in the retail general insurance context. It is worth reiterating that the DPA will only have effect where the data being processed is ‘personal data’ as defined in section 1(1) of the DPA. That is:

“... data which relate to a living individual who can be identified—

(a) from those data, or
(b) from those data and other information which is in the possession of, or is likely to come into the possession of, the data controller,

and includes any expression of opinion about the individual and any indication of the intentions of the data controller or any other person in respect of the individual;”

The DPA does not govern the processing of data that falls outside of this definition, for example data that is not about individuals or data that has been fully anonymised.

---

2 Available at: https://ico.org.uk/media/for-organisations/documents/1591/personal_information_online_cop.pdf

13/01/2016 Version 1.0 (final)
Overview of data protection principles

7. We have previously stated that ‘big data is not a game played by different rules’, and organisations need to ensure they comply with the DPA and the eight data protection principles (see annex) when processing personal data. We have sought to highlight below several areas where processing big data raises potential challenges for organisations in complying with the data protection principles.

8. The first data protection principle requires the processing of personal data to be fair and lawful, and, in addition, that organisations are able to satisfy an appropriate prescribed legal condition. Transparency is a key aspect in meeting the fairness requirement of the first principle, and an element of this relates to giving individuals sufficient information about the data processing being undertaken.

9. In accordance with the sixth principle, organisations need to ensure they respect the rights granted to individuals under the DPA, such as the right of subject access. The processing of big data can make providing individuals with the information they are legally entitled to a challenge.

10. The second principle concerns the purposes for which data are collected and places restrictions on further processing.

11. The third and fifth data protection principles relate to data minimisation, that is the requirement that data is adequate, relevant and not excessive, and kept for no longer than is necessary.

12. The fourth data protection principle is also relevant, as the greater the volume of data being processed, the harder it is for an organisation to ensure the data is accurate and remains up to date. This is a particular concern with the use of unstructured data that needs further analysis in order to be relevant to the big data processing operation. For example, when analysing social media data it is possible that algorithms will misinterpret the meaning, or miss the true sentiment behind, a post. The fourth section of this response will deal with other issues that can arise from the use big data, for instance the challenges that arise when utilising machine learning within a big data system.

13. This response will specifically answer questions 14 and 15 of the CFI, which consider consumer attitudes and the effect that big data may have on consumer behaviour.
Fair processing

14. The first data protection principle requires any processing of personal data to be fair and lawful, and the processing must satisfy one of the conditions for processing contained with Schedule 2 of the DPA (and Schedule 3 if sensitive personal data is being processed). One of the key aspects of fair processing is that the individual is made aware of how an organisation is going to use the data and how it will affect them. For this purpose, organisations processing personal data need to clearly explain what data they collect and the reasons it is being processed. It is important for any insurer looking to utilise big data to consider how best to communicate this to consumers in an understandable way. Big data analytics can be very complex and there are inherent challenges in providing meaningful explanations to individuals. However, this challenge should not be viewed as an excuse to fail to provide an appropriate level of information to individuals.

15. Organisations may consider they should seek consent of the individual before processing their data to calculate, for example, the premium for a particular policy. There are two levels of consent set out in the DPA. For personal data ‘consent’ is a valid condition, whereas the processing of sensitive personal data requires ‘explicit consent’. Whilst the meaning of consent is not further defined in the DPA, the Data Protection Directive, from which the DPA is derived, gives some indication as to what should be considered valid consent:

“...any freely given specific and informed indication of his wishes by which the data subject signifies his agreement to personal data relating to him being processed.”

16. Sensitive personal data is set out in section 2 of the DPA and covers data that relates to, amongst other matters, an individual’s racial origin, health, and religion. It is possible that inferences made through the use of big data could include sensitive personal data, and would require the higher threshold of explicit consent. The ICO’s guidance on what constitutes explicit consent clarifies the difference, stating that explicit consent:

“...suggests that the individual’s consent should be absolutely clear. It should cover the specific processing details; the type of information (or even the specific information); the purposes of the...
processing; and any special aspects that may affect the individual, such as any disclosures that may be made.”

