The Commercial Use of Consumer Data

A research report for the CMA

DotEcon with Analysys Mason

June 2015

DotEcon Ltd
17 Welbeck Street
London W1G 9XJ
www.dotecon.com

Content

Tables & Figures

1 Introduction and overview

1.1 Purpose of this study

The study examines how and why consumer data is collected and used

The Competition and Markets Authority (CMA) has commissioned DotEcon, together with Analysys Mason, to research how and why consumer data is collected and used through case studies in a small number of sectors. Whilst these sectoral studies are primarily factual, and circumstances identified in a particular sector may not hold true more generally, they provide insight into how markets for consumer data work in practice. This is intended to help the CMA better understand the scope and nature of issues arising in markets making use of consumer data.

In order to understand typical patterns of consumer data collection and use in different sectors, we have investigated:

- what kinds of consumer data are collected;
- how they are collected;
- how consumers are informed about what data is collected and how it will be used;
- how and why data is used;
- whether uses rely on consumer-specific or aggregated data, and whether consumer-specific data is identifiable to an individual or not;
- how data is sold and exchanged between companies and combined with other sources; and
- how parties realise the value of this data.

Three sectors have been selected for case studies

This report focuses on the collection and use of consumer data in three sectors (whose geographic scope for the purposes of this study is the UK only). Together with the CMA, we considered a range of possible sectors for the case studies, taking into account factors such as the characteristics of the sector and extent of data collection. The final selection was not based on any particular concerns about the sectors; instead, the aim was to identify three case studies that would provide a wide range of factual evidence as well as deeper insights into specific examples of data collection. On this basis, the sectors selected include:

- motor insurance – in part because of its long history of collecting and using consumer data offline and online to assess risk and set premiums, as well as recent developments such as the collection of telematics data;
- clothing retail – to provide evidence for the retail sector more generally and in part because of the growing role of

specialist service providers, social media and possible use of data to inform development; and

- games applications (or 'apps') – due to the relatively young and fast-moving nature of the sector and to help understand the growth in online mobile data collection and use. We consider games apps accessed either directly as installed apps on mobile devices or indirectly via social media networks.[1]

In addition to commissioning this research study, the CMA launched a call for information (CFI) about the commercial use of consumer data in January 2015.[2] We understand that this research and the responses to the CFI will provide inputs to the CMA's wider project to understand better the commercial use of consumer data, including how it affects customers, businesses, competition and the wider economy.[3]

1.2 Stakeholder engagement

Interviews were held with suppliers and 'infomediaries'

The case studies gather information from existing literature and publicly available sources, supplemented by additional evidence obtained from interviews with key stakeholders. We sought to conduct 3–4 interviews with suppliers in each of our chosen sectors and around 10 interviews with 'infomediaries'[4], some of which specialise in supplying services to firms in one of the sectors specifically, and others which operate across a large range of sectors.

All of the interviews we conducted proved to be an important complement to our initial desk research; the use of consumer data has developed rapidly and interviews allowed us to probe current commercial practices. We are grateful to the various companies and industry bodies who have helped us. For confidentiality reasons, we do not attribute information or opinions directly to the

[1] The research on games apps focuses on the collection of consumer data through games applications used by adults (excluding gambling).

[2] See CMA, 27 January 2015, 'The commercial use of consumer data: call for information', https://www.gov.uk/government/consultations/commercial-use-of-consumer-data

[3] The case timetable is available at: https://www.gov.uk/cma-cases/commercial-use-of-consumer-data

[4] An 'infomediary' refers a third party whose role is the exchange or processing of data, typically on behalf of other market participants.

stakeholders we interviewed. References to specific companies or service providers are on the basis of publicly available information.

The three case study sectors differ significantly in the amount of public domain information available concerning the collection and use of consumer data. We also found differences in the willingness of stakeholders to contribute information across the three sectors.

A comprehensive body of evidence was found for the motor insurance sector

There is a large body of research and well-documented evidence of commercial practices within the motor insurance sector. Further, we found considerable willingness on the part of market participants to assist in the study. We were able to gather evidence, either through interviews or written responses, from eight stakeholders who included insurers and price comparison websites (PCWs). In addition, we interviewed two infomediaries that specialise in serving the motor insurance sector. As a result, we have reasonable confidence that the full breadth of issues for consumer data in the motor insurance sector has been explored.

In the other two sectors there was a lower response rate from stakeholders

In contrast, whilst there is a large amount of high-level information about both the clothing retail and games apps sectors in the public domain, there is limited information about the incidence of use and degree of success of new business practices related to consumer data. A large number of stakeholders were contacted in both sectors. However, due to a low response rate we were unable to secure as many interviews in these sectors as for motor insurance. Nevertheless, in the clothing retail sector we interviewed a major retailer with a large number of high-street stores and an online presence and also two third-party specialist service providers. In the games apps sector, we interviewed three successful developers, one app store and one infomediary specialised in serving the games apps sector. As a result, our findings in these sectors may not be as comprehensive as those for the motor insurance sector; in particular, there may be suppliers with rather different business models who we have not covered.

1.3 Key concepts

By 'consumer data', we mean all forms of information that a 'data collector' might collect about individuals, including both customers and potential customers. This might be gathered through online interactions – on which we focus – or by other mechanisms. The data collector may be:

- a *first party,* with a direct relationship to the customer; or
- a *third party* collecting data on behalf of the first party.

Typically, but not always, the data collector is also the 'data controller' in the sense defined in data protection legislation as "*the party who determines the purposes for and the manner in which (personal) data is processed*"[5].

Consumer data can be classified into two main types:

- **Personal data** related to an individual who is identifiable from that data, either from the data itself or in combination with other data available to the data controller.[6] This data often involves the individual's characteristics, location or transactions. For example, name, address, email address and payment method details of a customer are typically needed to make purchases online. The processing of personal data, including collection, storage, usage and transfers, is governed by data protection and privacy rules.[7] Personal data is often commercially valuable as it allows analysis at the level of individuals; for example, personal data on demographic characteristics, location and previous motor insurance claims, combined with vehicle information, enable motor insurance providers to provide insurance quotes that reflect that individual's risk profile.
- **Non-personal data** that does not contain personally identifiable characteristics. There are two main types of non-personal data:
    - **Pseudonymous data**: This may include details such as age range, gender and approximate location, but does not contain sufficiently specific information to identify a particular individual (either from the data itself or if combined with other datasets that may be in the possession of the same data controller). Data may be collected directly as pseudonymous data. Alternatively, personal data may be transformed into pseudonymous data by removing certain data fields (e.g. email address) so that it can be freely exchanged and processed by other parties as non-personal data.
    - **Aggregate data**: This results from combining data – personal or pseudonymous – on a group of individuals. It is used to provide high-level insights about a group of individuals. For example, a games

[5] As defined in Part 1 of the UK Data Protection Act 1998.

[6] As defined in Part 1 of the UK Data Protection Act 1998.

[7] The Information Commissioner's Office provides an overview of data protection rights and duties: https://ico.org.uk/for-organisations/guide-to-data-protection/key-definitions/#personal-data

developer might be interested in aggregate information such as the average age of players that regularly play its games apps.

Broadly speaking, consumer data may be collected in a number of ways. Consumer data may be:

- **declared** explicitly by a consumer as part of the interactions that consumers initiate with organisations, for example by seeking a motor insurance quote or making a purchase from an online clothing retailer. In some cases, more than one party is involved: for example, a consumer using a Facebook login option can allow Facebook to share some of its previously declared data with another party;
- **observed** in the course of interacting with a website or using a service, device or application. For example, browsing histories may be observed through cookies and information on players' in-app activity might be collected by an app developer; or
- **inferred**; for example, the gender of a regular visitor to a retail clothing site might be inferred based on the clothes viewed.

A substantial area of interest is behavioural data

A wide range of stakeholders collect and use **behavioural data**, which may occasionally be declared (e.g. when an insurance customer states that a vehicle will be used for commuting) or, more commonly, observed by recording consumer interactions and decisions (e.g. through cookies). For example, many developers of games apps consider the collection of in-app activity a vital requirement in support of their business model. Behavioural data is potentially useful for testing and refining analytical techniques and also for measuring the impact of changes in business models, marketing strategies and so on. For example, a third party providing targeted advertising or recommendations to customers will track whether this results in a sale and use this behavioural data to refine its recommendation engine or provide evidence for payment-by-result business models.

Behavioural data may be collected as personal data, for example when a registered user is logged in to a retail website that is capturing behavioural information through the use of 'cookies' (discussed in Box 1 below). Alternatively, this data may be pseudonymous, where consumer behaviour on a website is linked to an identifier (such as a device ID), but not to a specific individual. Pseudonymous behavioural data may be sufficient for many purposes. For example, when a device is used to browse – but not purchase – red dresses on a retailer's website, cookies can allow the retailer to display advertising for the red dresses when the same device is being used to browse other websites ('re-targeting'), without requiring the identity of the device user.

Box 1: Cookies and identifiers

Many websites make use of devices such as 'cookies' or 'beacons', which store behavioural or contextual information about users. Cookies are small text files placed on a user's device that can be interrogated by websites. Session cookies are created temporarily when a particular website is visited and then deleted when the user leaves the website, whereas persistent cookies can remain stored over time as the user visits multiple websites. Cookies record a user's actions, such as movements within the website and, in the case of persistent cookies, actions prior to visiting the current website. For example, persistent cookies can be used to show where the user came from to arrive at the current website.

Beacons or 'tracking pixels' are very small images that may be embedded on a web page or an email. When the web page or email is loaded, the request for the image is made to the server, allowing the owner of the beacon to track this event, along with other basic items of information (e.g. the time of the visit to a web page, the browser used).[8]

Cookies and beacons can be placed by website publishers themselves, but many are placed by third parties such as analytics companies and advertising networks. Third-party cookies and beacons enable users' behaviour to be observed across multiple websites who use the same third-party provider (e.g. Google Analytics).

Cookies may be used for various purposes, which can be categorised as follows:[9]

- strictly necessary – e.g. allowing items to be added to a shopping basket;
- performance – e.g. allowing the website owner to improve how the website works;
- functionality – e.g. allowing the website to record users' choices, such having a username 'remembered' by the website; and
- targeting or advertising, which allow advertisements to be targeted on the basis of browsing behaviour, while also being used to limit the number of times an advert is shown and to track advertising effectiveness.

Since 2011, website publishers in the UK and the rest of Europe are required to set out their policies on the use of cookies, data collection, data

[8] This information may be of limited benefit to a website owner, who can already record requests for particular webpages made to the server, but beacons may be more useful for third parties (e.g. ad networks) to track activity across multiple websites.

[9] See International Chambers of Commerce, November 2012, 'ICC UK Cookie Guide', http://www.international-chamber.co.uk/images/ICC_Documents/icc_uk_cookiesguide_revnov.pdf

use and data sharing on their websites, in accordance with revisions to the ePrivacy directive.[10] In order to place and process cookies, websites must either obtain explicit consent or satisfy themselves (and the Information Commissioner's Office) that users have been properly informed about these devices so that consent can be presumed (e.g. by their continuing to use the service). Most websites do so through pop-up windows on the home page, explaining that they use cookies and that continued use of the site implies consent for them to do so. Some websites[11] provide an opt-out possibility, but often warn that the customer experience on the website may be degraded as a result.

Consumers can prevent their browsers from accepting cookies and often have various options available in order to do so – e.g. blocking all cookies, blocking third-party cookies only, or creating exceptions for particular websites.

1.4 Data flows in the collection and use of consumer data

Consumer interaction is the starting point for the collection and use of data

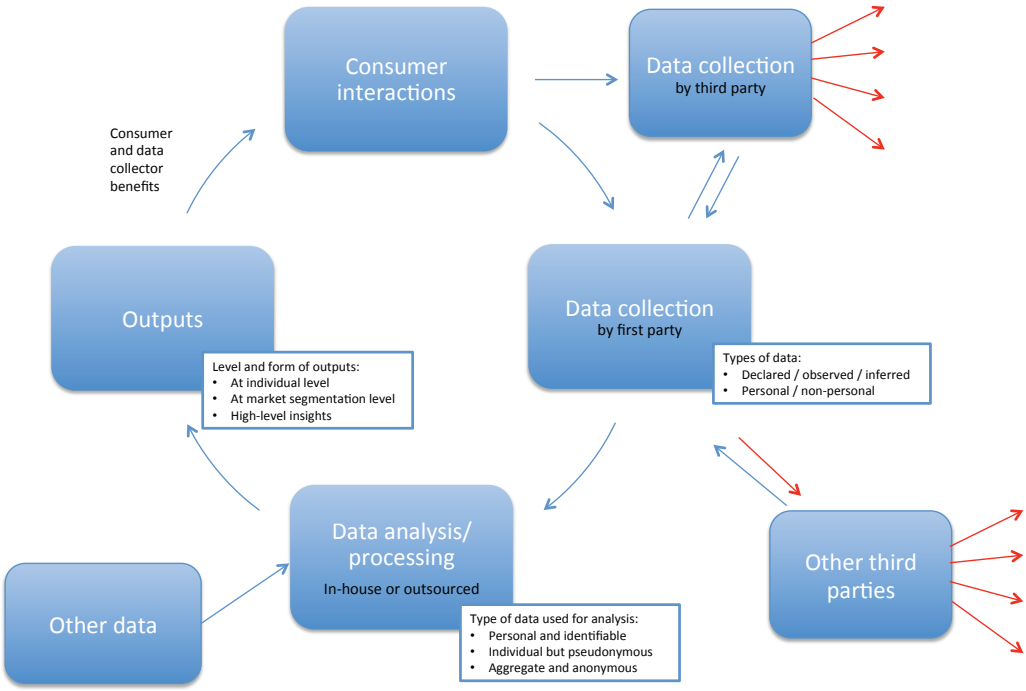
When consumers interact with first parties, they generate data. This data can be collected directly by the first party or indirectly by a third party acting on its behalf. First parties may enrich this consumer data by linking it with additional data they hold themselves or sourced from third parties, before conducting analysis, which may be in-house or outsourced. The ways in which data is processed, analysed and used is driven by the commercial interests of the first party. The outputs may in turn create benefits for the consumers and influence their future interactions with the first party (for example, by improving the service being offered). We note that in this report, we discuss the flows of data and the commercial agreements in place, where possible, however information on exact values and volumes is not always available, partly because this is commercially sensitive to stakeholders.

Figure 1 below provides a generalised, high-level illustration of these data flows.

[10] Unless the use of cookies is exempt from this requirement, for example in the case where the use of cookies is strictly necessary for the provision of a service requested by the user. See https://ico.org.uk/for-organisations/guide-to-pecr/cookies/

[11] For example BT's website, which provides three options: users may choose to allow only strictly necessary and performance cookies, or they may additionally allow functional cookies, or they may allow all types of cookies (including targeting).

Figure 1: Data flows involved in the collection and use of consumer data



Source: DotEcon and Analysys Mason

Data may be transferred between first parties and third parties

The closed cycle at the 'core' of the diagram shows the key steps of data collection and analysis, with the outputs providing benefits to the data collector and consumers, affecting future interactions with consumers. However, as illustrated by the links to boxes outside of this cycle, firms may also collect data from outside this cycle, including data sourced from third parties. These external links also illustrate the possibility that data collected as a result of consumer interactions with first parties could leave the cycle, being transferred to third parties for uses unrelated to the core purpose of the first party in interacting with its customers or potential customers. The red arrows illustrate this possibility. An innocuous example could be the supply of pseudonymous data by a price comparison website to a third-party firm, which then conducts analysis to show how the average motor insurance premiums are changing over time for different demographic groups of consumers.[12] However, the possibility of data leaving the cycle could raise concerns, for example if personal data is involved, if there is the potential for harmful uses of data, or if consumers lack awareness of and control over this flow of data.

In the following two sub-sections we provide a high-level overview of:

[12] See http://www.confused.com/car-insurance/price-index/about

- how the interaction of consumer and firms can generate data and how that data might be collected; and
- how data might be analysed and the outputs of that analysis used.

In each case, we provide some simple illustrations from the three case studies.

1.5 Consumer interaction and data collection

There are many ways in which consumers and firms might interact, in turn influencing the nature of consumer data that can be collected and the value of that data. This is illustrated by the different kinds of consumer interaction in the three sectors:

- Motor insurance providers collect data from existing and would-be customers relatively infrequently: once a year, at or near the time of policy renewal, and occasionally in the case of a claim or a change of details. The information collected provides a snapshot of an individual's situation, which may change over time. The data is declared by the individual and includes both personal information (on the individual, their location and the vehicle being insured) and behavioural information (e.g. information on previous claims). As individual circumstances can quickly change, the value of this data for assessing an individual's risk can degrade quickly, even after a year has passed. Therefore, there is an incentive to obtain a new snapshot of consumer circumstances to aid assessment of the risk associated with an individual. Nevertheless, large volumes of historical data are valuable when used to test and refine predictive models of risk.
- Online clothing retailers have an on-going but intermittent relationship with repeat customers (or repeat website visitors who are would-be customers). A retailer may collect behavioural data on browsing, purchasing and personal information related to payments via cookies or through a customer's online profile on the retailer website where applicable. Over time, this may allow an online clothing retailer to collect a rich dataset of an individual's browsing and purchase history and the items that they have returned. Data may degrade in value over time (as consumer tastes may change), but relatively recent data may allow clothing retailers to maintain their relevance to consumers through recommendations and personalisation. For instance, a clothing retailer told us that it uses only customer history from the past six months for the purposes of recommendations.

- Games developers may also have an intermittent but on-going relationship with games users. Behavioural data is collected from the in-game activity of users (e.g. for how long a user plays, in-game progress, whether the customer makes in-app purchases[13]). In the case of games played through mobile applications, developers may collect data that is pseudonymous, where there is an 'identifier' linking the data to a device or to an IP address as opposed to a known individual. Through tools such as Google Analytics[14] and Facebook Insight, games developers may also collect aggregate data on their users, such as gender and age range of players. The case of games played through social media (or on mobile apps using Facebook login) is somewhat different, as developers also access personal information linked to the user's social network profile. Where users register or create an account with the developer, this may also allow personal information to be collected.

Third party data collection

In addition to data collected by first parties, there are often other parties collecting data on the first parties' behalf. For example, third-party companies may embed and control cookies on first party websites. A further example is that credit reference agencies collect a large amount of personal behavioural data about an individual's financial history.[15] This data can then be used by various firms and organisations, such as motor insurance providers seeking to verify personal information provided to them by individuals and to establish an appropriate APR for monthly payments.

There appear to be a growing number of companies that specialise in data collection and data processing, including personal data, without necessarily having direct interactions with consumers. Business models are developing based on data enrichment activities and aggregating consumer data in new and different ways. For example, some companies provide 'social listening services' where they scrape and index information on the internet, in particular on social media websites, to gather comments or messages linked to its client's name, brand or products. Tools or

[13] We note that where in-game purchases are made the developer will know through the collection of behavioral data. However, payment data (such as credit card information) is not provided directly to the app developer. This is controlled by the platform (that is, from the app store the app was downloaded from (e.g. Apple's App Store or Google's Play store) and/or the social media platform the game was played on (e.g. Facebook)).

[14] See https://support.google.com/analytics/answer/2956981?hl=en

[15] For example, using data collected either from third-party public sources (e.g. electoral rolls) or transferred to them by third parties in accordance to the Terms & Conditions of services – e.g. credit card data.

dashboards are then provided for organisations to conduct tailored searches. A firm might use a specialised service such as this to assess the effectiveness of a marketing campaign in achieving a certain goal (e.g. re-positioning a brand or creating awareness of the brand amongst a targeted demographic).

In some instances, personal data can be 'repurposed' by firms who find new uses for it. As long as the user has consented to this repurposing and/or it falls under the other legitimate grounds for processing, this would typically fall within the law. By nature, however, the new purposes are not necessarily visible to customers at the point of providing their data and may in any case be difficult for them to understand.

1.6 Data analysis and output of analysis

Data analysis is ultimately driven by a desire to improve decision-making processes. One can expect the collection and analysis of data within a particular sector to be focused on areas in which there are the greatest commercial returns from making such improvements. Our three case studies demonstrate this.

Motor insurance

Data is needed to assess risk

The motor insurance sector relies heavily upon its ability to use consumer data to support its decision-making and business practices. Use of consumer data helps firms assess individual risks and set premiums accordingly. Firms collect a range of information about the driver and about the vehicle. Much of this data is provided directly by the consumer at point of quote, but firms may enrich the data with additional information, either held internally or obtained from a third party. In order to evaluate risk, insurers rely on predictive models that are built and continually refined through a resource-intensive process of data analysis to establish correlations between particular variables and risk. Insurers usually carry out such analysis in-house.

The result of this analysis, where successful, is that insurance premiums for individual customers more accurately reflect their risk profiles. Better evaluation of risk allows insurers to offer competitive quotes to individual drivers, while ensuring that premiums are sufficiently high to cover expected costs overall. Therefore, a firm's predictive risk modelling can be a key competitive asset.

Data can also support fraud prevention

In addition to supporting the risk evaluation process, consumer data is used by insurers to verify customer information and detect mistakes or possible instances of fraud, both at point of quote and thereafter, in relation to claims. To do this, firms typically cross-check information against external data sources. For example, credit reference agencies hold data that can help validate identity;

fraud protection groups hold data that can help validate previous claims information or establish any links to previous fraudulent activity; DVLA data can help validate information about previous driving convictions. Through such processes, firms can establish the likelihood of declared information being accurate and genuine.

The potential benefits from improved fraud detection are substantial – fraud remains a key priority for the sector and it is estimated that fraud adds an estimated £50 to the average general insurance premium.[16] Therefore, any savings achieved by fraud prevention could benefit both firms and consumers.

Data can be used for targeted marketing purposes

Though of somewhat lesser importance, motor insurers use consumer data to analyse their customer base (e.g. segmentation analysis) and reach consumers through targeted marketing, potentially generating additional revenue. This includes the use of personalised communications (e.g. email) or targeted advertising, though the scope for other marketing techniques may be limited in this sector; for example personalised recommendations may be important for firms selling fashion products but less relevant to firms selling insurance policies. Since consumer purchases are infrequent but regular – typically once a year – firms will use information about consumers' renewal dates and target their marketing efforts around the time of renewal; this may increase response rates and might arguably be viewed as convenient from a consumer perspective (as compared with receiving untargeted marketing all year round).