17. This sets a high standard in terms of the information that must be provided to the individual, which could prove challenging for organisations in a big data context. As highlighted above, big data processing is a complex area and providing enough information to customers in a way that leaves them fully informed about the collection and processing of personal data presents a challenge.

18. We should also highlight some of the challenges of reusing publicly available information, and in our Personal Information Online Code of Practice we outline our perspective on this matter. Regardless of where the data has originated from, there is still “an overarching duty to handle it fairly and comply with the rules of data protection”, meaning an organisation may breach the DPA if they reuse the data in a way that could be considered unfair. We suggest in the Personal Information Online Code of Practice that best practice would be to only use publicly available information in a way that the data subject would be likely to expect and be comfortable with. In terms of further processing and profiling using publicly available data, the inferences, opinions or predictions about individuals and their behaviour gleaned from this processing are likely to constitute personal data and subject to the safeguards in the DPA.

19. Organisations must carefully consider the legal and ethical dimensions of processing social media data, and how individuals may feel if they knew social media content about them was being used in a given context. Even well-intentioned uses of social media data by third parties can cause privacy concerns, as the recent example of the Samaritans’ Radar app demonstrated. Insurers must also be mindful of the varied level of understanding and application of privacy settings on social media accounts and bear in mind that some users may not realise that social media data relating to them is potentially available to be scraped from the internet. With these challenges in mind, insurers should consider whether they have legitimate grounds to use data that may have been gathered from social media platforms or other online sources for insurance purposes, rather than merely relying on the fact that some content is accessible.

3 https://ico.org.uk/for-organisations/guide-to-data-protection/conditions-for-processing/
4 Personal Information Online Code of Practice, page 18
As stated above, if making inferences from data that can be accessed publically, these inferences are likely to be new personal data themselves and will be covered by the DPA. Additionally, using social media data to create profiles of individuals that take into account their social connections could also amount to processing the data of third parties.

**Individual’s rights**

20. When processing personal data insurers must take into account an individual’s rights under the DPA, in particular the right to access the data, and rights related to the automated processing.

21. When processing personal data insurers are obliged to provide the data subject with a copy of their personal data upon request. This could pose a significant challenge given the voluminous nature of the datasets being employed, and the potential for unstructured data to be held. It is important that any organisation undertaking big data analytics considers how they would respond to a subject access request (SAR) from the outset, and builds-in the capability to be able to collate the requestor’s personal data in order to respond effectively. Additionally, in most circumstances a data subject is entitled to be informed about the logic behind a decision that has been made about them.

22. The DPA also confers upon individuals certain rights in relation to automated decision-taking, which is clearly relevant for organisations employing big data analytics, and especially where machine learning is being used. Further, an individual has the right to require that no decision that significantly affects them is made solely by automatic means. Insurers that employ fully automated processing without human intervention should consider how their processing takes these rights into account.

**Purpose specification**

23. The second principle relates to the purposes for which personal data are collected and processed. It can be seen as a two-part test, whereby (depending upon the context of the processing) the purpose for which the data is being collected and processed must be disclosed to the data subject, and any further processing must not be incompatible with the original purpose. The second element raises issues for using big data in the insurance context. The example the CfI sets out regarding the use of purchase data from supermarket loyalty schemes highlights the potential for data collected for one reason being
repurposed to gain insights into allocating risk for insurance purposes. In this instance it would be a matter for the organisation in question to decide whether using data for calculating insurance premiums is incompatible with the original purpose of marketing and customer loyalty. When considering this, a key factor should be whether the new purpose would be within the reasonable expectations of the data subject. If the new purpose is incompatible with the original purpose then the organisation will have to inform the data subject of this further processing and gain their consent.

**Data accuracy**

24. The fourth data protection principle requires that data must be accurate and up to date. The use of big data raises issues around the accuracy of the data that is being used in the analysis (for example the provenance of data sources, such as social networks) and, consequently, the accuracy of the inferences garnered from that data. There should also be some way for data subjects to rectify any errors that may exist within the insurer’s datasets. Some thought must also be given as to how to keep data up to date. Profiles, inferences and insights derived from the data may need to be altered over time to keep up with a data subject’s changing profile.