Clothing retail

We have found that the main uses of consumer data in the clothing retail sector are to increase the conversion of consumer interest into actual purchases, to reduce returns and, to a lesser extent, to use high-level insights to inform business strategy.

Data enables personalisation of search results and recommendations

Retailers use personal data (usually provided by the customer when they make a purchase or register on a retailer's website) and pseudonymous data (such as purchase history and browsing history) to develop a data-rich customer profile and make inferences about consumer preferences. This data is then used to provide personalised search results and product recommendations online in real time. Retailers can also use this data to serve targeted adverts to individuals or to provide targeted emails displaying personalised content.

Whilst some retailers carry out these activities in-house, there are also third parties who provide data analysis services and provide product recommendations on behalf of the retailer. Third parties

[16] See https://www.abi.org.uk/Insurance-and-savings/Topics-and-issues/Fraud

will collect behavioural data to help them understand the customer's preferences, and in some cases may be able to augment this with pseudonymous data provided by the retailer.

The benefit to consumers of personalised recommendations is that items of interest may be displayed more prominently, reducing search costs and improving the online shopping experience. If customers engage with the retailer as a result, and respond to the recommendations, the retailer also benefits from increased sales.

Specialist providers use data to enhance the user experience

Other specialist service providers have also emerged to support the online clothing retail sector, helping retailers to improve the customer experience and provide further ways of driving conversions, averaging order values and decreasing the number of returns. For example, there are services that allow customers to better understand how a particular garment will fit someone of their size and shape, or to determine what size to order based on key body measurements or measurements of favourite clothing items.

Such service providers typically analyse pseudonymous data including declared data on body measurements, body shape and style preferences, supplemented by data about the retailer's clothing line, to provide size recommendations and product recommendations that match the customer's preferences. The services are valuable to retailers as they can help reduce size-related returns. This is particularly valuable to retailers in the UK where returns are often provided for free (with the retailer covering the costs of return postage) and with an average of 1 in 4 online clothing purchases being returned.[17] However, these services also lead to increased numbers of sales as customers have a better idea of how the item of clothing will fit or look, reducing uncertainty and removing a potential source of disadvantage for online shopping relative to visiting a physical store. We understand that this motive is especially relevant for visualisation services (that show how the garment will fit on someone of similar size and shape to the customer) where increasing sales is a particular target.

Data can also provide useful high-level insights

Retailers and third-party service providers also conduct data analysis to identify from pseudonymous and aggregated data any customer segments or patterns of usage that might hold valuable insights, e.g. the typical demographic or purchasing patterns of the retailer's customers, the brands or items that have proved popular amongst its customers and those that are underperforming. These insights can then be used to inform key decisions that influence the

[17] See Fits.me, 'Whitepaper – Garment Returns and Apparel Ecommerce: Avoidable Causes, The Impact On Business And The Simple Solution', http://communication.fits.me/acton/attachment/8391/f-0012/1/-/-/-/-/Whitepaper_A4_garment_returns_UK.pdf

services and the style of garments it provides to its customers. These high-level insights can influence the retailer's purchasing decisions, stock levels and, for retailers with high-street stores, determine in-store space allocation.

Mobile and social media games apps

The predominant use of consumer data amongst game developers, whether games are accessed through mobile apps or through social media, is to gain insights about app usage to inform game design and drive improvements in the user experience. The main method of monetisation of games apps usage is through at least some users making in-app purchases. The expectation of developers is that optimising the user experience will increase the number of users, create strong levels of engagement and keep users playing for longer, in turn increasing the volume of in-app purchases. There is also some use of consumer data for targeted advertising purposes, but this often appears to be a secondary concern for developers relative to driving in-app purchases.

Pseudonymous behavioural data can be analysed to optimise game design

Developers of games for mobile apps generally collect pseudonymous data about user activity within their games and have access to aggregate data on the user base of their games from the platform hosting the sales of apps. Data analysis predominantly seeks to identify from this behavioural data any patterns of usage that might hold valuable insights, e.g. about the points where users leave the game, potentially indicating that the preceding level was too easy, too hard or too long, causing users to lose interest. Optimising the difficulty level is seen to be key to driving user engagement and retention. Modifications of games may also be guided by high-level insights drawn from aggregate data, such as the gender and age ranges of its users and locations of active users.

There is little evidence that personal data is commonly used

In addition to pseudonymous and aggregate information, games played through social media also result in collection of personal information linked to an individual's social network profile. This is an important distinction between mobile apps and apps accessed through social media platforms (predominantly Facebook). Developers of mobile apps may also be able to collect some personal data (for example, if users have created an account with the developer or if the game is being played through a social media network) but only for the group of users actively opted in to this option. However, from our research and interviews with app developers we have no evidence to suggest that developers are commonly using personal data for analytical purposes or sharing this with third parties. Analysis of pseudonymous or aggregate data for game improvement purposes appears to take place largely in-house, though third-party analysis tools may be used. Indeed, given the close link between game usage insights and monetisation, games developers appear to be keen to conduct this type of analysis themselves, as gleaning insights from game user behaviour may result in a competitive advantage. Insights from data analysis on the usage of one app may help not only in improving game

design and monetisation strategies in that app, but also in doing so for future apps.

Data allows targeting of in-app advertising

In terms of advertising within mobile games apps, developers may display adverts from third-party 'ad networks', advertise other games in their portfolio, or advertise other developers' games as part of a reciprocal advertising arrangement. Third-party ad networks may be used to target advertising on the basis of pseudonymous data collected by the ad networks and linked to a device identifier. In the case of games played through social media, games developers can present adverts 'in-app' using ad providers that follow Facebook's advertising guidelines.[18] In the case of games developers only advertising their own games, simple logic may be used to select ads for display to a user in combination with information such as: (i) its own games already downloaded by a device and which should not therefore be advertised; (ii) previous advertising to the user (for example, so as not to repeat ads already displayed the previous day); and (iii) the user's propensity to make in-app purchases.

1.7 Transparency of data use

Websites typically publish cookie and privacy policies

We have found that there appears to be a fair amount of information made publicly available by data collectors in the three case study sectors about their activities related to data collection and use. To a large extent this reflects the legal requirements that data collectors must abide by. For example, as discussed in Box 1, firms typically notify consumers about the use of cookies on their websites, in accordance with the ePrivacy Directive.

There is also widespread use of privacy policies on the websites of firms that collect and/or use consumer data, explaining in varying degrees of detail how it is used, who it is shared with and why, especially when the types of data involved include personal data. Where data is shared for analysis purposes, policies generally state that it is shared with 'fitness for purpose' and with privacy concerns in mind, with identifiers removed and fields aggregated where this is compatible with the required data analysis.

Privacy policies can be ambiguous

Nevertheless, there is sometimes ambiguity within privacy policies with respect to the role of third parties and the degree to which data may be shared with third parties (e.g. 'we may share this data with third parties for purposes including analysis'). There may be good reasons for this – if the scope for use of consumer data is

[18] Facebook provides a list of ad providers that have agreed to Facebook's terms and conditions: https://developers.facebook.com/docs/adproviders

changing relatively quickly, the types of analysis that might become possible and valuable in the future may not yet be fully known. Nevertheless, this means that consumers are unlikely to be able fully to anticipate how their data might be used by third parties, which raises questions about the nature of their initial consent.

In our three case studies, we consider in more detail the level of communication by data collectors about the collection and use of consumer data, and whether there appears to be consumer awareness of these practices.

1.8 Future uses of consumer data

Views differ on future prospects for data use

Both within and across sectors, we have found a diversity of views about the additional value that new developments in the collection and use of consumer data might yield. While the general sentiment is positive, there is considerable variation in the extent of investment in initiatives seeking to make innovative use of consumer data. This variation appears to be driven by a number of factors:

- differing views on the scale of likely benefits from new or more effective use of data. For example, in the motor insurance sector, stakeholder views differ on the extent to which 'big data' can significantly improve the accuracy of predictive risk models;
- the nature and scope of firms' activities. For example, firms with many products may view any increased possibility for cross-selling as a relatively large potential benefit of consumer data use (which we see to an extent with gaming apps);
- the perceived downsides of potential new approaches to consumer data collection and use. This may include privacy concerns and the potential for consumer backlash against practices perceived as invasive; and
- uncertainty about future changes to regulation, which may influence the feasibility and costs associated with particular uses of consumer data. For example, smaller players may not have the resources or capabilities to meet stricter compliance requirements.

2 The motor insurance sector

2.1 Summary

Uses of data and potential value

The motor insurance sector has always relied upon data about customers to support its decision-making and business practices:

- Data helps firms to assess individual risks and set premiums accordingly. Each individual element of data may only be a crude predictor of risk, which is why firms seek to collect a range of information. Accurate risk evaluation helps to offer the most competitive quotes while ensuring that customers' premiums are sufficiently high to cover expected costs. Therefore, a firm's predictive risk modelling can be a key competitive asset.
- Data allows firms to verify customer information and detect mistakes or possible instances of fraud. There are strong incentives to reduce losses related to fraud, which add an estimated £50 to the average general insurance premium.
- Data allows firms to reach consumers through targeted marketing, potentially generating additional revenue.

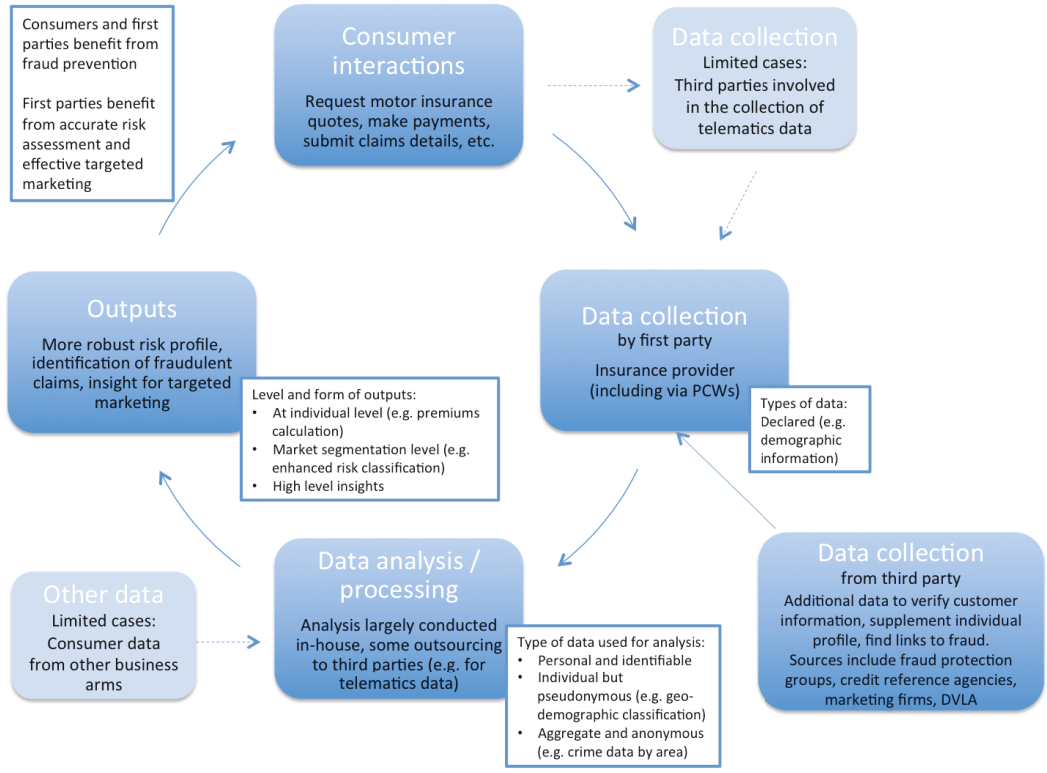
How is data collected?

Therefore, insurance firms have strong incentives to collect and use data, which they gather:

- from the consumer (e.g. at point of quote or when making a claim), including directly or through price comparison websites (PCWs);
- from commercial third-party sources (e.g. credit reference agencies, marketing firms), including on a subscription basis or through a pay-per-inquiry model; and
- from non-commercial third-party sources (e.g. the DVLA, non-profit fraud protection membership groups).

The processes by which the motor insurance sector collects and uses data are summarised by the diagram below.

Figure 2: Data flows involved in the collection and use of consumer data in the motor insurance sector



Source: DotEcon and Analysys Mason

Recent developments

Technological developments have facilitated changes to the data flows involved in the sector:

- The prominence of PCWs has allowed consumers to provide their information and request quotes from a large number of providers.
- Location-based driving data can be collected through telematics devices or apps, providing data to inform risk evaluation and claims management.
- Shared fraud databases have existed for many years, but some innovations may increase the range of relevant third-party information and technologies available to combat fraud, including in real time (e.g. from the DVLA).
- There is evidence that specific types of data (e.g. social media information) have the potential to add value by enhancing risk evaluation and fraud detection.

However we understand that some of these are relatively new developments and their impact so far is limited:

- Telematics policies are a niche (but growing) segment of the market. The use of telematics data is still evolving and third-party specialist firms are involved in developing more sophisticated analytical models. Barriers may exist in the form of consumer privacy concerns and a lack of standardisation for telematics data.

- Although there is evidence of firms increasingly using or experimenting with new types of data (e.g. grocery purchasing behaviour), the significance of this trend may be over-emphasised by third-party data or analytics providers. We have found that these new areas have so far had a limited impact on how firms operate and there is uncertainty over any future impact. Firms may be mindful of data protection concerns, of any risks of negative public reactions to expanding data collection and use, and of any impact of enhanced risk classification on risk pooling.
- Some innovations in information verification and fraud detection are still gaining traction – e.g. firms are only gradually adopting the DVLA's MyLicence service and social media information may only be useful in rare cases.
- For marketing (e.g. targeted communications and cross-selling), firms often do not rely on large volumes of consumer data or on extensive analysis, though there is some evidence that this is an area of increasing focus for firms. Targeted marketing, for example based on renewal date information, appears common with third-party firms involved in providing renewal date information or other leads, as well as administering campaigns.

2.2 Sector overview

In this section we present a review of the motor insurance sector in the UK, including an overview of the size of the sector, the firms involved and the competitive dynamics, along with recent trends.

2.2.1 Scale of the sector

All vehicles in use on public highways are required by law to have motor insurance. Following its market investigation into the private motor insurance (PMI) sector, the CMA noted that there were 34.5million vehicles registered in the UK in December 2012.[19] This figure had increased to 35.9 million by September 2014.[20]

[19] CMA, 24 September 2014, 'Private motor insurance market investigation Final report', §2.2.

[20] Department for Transport, December 2014, 'Vehicle licensing statistics: July to September 2014, https://www.gov.uk/government/statistics/vehicle-licensing-statistics-july-to-september-2014

The size of the sector is in excess of £10 billion

The CMA focused on the largest segment within this sector, private motor insurance,[21] estimating that the total gross written premium (GWP) for vehicles in this segment was in excess of £10 billion in 2012. Similarly, Datamonitor estimated that the GWP for private motor insurance was £10.8 billion in 2013.[22] Deloitte has estimated that the GWP for the entire UK motor insurance sector (not confined to private motor insurance) was around £13 billion in 2013.[23]

Statistics from the Association of British Insurers (ABI) show that the UK motor insurance industry made a £53 million underwriting loss in 2013, not having made an underwriting profit (based on revenues from premiums)[24] since 1993, though the broader UK general insurance industry saw an underwriting profit of £1.4 billion.[25] Alongside revenue from premiums, motor insurance providers receive revenue from other sources, such as investment income and the receipt of referral fees from third parties.

Premiums have declined in recent years, following sharp increases

According to Deloitte's figures, there was relative stability in the GWP between 2003 and 2009, followed by sharp increases in 2010 and 2011 and then some decline subsequently. The CMA's analysis showed that the price of motor insurance had historically tended to increase faster than the rate of inflation, with a particularly rapid increase from 2009 to 2010. Increasing premiums have been the subject of scrutiny, for example by the House of Commons Select Committee,[26] though average premiums have since entered a downward trend.[27] During times where prices increased, possible causal factors include:[28]

---

[21] This was defined as insurance for privately owned cars used for non-business purposes, of which it was estimated there were 25.7 million in the UK (around 75% of all the vehicles registered in the UK).

[22] Datamonitor, July 2014, 'UK Private Motor Insurance: Market Dynamics and Opportunities', http://www.datamonitor.com/store/Product/uk_private_motor_insurance_market_dynamics_and_opportunities?productid=CM00248-011

[23] Deloitte, September 2014, 'Gibraltar Motor Insurance Seminar 2014', http://www.gii.gi/wp-content/uploads/2014/10/Deloitte-Gibraltar-Insurance-Presentation-for-GII_30-Sept-2014.pdf

[24] An underwriting loss occurs when expenses incurred and claims paid out exceed premiums collected on insurance policies by the insurer.

[25] ABI, 2014, 'UK Insurance KEY FACTS'.

[26] Transport Committee, September 2011, 'The cost of motor insurance', and follow-up reports, http://www.parliament.uk/business/committees/committees-a-z/commons-select/transport-committee/inquiries/cost-of-motor-insurance/

[27] See e.g. ABI, April 2014, 'ABI average motor insurance premium tracker – Q1 2014 data', https://www.abi.org.uk/News/Industry-data-updates/2014/04/ABI-average-motor-insurance-premium-tracker-Q1-2014-data

[28] http://www.parliament.uk/briefing-papers/SN06061.pdf

- the effect of cyclical patterns within the insurance industry (where rates may be set relatively low for several years, followed by a 'catch-up period'[29]);
- an increased incidence of whiplash injuries (including bogus or exaggerated claims), increasing claim costs; and
- increased levels of fraud in general, which can occur particularly during periods of poor performance of the economy.

Fraudulent claims remain a key issue

In particular, the issue of fraudulent whiplash claims and the role of 'no win, no fee' personal injury lawyers has received significant attention. Government reforms have aimed to address this problem, including by banning the payment of referral fees in relation to personal injury cases from 2013.[30] Nevertheless, the volume of fraudulent motor insurance claims continues to be a key issue in the sector. To put this into context, in 2014 motor insurers paid out over £6 billion for a total of almost 3 million claims;[31] by comparison, in 2013 insurers detected a total of 59,900 bogus or exaggerated motor insurance claims, with an associated value of £811 million.[32]

2.2.2 Firms participating in the sector

Various types of firms offer motor insurance to consumers

Various types of firms participate in the motor insurance sector:

- insurers;[33]
- brokers;
- price comparison websites (PCWs);
- third parties involved in policy administration and in the claims management supply chain (e.g. vehicle repairers, solicitors); and

[29] One explanation for such patterns is that one insurance product may temporarily be used as a loss leader to attract customers and cross-sell other products.

[30] The Legal Aid, Sentencing and Punishment of Offenders Act 2012, http://www.legislation.gov.uk/ukpga/2012/10/contents

[31] ABI, 2014, 'UK Insurance KEY FACTS', https://www.abi.org.uk/Insurance-and-savings/Industry-data/Key-Facts-2014

[32] ABI, 30 May 2014, 'Insurance cheats feel the heat – value of fraudulent claims uncovered by insurers hits record level', https://www.abi.org.uk/News/News-releases/2014/05/Insurance-cheats-feel-the-heat-value-of-fraudulent-claims-uncovered-by-insurers-hits-record-level

[33] We adopt the terminology used in the CMA's PMI market investigation, where 'insurers' are those firms that underwrite insurance policies, such as Direct Line Insurance Group PLC and Aviva PLC.

- other third parties that provide data or services to the motor insurance sector (e.g. credit reference agencies, anti-fraud groups, marketing firms, technology firms).

Insurance is provided by an insurer or a broker

From a consumer's perspective, insurance policies may be provided directly by an insurer, or indirectly by a broker who acts on the insurer's behalf but does not underwrite the policy. Brokers usually offer products from multiple insurers, using their knowledge of risks and the insurance market to advise the customer and arrange suitable policies. However, some brokers may act as agents on behalf of only one specific insurer.[34] For simplicity, insurers and brokers can both be termed 'insurance providers', since they are the firms selling the policy to the customer.

The sector has been affected by the growth of PCWs over the last decade or so.[35] PCWs do not provide insurance; they provide a platform that allows consumers to request quotes from a vast range of insurers and brokers (in excess of 100).[36] They obtain the majority of their revenues from fees paid by insurance providers when a sale is made through the PCW.[37] The CMA found that the four largest PCWs have fairly similar market shares and together account for the vast majority of the PCW market.[38]

Online has become a major sales channel

Since PCWs may act as intermediaries between consumers and insurers or brokers, while brokers also act as intermediaries between consumers and insurers, there are various possible routes by which consumers may purchase a policy. The CMA found that, for the largest ten insurers, 37% of consumers bought motor insurance (including renewals) direct from the insurer – 20% online and 17% by telephone. Purchases were also made through brokers (31%)

[34] For example, Drivology is an appointed representative of AXA Insurance UK PLC.

[35] The first PCW for motor insurance was launched in 2002. See http://www.confused.com/about-us

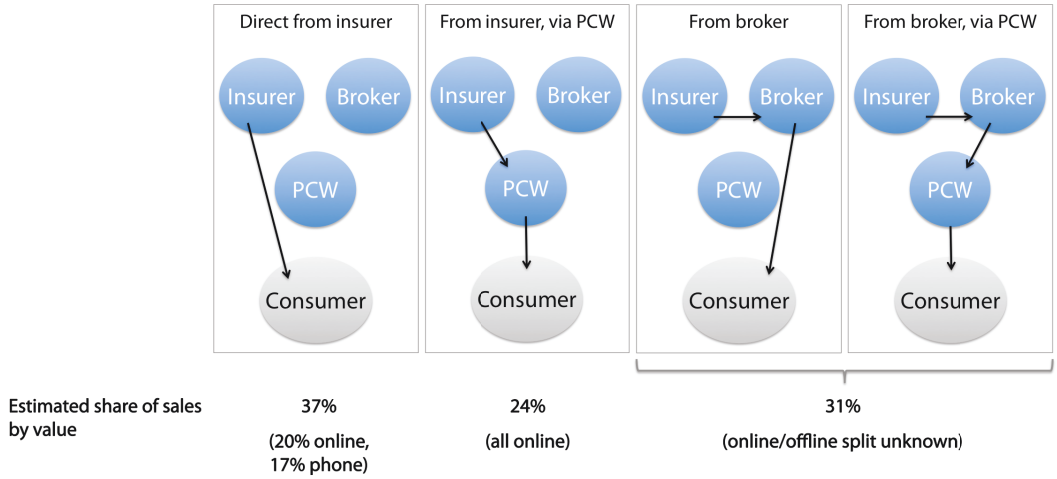
[36] The CMA noted that, though disintermediation has been a trend in the last 20 years, competition between direct and broker businesses has been blurred in recent years by PCWs, as PCW users will be largely unaware of whether the policy is being arranged through a broker or directly with the insurer. We note that, since PCWs obtain quotes from brokers and insurers, a given insurer may effectively provide multiple quotes (through different brokers) in response to a single consumer request on a PCW.