**Data minimisation**

25. Organisations will also need to consider the third and fifth data protection principles when considering using big data for insurance products. The third principle says that “personal data shall be adequate, relevant, and not excessive in relation to the purpose ... for which they are processed.” whereas the fifth principle requires that data is not kept longer than necessary. The nature of big data encourages the use of as much data as possible to see what connections can be made. If an organisation is unable to justify the use of certain categories of data then they risk failing to comply with this principle.

26. A key data governance consideration with big data is the temptation to keep data for an indefinite or excessive period in case it may be found to be relevant later on. This engages the fifth principle, and it is important for organisations to be mindful of retention periods, and the restrictions that apply when processing personal data.
Other aspects that should be taken into account

27. The use of machine learning also poses some questions that organisations will need to consider. Part of the role of big data analytics is to find correlations within the multiple data sets rather than testing a particular hypothesis. The use of such tools therefore has the potential to raise the question of whether these correlations are reasonable. Leaving it down to a computer to make these connections has the potential to lead to unfair decisions being taken, unless there is some degree of oversight and human intervention. There is also scope for systems based on machine learning to apply certain criteria that may be discriminatory, for instance taking into account an individual’s race or religion when calculating an insurance premium.6

28. It is worth noting that the General Data Protection Regulation will be likely to affect the regulatory framework with regards to big data analytics. It is possible, at this stage, to highlight some provisions of the Regulation that insurers will need to be mindful of in the future. Article 20 concerns profiling, and affords users with similar rights to those found under the existing section 12 of the DPA. However, profiling is included as a specific example of data processing. This provision sets out justifications for utilising automated processing, which insurers will have to consider. There is also emphasis given to the ability of individuals to have a decision that has been made about them looked at by a human being. There have also been changes made to the definition of consent, which has been strengthened. There are also significant obligations under Article 14a placed on data controllers where data is sourced from third parties. Big data analytics may involve utilising many data sets that have been sourced from third parties, and where this is undertaken the data controller, in this case the insurer, is required to provide certain information to the individual concerning the data they now have. It should also be noted that data relating to children will be subject to greater protection. This is especially the case with regards to profiling and the right to erasure.

29. In addition to providing a general overview of how data protection impacts the use of big data in the general insurance market, we would also like to take this opportunity to provide some more specific answers to questions 14 and 15 of the CfI.

6 The potential for big data to result in exclusion and discrimination was discussed in a recent paper by the FTC titled Big Data: A Tool for Inclusion or Exclusion? Available at: https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf
Paragraph 4.16 refers to the issue of consumer trust, and the negative consequences for business of losing that trust.

**Q14: Do consumer attitudes towards the potential use of data impact firms’ decisions to invest in Big Data?**

**Q15: To what extent is consumer behaviour influenced or affected by the use of Big Data?**

30. There does not appear to be evidence of a specific consumer reaction against organisations using big data analytics, but there is evidence of consumer concern about the use of personal data. For example, a Digital Catapult survey in July 2015\(^7\) asked people which sector they trusted with their personal data. After the public sector (43.5%), financial services was the most trusted, but still with only 28.6%. The survey also asked people what most concerned them about organisations’ use of their personal data; 76% chose “that I have no control over how it is shared or who it is shared with”.

31. By contrast, the Direct Marketing Association commissioned the Future Foundation to look into attitudes to personal data in 2012 and 2015\(^8\). They found that the percentage of so-called ‘fundamentalists’ who won’t share their data fell from 31% to 24% and the percentage of ‘not concerned’ increased from 16% to 22%.

32. On a complex issue such as this, survey evidence is not entirely conclusive and may produce apparently contradictory results. Also, people may take a particular view in a survey which is not borne out by their day to day practice. However, even if people appear to be increasingly willing to make their personal data available, both through social media and by providing it to organisations when asked, and they do not actively voice concerns about big data, this should not be a cause for complacency on the part of those handling that data, for the following reasons:

33. The processing involved in big data analytics is generally invisible to the data subject. Individuals are unlikely to voice concerns unless they are negatively impacted by decisions made as a result of the analysis, for example if their insurance premiums

---

increase because the risk they represent has been more accurately assessed.