[37] FCA, 2014, 'Price comparison websites in the general insurance sector', TR14/11, p. 13, https://www.fca.org.uk/your-fca/documents/thematic-reviews/tr14-11

[38] These are Comparethemarket.com, Confused.com, Gocompare.com and Moneysupermarket.com.

and through PCWs (24%).[39] When purchasing through brokers, consumers may still purchase online (either from the broker's website or via a PCW); therefore, online purchases seem likely to make up the majority of overall purchases, though exact figures for this is not available.

Figure 3: Main purchase channels for motor insurance policies



Source: CMA, 24 September 2014, 'Private motor insurance market investigation Final report'

Price comparison websites are widely used

The CMA noted that sales through PCWs were particularly prevalent for new business (i.e. excluding renewals) – around 55% of new motor insurance business was being initiated through PCWs. PCWs are now widely used by consumers looking for motor insurance; according to the CMA's research, around 77% of consumers make use of PCWs.

Some firms in the insurance sector are vertically integrated

There is also some vertical integration between firms. At the time of the CMA's report:

- of the four large PCWs, three were owned or part-owned by insurance providers;[40] and
- three of the largest ten insurers also owned brokers.

[39] Percentages reflect the share of GWP. The remaining 9% is accounted for by other channels, including sales made through partnerships with retailers or banks. See Table 2.4 of CMA, 24 September 2014, 'Private motor insurance market investigation Final report'.

[40] Since the market investigation report was published in September 2014, the CMA has decided not to refer the anticipated acquisition by esure Group plc of the remaining 50% of Gocompare.com that it does not already own. See CMA case ME/6495-14, https://www.gov.uk/cma-cases/esure-group-plc-gocompare-com-holdings-limited

Many other third-party firms are involved with the motor insurance sector

As well as insurance providers and PCWs, many other types of firms may have some indirect involvement in the motor insurance sector. In relation to accidents and claims, insurance providers may engage with other firms such as claims management companies, car hire companies and repairers. In relation to the collection and use of consumer data, firms in the motor insurance sector may engage third-party firms, as well as interacting with public bodies such as the DVLA. The collection and use of consumer data is explored in Sections 2.3 and 2.4.

2.2.3 Competitive dynamics

The competitive dynamics in the market have been examined as part of the CMA's PMI market investigation. Competition occurs between PCWs as well as between insurance providers. To a limited extent, insurance brands that choose not to be listed on a PCW will also compete with the PCW in terms of vying for direct contact with consumers.[41]

Competition between PCWs

The motor insurance services offered by the four main PCWs do not appear to be strongly differentiated with regard to functionality, though the PCWs spend heavily on advertising, which may create some differentiation between PCW brands. The CMA found that price competition between PCWs was being weakened by 'wide most favoured nation (MFN) clauses' – agreements between a PCW and an insurance provider that restrict the insurance provider's ability to charge a lower premium through a different sales channel. In negotiating such agreements, the CMA found that each PCW benefited from unilateral market power in negotiating these clauses with providers, because some consumers only used one PCW and therefore could only be reached through that PCW. Partly as a result of the clauses, PCWs faced a limited threat of entry and expansion.[42] In practice, such agreements could have the effect of limiting price competition both between different PCWs and between PCWs and insurance providers, though the use of wide MFN clauses or equivalent arrangements has been banned from April 2015.[43]

[41] CMA, 24 September 2014, 'Private motor insurance market investigation Final report', §8.5

[42] Ibid., §56

[43] See https://www.gov.uk/government/news/cma-publishes-final-motor-insurance-order

Competition between insurance providers

The market for motor insurance exhibits relatively low concentration; according to the CMA's report, the largest ten insurers accounted for around 70% of the GWP in 2012. Moreover, the widespread use of PCWs has the potential to reduce consumer search costs to the benefit of competition. Indeed, the CMA "*found evidence that price competition between insurers on PCWs was strong and that PCWs had increased competition between PMI providers*".[44]

Nevertheless, the CMA found that providers sometimes limit the information available to consumers in ways that could dampen competition. Insurance providers sometimes offer add-ons that are not easily compared by consumers and which may hamper customer search efforts to identify the best value packages. Similarly, insurance providers sometimes do not provide clear information about the benefits of no-claims bonus protection, which can prevent consumers from choosing the best policy. The CMA has acted to improve the information available to consumers, effective from August 2016.

Differentiation between providers

Insurers may compete on non-price factors and differentiate themselves through their business models. Examples include insurance brands such as Marmalade and Drive Like A Girl, which appear to focus on the offer of telematics policies and on attracting particular consumer segments (respectively, young or inexperienced drivers, and female drivers). There may be other forms of differentiation, such as providers' network of approved repairers, not easily visible to policyholders and therefore not a core dimension of competition.

Similarly, the broker segment of the market is diverse, with small and large brokers co-existing and with considerable differentiation between companies. The CMA identified three main broker types:

- traditional brokers, which use a range of distribution channels, including branches, and might be marketed under a well-known brand (e.g. M&S);
- specialist brokers, which use a range of distribution channels, not including branches, and cater to specialist needs (e.g. Endsleigh targets students in particular); and
- online direct brokers, which use online distribution only.

Vertical integration and competition

In theory, the prevalence of vertical integration between PCWs and providers might have the potential to affect competition, though the CMA has not identified adverse effects on competition at present.

[44] CMA, 24 September 2014, 'Private motor insurance market investigation Final report', §53

Box 2: CMA decision on Gocompare/esure merger

The CMA's recent decision of 2 March 2015 provides an insight into the possible competitive effects of vertical integration between a PCW (Gocompare) and a provider (esure).

There might be incentives for the merged entity to engage in anticompetitive strategies, for example:

- customer foreclosure – favour esure over rival providers on the PCW (e.g. excluding rivals from the PCW or manipulating rankings); and
- information sharing – use information gained through the PCW to increase esure's margins or gain a competitive advantage (e.g. by inferring competitors' pricing algorithms).

However, the CMA found that existing regulation and monitoring within the motor insurance sector, together with providers' ability and tendency to monitor PCWs' conduct continuously, would be a sufficient countervailing force. The strategies would present risks for Gocompare (e.g. reduced quality of service, leading to lower market share) that significantly outweigh the potential benefits for esure. Therefore, the CMA found that the merger was unlikely to substantially lessen competition.

2.3 Collection of consumer data

The sector has always been data-intensive

Insurers, brokers and PCWs collect detailed information about consumers. Data is provided directly by consumers at the point of quote, as well as post-quote, for example as part of the claims process. Insurance providers will routinely receive additional information from third parties for such purposes as risk profiling, identify verification, fraud detection and marketing. With regard to risk specifically, each individual element of data collected (e.g. age) may only be a crude predictor of the likelihood of a claim. For this reason, firms have always sought to collect a wide range of information that jointly provides a fuller picture of risk.

In the following sub-sections, we provide an overview of these data collection practices.

There are new trends and a possible expansion of data collection

There is evidence of the sector undergoing a process of change, with insurance providers seeking to collect more data as enabling technologies have developed. A clear example is the use of telematics devices or smartphone applications (apps) to collect driving data. However, this is arguably part of a broader trend of increased data gathering and analysis, reflecting the direct commercial benefits to insurance providers from improved data availability (for purposes such as risk assessment, fraud detection and marketing). For example, while insurers routinely collect various types of data to predict individual risk, the data remains an imperfect proxy and there could be a substantial commercial advantage from collecting an additional piece of consumer data

that produces even a small improvement in the accuracy of risk assessment.

Box 3: A data-gathering 'arms race'?

Paul Evans, CEO of AXA UK and chairman of the Association of British Insurers (ABI), recently claimed that "*the sector is locked in a data-gathering "arms race" as they seek ever more information to help predict whether individual policyholders are likely to put in claims*".[45]

An Aviva representative recently stated that Aviva intends to "*expand our data footprint, as all organizations are doing at the moment, and taking advantage of all the new data sources that we get from Web, social, call data, etcetera*".[46]

RSA has stated that "*[a]s the volume and availability of data around the world continues to grow, we are increasingly looking to integrate new data into our processes*".[47]

Reflecting these statements, a recent global survey of insurance CEOs by PwC found that "*analytics and data are insurance CEOs' top transformational priority*".[48]

This trend may have negative perceptions among consumers. A recent survey of UK general insurance policyholders found that 72% were worried about the amount of information insurers have access to.[49]

Future data collection practices remain uncertain

Our interviews with stakeholders indicate that firms in this sector have always had strong incentives to gather data, but they are mindful of technological developments that can make it easier and cheaper to collect and process a wider range of data. While there do appear to be current moves to investigate new data sources and integrate more data into various business processes, this trend may well still be in its infancy. The role of 'big data' in the industry has been emphasised by third-party data and analytics providers with

[45] FT, 1 February 2015, 'Insurers warned to use 'big data' responsibly', http://www.ft.com/cms/s/0/08f6049c-a7cd-11e4-8e78-00144feab7de.html

[46] Teradata, 1 May 2014, 'Aviva: Driving Forward with Data to be a World Class Digital Insurer', http://blogs.teradata.com/customers/aviva-driving-forward-with-data-to-be-a-world-class-digital-insurer/

[47] RSA, 2014, 'Annual report and accounts 2013', http://rsaar13.g3dhosting.com/system/files/ar13/rsa_annual_report_2013_final.pdf

[48] PwC, 2014, 'Doing more with more: How P&C insurers are creating an information advantage with 3rd party data', http://www.pwc.com/en_US/us/insurance/publications/assets/pwc-third-party-data-insurance.pdf

[49] Consumer Intelligence research, as reported in Your Wealth, November 2014, 'Third of UK insurance customers unaware insurers use social media to check claims', http://www.yourwealth.co.uk/features/third-of-uk-insurance-customers-unaware-insurers-use-social-media-to-check-claims

an interest in selling their services. However, our interview evidence with insurers themselves indicates that the impact on firms' operations and processes thus far is relatively limited.

Insurers told us that they are aware of public concerns that could be caused by the use of data in the industry, and of the need to respect the relevant permissions and legislation. The possibility of some form of self-regulation regarding the collection and use of data has been mooted, but it is not clear at the time of writing how this idea might progress in the sector.[50]

2.3.1 Data collected directly from the consumer

We first outline the ways in which first parties collect information directly from individuals, either at the point of quote or thereafter.

At the point of quote

Firms collect information about the driver, vehicle and location

A variety of information is provided directly by the consumer at the point of quote, regardless of whether the consumer is interacting with a PCW, a broker or an insurer. The types of information collected in this way can be classified into three broad categories:

- information about the driver (e.g. date of birth, occupation, no claims bonus (NCB), claims and convictions record, home ownership, annual mileage);
- information about the vehicle (e.g. make, model, value, transmission, security devices, modifications, year, colour); and
- information about the location (e.g. address, whether the vehicle is kept on the road, on a driveway or in a garage).

In addition to the types of information above, primarily required in order to evaluate risk and detect potential fraud, other information collected will usually include:

- the policy renewal date;
- the consumer preferences required to produce a quote (e.g. type of cover, voluntary excess); and
- any selected pieces of information unrelated to the motor insurance policy (e.g. home insurance renewal date).

[50] Pinsent Masons, February 2015, 'Big data and insurance: should there be a code of practice?', http://www.out-law.com/en/articles/2015/february/big-data-and-insurance-should-there-be-a-code-of-practice/

Consumers will usually be asked for a preference regarding the use of their personal information for marketing purposes. Online, this typically takes the form of a 'soft opt-in', where consumers may be contacted by email for marketing purposes unless they opt out.[51]

Impact of PCWs on data collection

PCWs collect the above information and pass it on securely to a large number of insurance providers simultaneously, who then return quotes to the PCW that are displayed to the end user. This has a number of implications.[52]

Providers now receive more requests for quotes

First, as the majority of consumers now use PCWs, which pass their information on to a large number of providers (PCWs allow quotes to be requested from over 100 insurance providers), insurance providers are now typically collecting information from a greater number of potential customers than would have been the case previously (even though only a fraction of these will be converted to actual customers).

PCWs must collect all information required by all providers

Second, PCWs must collect data if it is required by *any* of the participating providers in order to produce quotes, meaning that the range of information being collected from each consumer may have expanded due to PCWs. One insurer expressed the view that this has produced a large increase in the range of consumer information available to insurers; in contrast, another insurer told us that it would routinely have collected roughly the same range of information that is collected by PCWs, such that there has not been a material expansion. The divergent views might reflect different approaches with regard to the breadth of information considered necessary at point of quote; insurers who previously collected less information may have found that the information available at the point of quote has increased due to PCWs, whereas those insurers already collecting a broader range of information may have seen little or no effect.

PCWs and providers are data controllers

Third, the role of PCWs has potential implications from a privacy and data protection perspective. PCWs act as data controllers and hold responsibility for their use of that data, but once the data is passed on to a large number of providers, each of those becomes a data

[51] For a full explanation, see https://ico.org.uk/for-organisations/marketing/

[52] In contrast, an alternative service, 'Google Compare', collects similar data to PCWs and uses this to provide a price comparison service akin to those provided by the four main PCWs. However in contrast to the PCWs, it does not share personal information with any insurer at the point of quote. Instead it generates the quotes by using the information to "look up" quotations from different insurers using tables provided by the insurers. If (and only if) the user decides to apply for a policy with one of the insurers listed or is happy to be contacted by the insurer directly, Google will pass on the relevant information to the insurer in question to allow the user to proceed with the application.

controller with regard to its own use of the data. This complexity is reflected in some PCWs' privacy notices.

Box 4: Extracts from PCW privacy policies

Extract from Confused.com Privacy Policy[53]

"When using the insurance or other services offered through our site, it is necessary for us to pass your information to the organisations from whom you will receive quotations you have requested from our site. ***While we remain data controllers in relation to that information, in connection with everything that those organisations do with that information, they will also be data controllers. Their privacy statement, setting out what they do with your personal data, will be available to you on their site****."* [emphasis added]

Extract from MoneySuperMarket.com Privacy Policy[54]

"*As part of using our Services you consent to us disclosing your personal information to the following parties:*

a. Third parties, including but not limited to companies whose products or services are included on our website with a view to providing you with an online quotation for the product and/or service requested by you. When these companies use your data in this way, ***they will be acting as data controllers of your information and therefore you should view their privacy policies*** *or contact them directly for further information.*" [emphasis added]

One insurer told us that, since PCWs are now largely responsible for consumer data collection at the point of quote, insurers rely on PCW privacy policy statements being adequate and including the relevant permissions. For this reason, the policies are monitored carefully.

Post-quote

A limited amount of additional information may be collected from the consumer at the point of sale, where the additional information is needed to process payments and administer the policy. This will include payment details and the vehicle registration number (which does not necessarily need to be provided at the point of quote).

Post-sale, the insurance provider will typically keep a record of all interactions with the consumer, such as changes of details.

When a claim is being made, the insurance provider will collect the necessary details needed to process the claim, such as the time, location and description of the accident, and the details of the other

[53] http://www.confused.com/privacy-and-security/privacy-policy

[54] http://www.moneysupermarket.com/legal/privacypolicy.asp

party or parties involved in the accident, as well as confirming details such as the policyholder's demographic information and vehicle details. Collecting such information is also important in order to detect possible cases of fraud.

2.3.2 Data collected via third parties

Providers have incentives to 'enrich' data about the driver, vehicle and location

Insurance providers typically collect additional information from a number of third-party sources, supplementing the information provided directly by the consumer. This practice may be referred to as 'data enrichment'. This can take place in real time at the point of quote, but also post-quote including when a claim is made, and serves various purposes: to verify information provided by the individual, to detect possible fraud and to build more accurate risk profiles.

One insurer told us that collecting data from third parties, rather than from the individual, presents various possible advantages:

- it allows cross-checking of data provided by the individual, helping to detect mistakes or deliberate misrepresentation;
- it can streamline the quote process by reducing the volume of information that the individual is required to provide; and
- it does not reveal the use of particular variables, which can be a source of competitive advantage for an insurer.

Again, the types of information that are obtained from third parties can be grouped into broad categories, according to whether it relates to the individual, the vehicle or location. These topics are outlined below, followed by a discussion of how data enrichment activities differ among firms in the sector.

Information about the driver

Additional information is gathered 'within industry'

Various initiatives exist among insurance providers, and also other stakeholders, to enable information about individuals to be collected and shared. This is predominantly with a view to validating information provided by the consumer and tackling fraud. Examples of widely used data sources are provided in Box 5 below.

Box 5: Data collection and sharing to combat fraud

Various non-profit organisations have been established, with insurance providers as members, collating fraud-related data and making it accessible to members individually. Typically these operate on a principle of reciprocity, requiring members to participate by submitting relevant information in order to be allowed to query the database.

For example:

- The Claims and Underwriting Exchange (CUE) database is widely used across the sector. It is a central database established and managed by a non-profit company, Insurance Database Services Limited, on behalf of its members, which include many insurance providers and local councils. The database holds details of motor, home and personal injury incidents reported to insurance companies, which may or may not give rise to a claim. Thus, it can help prevent multiple claims fraud and the misrepresentation of claims histories. Participating insurance providers will both contribute data to the database and receive individual consumers' records upon request. This may be done through a third-party intermediary with a licence to offer CUE-related services (e.g. Experian and CRIF).
- Another widely used database is the National Fraud Database managed by Cifas, another non-profit organisation. The database holds information on hundreds of thousands of cases of confirmed fraud. Again, many insurance providers are among the members, who contribute information to the database as well as accessing it.
- The Insurance Fraud Investigators Group is another member organisation that facilitates the sharing of fraud-related information, reportedly covering the majority of the insurance industry.
- The Insurance Fraud Bureau is funded by the insurance industry and, among other things, it manages the Insurance Fraud Register. The Register, sponsored by the ABI, is described as the first industry-wide database of known insurance fraudsters.

The arrangements above can be seen as within-sector-sharing arrangements, in the sense that they effectively allow an insurance provider to access information that has been contributed by other insurance providers.

Further information also comes from 'external' third parties

In addition to such arrangements, insurance providers routinely access information about individuals from third-party sources external to the motor insurance sector. For example, credit scores and other information from public credit records is widely collected by insurance providers from credit reference agencies. Other types of third-party firms, such as data brokers, may supply further information about consumers, such as renewal dates. Increasingly, the DVLA is providing information about the individual (driving entitlements, length of licence and any motoring convictions) through its MyLicence service.

Information about the vehicle

Insurance providers may obtain additional vehicle information from third parties. The DVLA makes various pieces of information available based on the registration number, such as the year of manufacture, fuel type and colour. Commercial third-party data providers can also provide vehicle information, potentially with greater detail, such as the vehicle's acceleration characteristics, estimated value and history. For example, one commercial service offers up to 200 fields of data per vehicle.[55]

Information about the location

Insurance providers can use third-party data sources to gain further information based on the individual's location.

Some freely available data may be collected from Government websites to obtain information about a specific location (e.g. based on postcode). This can include relevant crime data and data about road accidents in each location.

Some insurers purchase additional location-based information from third-party data providers. Companies such as Experian and Callcredit have built geo-demographic segmentation systems which classify consumers into different consumer 'types', for example at postcode level or at household level. As such, this may be regarded as inferred information about the individual, although in practice it is attached to a location rather than to a named individual.

Business models of third-party data providers

Where firms in the motor insurance sector obtain third-party data, a variety of business models and financial arrangements may apply, though details in this area are scarce and subject to commercial confidentiality.

Data providers may charge subscription fees or fees per enquiry

Some types of third-party data may be accessible for free, such as publicly available data (e.g. crime statistics), or data supplied to insurance providers by the DVLA or by fraud protection groups (contingent on being a member). Even in such cases, commercial arrangements with third parties may exist, for example where third

[55] CDL Vehicle Information Services, see http://www.cdl.co.uk/solutions-vehicle-data.php

parties provide the enabling technical and IT infrastructure for real-time lookups, or any value-added services that leverage the freely available data. In cases where data or data-related services are supplied for a fee (e.g. from credit reference agencies or marketing firms), insurance providers might pay an annual subscription or licence fee for unlimited access to a database/use of a service, or they may be charged a fee per enquiry.

2.3.3 Data enrichment across different types of firm

There are differences in the mechanisms by which the various types of firms in the sector receive data from third parties.

PCWs' data enrichment activities are relatively limited

PCWs' data enrichment activities are fairly limited, since their primary role is simply to pass on the data collected from the individual to the insurers and obtain quotes. However, PCWs may still enrich the data to some extent. One PCW told us that it routinely obtains additional vehicle information from a third-party provider, as well as receiving data that helps to verify identity and detect fraud.

Data enrichment capabilities might differ between insurers and brokers…

There are also some practical differences in how brokers and insurers operate at the point of quote that can affect data enrichment. Based on industry interviews, we understand that when brokers provide quotes for different insurers, they usually do not contact insurers and pass on the relevant information at that point. Rather, brokers generate quotes using software platforms provided by third-party firms,[56] which apply risk-prediction models based on algorithms supplied and periodically updated by the insurers. Typically it is only at the end of each day that brokers notify insurers of policies sold on their behalf and only then is customer data shared with the insurers.

Brokers' reliance on software platforms can affect their data enrichment capabilities, as their ability to enrich data in real time at the point of quote is constrained by the functionality offered by the software. Evidence from interviews indicates that, at least for some brokers, quotes have tended to be based on the information provided by the consumer and on the 'static' algorithms used by the software, rather than using additional third-party data sources in real time to verify and supplement the information, which insurers are able to do when generating quotes.

[56] Software platforms are provided by companies such as SSP, CDL, OpenGI or Transactor. Such platforms are licensed to the brokers and typically take the form of enterprise software that executes in the broker's own IT infrastructure; however, at least in one case (CDL), this software is provided to brokers as a cloud service using infrastructure that is shared across brokers.