34. An individual’s apparent willingness to share their data may indicate that, consciously or not, they invest a level of trust in the organisation in question. This in turn places an onus on the organisation to maintain that trust and to not use the data in a way that would compromise it.

35. The fact that there hasn’t been a consumer reaction against big data so far doesn’t mean that it won’t happen in future. It is true that where there is evidence of consumer reaction to the misuse of data it has been as a response to information security breaches, rather than to big data processing\(^9\). Nevertheless given that consumers are well versed in using social media to publicise their complaints about services, there is a clear possibility of negative publicity if a ‘creepy’ use of big data comes to light, even where there has not been a data loss.

36. There is evidence that responsible companies are seeking to mitigate this risk by developing ethical approaches to big data analytics. Key features of these approaches are:

37. Understanding the customer’s point of view. Some big data practitioners talk about the ‘granny test’ – would you be happy if it was your elderly relative’s data that was being used?

38. Being transparent about what you’re doing with the data.

39. Building a relationship in which customers trust the organisation to handle their data properly.

40. Creating a value exchange, in which customers can see the benefit they receive in return for their data (e.g. financial rewards, improved services) and are encouraged to provide more data in return for more benefit.

41. Our paper on Big data and data protection refers to the examples of Aimia’s ‘new data values’ and IBM’s ethical framework for big data analytics. In the insurance context, Paul Evans of the ABI earlier this year called for an industry code of conduct on the responsible use of data\(^{10}\). In the US, automobile manufacturers

\(^9\) For example, the US retailer Target’s use of sales data to predict pregnancy was widely cited as an example of the intrusive and unexpected use of analytics, but when they suffered a significant loss in sales in 2013 this was seen as a reaction to a subsequent security breach, rather than to the analytics.

\(^{10}\) Gray, Alistair. Insurers warned to use ‘big data’ responsibly. FT.com 1 February 2015
have developed a set of privacy principles for the use of data from telematics devices in cars\textsuperscript{11}.

Summary

42. Whilst supportive of the use of big data to innovate in the insurance market, bringing benefit to customers in both the price they are paying and the services that they have access to, the Information Commissioner is keen to highlight areas that insurers must address their minds to. There are real challenges with processing information fairly, such as keeping individuals informed about the processing being undertaken. However, it is important that insurers do not avoid this by creating a black box where insurance premiums are calculated without the individual understanding how this happens. Insurers will also have to be clear about why they are using the data they have and avoid keeping data on a “just in case” basis. Finally, it is important that, when employing machine learning, insurers regularly review what the algorithm is taking into account when assessing individual risk.

43.

Annex

<table>
<thead>
<tr>
<th>The data protection principles</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Personal data shall be processed fairly and lawfully and, in particular, shall not be processed unless—</td>
</tr>
<tr>
<td>a. at least one of the conditions in Schedule 2 is met, and.</td>
</tr>
<tr>
<td>b. in the case of sensitive personal data, at least one of the conditions in Schedule 3 is also met.</td>
</tr>
<tr>
<td>2. Personal data shall be obtained only for one or more specified and lawful purposes, and shall not be further processed in any manner incompatible with that purpose or those purposes.</td>
</tr>
<tr>
<td>3. Personal data shall be adequate, relevant and not excessive in relation to the purpose or purposes for which they are processed.</td>
</tr>
<tr>
<td>4. Personal data shall be accurate and, where necessary, kept up to date.</td>
</tr>
</tbody>
</table>

\textsuperscript{11} Consumer privacy protection principles. Privacy principles for vehicle technologies and services. Alliance of Automobile Manufacturers Inc and Association of Global Automakers Inc., 11 December 2014

13/01/2016 Version 1.0 (final)
5. Personal data processed for any purpose or purposes shall not be kept for longer than is necessary for that purpose or those purposes.

6. Personal data shall be processed in accordance with the rights of data subjects under this Act.

7. Appropriate technical and organisational measures shall be taken against unauthorised or unlawful processing of personal data and against accidental loss or destruction of, or damage to, personal data.

8. Personal data shall not be transferred to a country or territory outside the European Economic Area unless that country or territory ensures an adequate level of protection for the rights and freedoms of data subjects in relation to the processing of personal data.