…Though innovative solutions may eliminate any 'mismatch'

Reflecting this, there may be something of a mismatch between the data enrichment capabilities of brokers and insurers, with insurers more likely to access a greater breadth and granularity of information at the point of quote, which might also result in competitive advantages for insurers selling policies directly. However, new infomediaries exist that work with insurance providers and software platforms to develop solutions that increasingly enrich the data available to brokers, replacing the traditional 'static' pricing approach with one that enables real-time data enrichment using third-party databases. Our interviews found divergent views on the extent to which a substantial 'mismatch' still exists today between brokers and insurers in general, though it is possible that smaller brokers are less able to take advantage of data enrichment and therefore might suffer some competitive disadvantages.

2.3.4 Collection of telematics data

Technology allows insurers to collect detailed driving data

A telematics device, commonly referred to as a 'black box', can be installed on a vehicle in order to record detailed driving data, including speed, distance, roads used, the times at which the vehicle is used, and accelerating and breaking patterns. To an extent, smartphone apps can also collect similar data. Apps may be less reliable as a means of data collection, e.g. because phones can be switched off, but they may have other advantages, e.g. lower costs. We refer to any policies that involve the collection of driving data as telematics policies, regardless of whether data is collected by an installed black box or a smartphone app (or both).

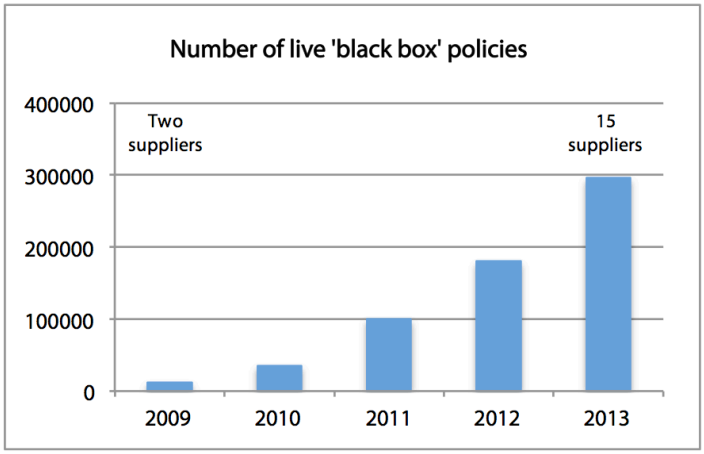
Telematics policies target young drivers

In the UK, black box policies have been marketed primarily at young drivers, a high-risk segment who typically face high premiums because of their age and inexperience, and who may be able to achieve substantial savings with such policies. Different business models have been built on the collection of driving data.[57] Nevertheless, the usual principle is for analysis of detailed driving data collected through black box devices or apps to produce driving 'scores' for individuals; premiums can then be adjusted accordingly. Sometimes there are incentives to encourage take-up of the policies, such as guaranteed discounts, whereas in other cases policies allow for both upward and downward adjustments to premiums on the basis of telematics data.

[57] For example, many policies collect telematics data continuously, but some providers may collect data for only a fixed period (e.g. for 200 miles; this is the case for some smartphone app-based business models).

According to the British Insurance Brokers' Association (BIBA), the penetration of telematics devices has been increasing rapidly and by 2013 there were around 300,000 black box policies. Multiple stakeholders in this sector quoted take-up of telematics devices being about 1–2% of the motor insurance market.

Figure 4: Take-up of black box policies in the UK



Source: BIBA, 2014, 'Research on Telematics Market'.

Overall, there is some consensus that the penetration of such devices will continue to increase, but opinions differ on how far this trend will go. According to the Association of British Insurers (ABI), "*[w]hile the market is expected to continue expanding rapidly in the immediate future, it is unclear whether Telematics Policies will remain a niche product or become mainstream*".[58] Our interviews with stakeholders confirmed that there are diverging views and a degree of uncertainty about future developments.

Third parties may be involved in data collection

As well as the insurance providers, specialist firms may have access to the data as they may be involved in providing the black box devices, collecting data and analysing it. For example, according to Admiral's privacy and security statement in relation to its LittleBox telematics policy, the data controllers are both the insurer (EUI Limited) and the technology company (Octo Telematics S.p.A.). Increasingly, car manufacturers appear to be installing in-built devices in vehicles that are capable of collecting driving data and

[58] ABI, April 2013, 'Selling Telematics Motor Insurance Policies, A Good Practice Guide', https://www.abi.org.uk/~/media/Files/Documents/Publications/Public/Migrated/Telematics/Selling%20telematics%20motor%20insurance%20policies%20-%20ABI%20good%20practice%20guide.ashx

manufacturers may come to play an important role in the data collection process in the future.[59]

Telematics and privacy

Some drivers may see the collection of detailed driving data as an invasion of privacy.[60] Such fears may not necessarily be specific to telematics insurance policies, but rather may reflect a general attitude towards the 'connected car' and real-time collection of data, which might increasingly occur through a variety of in-built sensors and monitoring systems.[61] However, telematics policies are offered purely on an explicit opt-in basis at present, which should mitigate privacy concerns, as long as policyholders are well informed about what data will be collected and how it will be used.

ABI guidance sets good practice for telematics policies

The ABI published consumer and industry guidance on telematics insurance policies in 2013.[62] The industry guidance has the status of a voluntary good practice guide and "*seeks to complement and reinforce the responsibilities set out elsewhere*", such as in the Data Protection Act 1998 (DPA). Good practice includes obtaining explicit consent, presenting customers with clear and comprehensive information about how the data would be collected and used and clarifying who would have rights of access to it. Moreover, data should not be recorded unless it is necessary for the purposes it is being used for, as declared to the customer.

[59] See Computerworld UK, August 2014, 'BMW to install usage-based insurance telematics into cars', http://www.computerworlduk.com/news/it-business/3534379/bmw-to-install-usage-based-insurance-telematics-into-cars/; Moneysupermarket, 21 March 2013, 'Telematics: The future of driving?', http://www.moneysupermarket.com/c/news/telematics-the-future-of-driving/0017203/; Pinsent Masons, 22 January 2015, 'Telematics insurance, market disruption and control of data', http://www.out-law.com/en/articles/2015/january/telematics-insurance-market-disruption-and-control-of-data-/.

[60] For example, privacy issues could relate to a general aversion to being monitored, or to more specific concerns such as a fear of third parties (e.g. hackers) accessing the data and using it for nefarious purposes, or a fear that data could be seized through official channels and used to incriminate drivers (e.g. for speeding). See e.g. Deloitte Center for Financial Services, 21 April 2014, 'Overcoming speed bumps on the road to telematics', http://dupress.com/articles/telematics-in-auto-insurance/

In our interviews, one insurer noted that, where there are multiple drivers of a vehicle, it is desirable to be able to distinguish who is driving at different times, but that even if technology allowed this it may be seen by consumers as intrusive.

[61] See e.g. BBC, 5 November 2014, 'Is your connected car spying on you?', http://www.bbc.co.uk/news/business-29566764

[62] ABI, April 2013, ''Pay How You Drive' Motor Insurance'; and ABI, April 2013, 'Selling Telematics Motor Insurance Policies', https://www.abi.org.uk/News/News-releases/2013/05/The-ABI-and-BIBA-publishes-consumer-guide

2.3.5 Collection of data via cookies

As in virtually any sector where firms have an online presence, first parties commonly use cookies or beacons. PCW and insurance providers' cookie policies show that often these include all four types identified by Analysys Mason as being part of the online data value chain[63] (that is, strictly necessary, performance, functionality and targeting/advertising). The purposes include those that are observed in most sectors:

- understanding usage of the website (e.g. how long spent on each page, commonly clicked links), which may help improve structure and functionality;
- filtering out irrelevant messages/pop-ups and personalising information; and
- evaluating and refining approaches to targeting and personalisation.

We have found no evidence that data collected using cookies is commonly used as part of risk evaluation or fraud detection. However, the motor insurance sector differs from other sectors in that firms request a large volume of consumer information before even offering a price. In other sectors, consumers would not provide firms with the same depth of information directly and therefore firms may have greater incentives to try to infer such information (e.g. demographic characteristics) from data collected via cookies.

2.4 Use of consumer data

In this section we focus on specific key uses of consumer data in the sector. Both historically and today, there is little doubt that insurance providers see risk evaluation as the primary focus of their data analysis activities. Moreover, fraud is recognised as a major on-going problem in the sector and using consumer data to try to detect and prevent fraud is also an area of high commercial priority. On top of this, data is used for other purposes, including marketing, but this appears to be an area of lesser focus compared to other sectors.

The following sub-sections examine firms' specific uses of data in the areas of risk evaluation, fraud prevention and marketing.

[63] Analysys Mason (for Ofcom), 2014, 'Online data economy value chain', http://stakeholders.ofcom.org.uk/binaries/research/online-data-value/online_customer_data.pdf

Following this, we present an overview of the ways in which data may be provided to third parties, and finally of some additional possible uses of telematics data.

2.4.1 Risk evaluation

Using data to predict risk is a high-priority area

An accurate assessment of individuals' risk, as reflected in its underwriting and pricing, is essential to an insurer's success. Therefore, insurers invest significant resources in developing and refining predictive models of risk, and even a marginal improvement in these models can be a source of competitive advantage.[64] To illustrate the importance of this use of data, one insurer told us that it had around 100 staff working in data analytics roles, of which around 85 were dedicated to risk evaluation. Another insurer confirmed that this was the most important aspect of data analysis because of the value that accurate risk prediction can create for the company.

With regard to the various factors that may be used to predict risk, there seems to be some consensus that those factors provided at the point of quote by the driver, such as age and details of driving history, are likely to remain of paramount importance in determining risk assessments. As outlined in Section 2.3.2, insurers will also typically obtain additional information from third-party sources at the point of quote and use this to cross-check and supplement the risk information collected from the consumer.

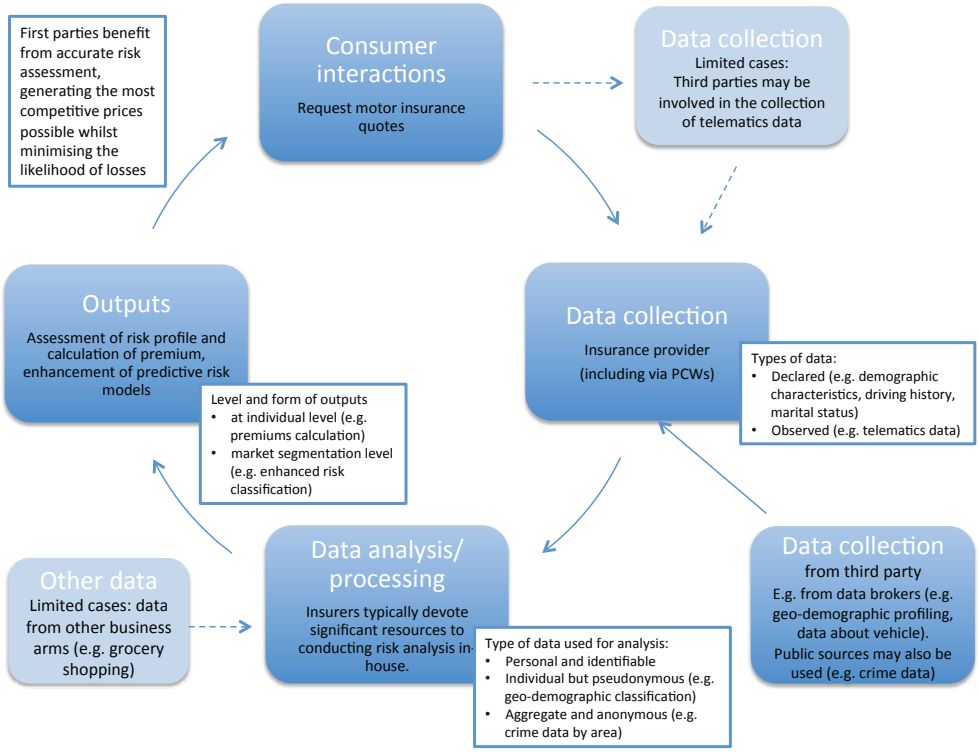
Insurers generally opt to carry out the data analysis necessary to build and refine predictive models in-house, though at times they may utilise third-party software, as well as third-party data sources. According to PwC, third-party providers "*are "mashing" various types of external and internal data when partnering with insurers and help the latter better understand and mitigate risks and predict future loss ratios*".[65]

The diagram below summarises the role of data as part of the risk evaluation process, which is then described in more detail.

[64] For example, "*[a]n insurer that discovers a new way to identify and exclude high risks improves its competitive position in two ways: it lowers its average risk and, assuming the people it rejects go elsewhere, it increases the average risk of its competitors*" (Baker, 2002, 'Containing the Promise of Insurance: Adverse Selection and Risk Classification', http://ssrn.com/abstract=322581). Similarly, an insurer with superior predictive models may be able to 'cherry-pick' low-risk individuals that its rivals have failed to identify as such.

[65] PwC, 2014, 'Doing more with more: How P&C insurers are creating an information advantage with 3rd party data', http://www.pwc.com/en_US/us/insurance/publications/assets/pwc-third-party-data-insurance.pdf

Figure 5: Data flows associated with the assessment of risk



Source: DotEcon and Analysys Mason

Use of new types of proxy data

'Big data' has been linked to risk evaluation

Insurers have incentives to experiment with new types of data in order to test for improved predictive power and consider incorporating new variables into their predictive models. While this process may take place on an on-going basis as insurers seek to refine their risk evaluation processes, it does appear to be an area of particular interest within the sector currently, even if it may have yet to produce many tangible changes. The interest relates to technological developments that have seen an increase in the breadth of consumer data being collected and an improvement in the tools available to process and analyse it.

Box 6: 'Big data' and risk evaluation

The potential impact of 'big data' on insurers' ability to predict risk has seen considerable attention in recent years. For example, an academic article argues that:

"*Big data promises new opportunities to fine tune risk classification by using algorithms to mine new and complex sets of data to find new correlations and make predictions about behavior. Carriers* [i.e. insurance providers] *can gather information about insureds from a variety of new sources, including phone records; the Internet; health records; sensors in cars and clothing, electrical grids, or communication devices. In this way, carriers' use of big data may be a natural evolution in risk classification*"[66]

Statements made by some stakeholders support the existence of this trend. For example, Aviva argues that "*[t]he increased volume, variety and velocity of data allows for greater use of predictive analytics*".[67] Therefore, a main area of focus is "*enhancing our analytics capabilities, increasing the use of internal and external data to drive superior performance across the business including underwriting, pricing and claims*".[68]

However, we have found diverging views on the extent to which new data sources will have such an impact. Some stakeholders seem more open to the idea that new data sources and variables, including from third parties, could enhance risk evaluation, while others remain more sceptical. In any case, assessing risk is likely to remain challenging and the array of data used – including any new data – will continue to be an imperfect proxy of risk.

New proxy data may be sourced internally or externally

In seeking to leverage new data sources, some insurance providers may be able to draw upon consumer data that they already hold. For example, MoneySuperMarket.com's privacy policy states that "*[s]ome product providers may carry out checks against data they already hold on you (or is held by the company whose brand they administer the product for, or members of their group of companies) such as data from existing products, account data or loyalty scheme data. They may use this data to help them assess and rate your application for a quote and determine your premiums.*"[69] On the other hand, insurance providers can also source additional data from third-party providers.

[66] Swedloff, 2014, 'Risk Classification's Big Data (R)Evolution', Connecticut Insurance Law Journal, Vol. 21, http://ssrn.com/abstract=2566594

[67] Aviva, 2014, Annual report and accounts 2013, https://www.aviva.com/library/2013ar/pdf/Strategic_report/Market_context.pdf; this is an oft-repeated statement in the literature around 'big data'.

[68] Ibid.

[69] See http://www.moneysupermarket.com/legal/privacypolicy.asp

Testing for predictive value

One insurer explained to us that, in considering whether a new type of data or variable could improve risk assessments, three questions would typically be asked:

- Is the type of data in question readily available (either within the company or from external sources)?
- Does the data have significant value?
- Based on the answers to the questions above, and taking into account any other considerations (e.g. potential privacy and data protection issues), should the data be used?

In order to establish whether a new type of data is valuable, an insurer may conduct retrospective analysis, making use of historical data held about its customers in order to establish whether the new type of data appears to be correlated with losses. In some cases, it may be necessary to transfer the data to a third party in order for that third party to attach the new data to existing customer records.

Understandably, insurers may not wish to publicise their use of particular datasets, variables or techniques in the context of risk evaluation, as this information is commercially sensitive. Therefore, providing a full account of the use of consumer data in this area is inherently difficult. Nevertheless, publicly available information allows certain candidate data types to be identified, as discussed below.

Credit ratings and financial information

An individual's credit scores and history are typically used when the individual opts to pay for insurance in instalments, as insurers seek to establish the individual's likelihood of keeping up with monthly payments and set the interest rate accordingly. However, this information might be valuable as part of individual risk assessments more generally, on the basis that empirically a correlation may exist between individuals' credit scores and their driving behaviour.[70] In the US, it was reported that "*auto insurers started to incorporate behavior-based credit scores from credit bureaus into their analysis when they became aware of empirical evidence that people who pay their bills on time are also safer drivers*".[71] Similarly, according to a UK press report, a "*strong link*" has been established between those individuals who are more prudent with their spending (e.g. stay within overdraft limits and do not miss credit card payments) and

[70] McKinsey, August 2014, 'Unleashing the value of advanced analytics in insurance', http://www.mckinsey.com/insights/financial_services/unleashing_the_value_of_advanced_analytics_in_insurance

[71] Note: 'credit bureau' is the American term for credit reference agencies. Ibid.

those that are relatively less risky drivers.[72] We note that recent research into UK general insurance holders found that 62% believed that quotes reflected individuals' financial status, though only 28% believed that it was fair for insurers to check this information.[73]

Box 7 below provides further discussion of use of credit and financial behaviour.

[72] Telegraph, July 2014, 'Thrifty drivers could save on insurance premiums', http://www.telegraph.co.uk/finance/personalfinance/10963992/Thrifty-drivers-could-save-on-insurance-premiums.html

[73] The statistics appear to refer specifically to financial information in credit records. Consumer Intelligence research, as reported in Your Wealth, November 2014, 'Third of UK insurance customers unaware insurers use social media to check claims', http://www.yourwealth.co.uk/features/third-of-uk-insurance-customers-unaware-insurers-use-social-media-to-check-claims

Box 7: Credit/financial behaviour as a predictor of driving risk

It is unclear to what extent, if any, credit scores and related information are being considered as predictors of driving behaviour in the UK. Insurance providers' privacy policies do typically state that credit checks may or will be carried out, though it is often not specified in what circumstances, or how the information would then be used. For instance, Covea states that it "*may also conduct credit reference checks in certain circumstances*".[74]

In cases where the uses of the data are specified, it may still be somewhat ambiguous whether any data is used as a driving-risk rating factor or whether it is only used for other purposes, such as identity verification and setting of the APR (where monthly payment options apply). For instance, Aviva states that "*[t]o ensure we have the necessary facts to assess your insurance risk, verify your identity, to help prevent fraud and to provide you with our best premium and payment options, the insurer may obtain information from third parties including a quotation search from a credit referencing agency.*"[75] Sainsbury's Bank states that "*[w]e carry out a consumer search when any application for insurance is submitted to evaluate insurance risks*".[76]

In interviews it was reported to us that credit reference agency data is primarily used for other purposes, such as identity verification and setting of the APR for payment by instalments, rather than being used as a rating factor in relation to driving risk, though the possibility of this practice existing in the sector was not ruled out.

One related area is the possibility of using more detailed data on financial behaviour and transactions, where available to the insurance provider – in other words, "*assessing fiscal responsibility by scrutinising bank accounts and bill payments*".[77] One motor insurer that is reported to have used consumer data in this way is Scottish Widows, Lloyd's insurance arm, seemingly using data that it holds on its customers who also bank with Lloyds, to identify individuals who are expected to be relatively safe drivers.[78] On the other hand, one stakeholder expressed scepticism to us about this practice generally being possible, since where there is separation between a company's – or a group of companies' – banking and insurance functions, it may not be possible for data to be transferred and used in this way.

Shopping data

Another type of consumer data reportedly being linked to risk is grocery purchasing data. Grocery retailers such as Tesco, Sainsbury's, The Co-operative and ASDA collect consumer data through their loyalty card schemes and also offer motor insurance

[74] https://www.coveainsurance.co.uk/a/pdf/Policyholder%20Notice.pdf

[75] http://www.aviva.co.uk/static/library/pdfs/motor/multi-car/NMDMG10248.pdf

[76] http://www.sainsburysbank.co.uk/library/default/pdf/standard_car_insurance.pdf

[77] FT, 1 February 2015, 'Insurers warned to use 'big data' responsibly', http://www.ft.com/cms/s/0/08f6049c-a7cd-11e4-8e78-00144feab7de.html

[78] Telegraph, July 2014, 'Thrifty drivers could save on insurance premiums',

to consumers. Such information might have the potential to allow inferences about household characteristics that could be useful to insurers.

Tesco's car insurance policy document states: "*[w]e may access and use information (including transactional information) from your Tesco Clubcard to allow us and your insurer to assess your premium at quotation, mid term amendment and renewal. This will only be used to have a positive impact on your premium*".[79] Indeed, it has been reported that Tesco is offering discounts of as much as 40% on insurance products to customers expected to be relatively low-risk, based on shopping habits.[80] Evidence from our interviews confirmed that analysis has been carried out on in-store purchasing data, collected from loyalty cardholders, with some evidence that purchasing patterns can have predictive value in relation to motor insurance risk, though insurers with access to such data may not yet all be using it for this purpose.

Social media information

Social media data may also be a candidate data type for predicting risk, insofar as it could reveal individual characteristics related to behaviour, lifestyle, interests and so on which an insurance provider would otherwise be unlikely to observe. This use of data has been reported to exist among life insurance providers,[81] though some insurers have denied this.[82]

In the context of the UK motor insurance sector, there is little tangible evidence to indicate that social media data is currently being used in this way, but the situation could change in the future. One insurer told us that it has conducted some research in this area, finding some evidence of social media information having predictive value for risk. However, at present it is not making use of any social media information in this way.

[79] See http://www.tescobank.com/assets/sections/carins/pdf/tesco-car-insurance-all-policy-booklet-1014.pdf

[80] Telegraph, July 2014, 'Thrifty drivers could save on insurance premiums', http://www.telegraph.co.uk/finance/personalfinance/10963992/Thrifty-drivers-could-save-on-insurance-premiums.html

[81] See for example BBC article, November 2013, 'How big data is changing the cost of insurance', http://www.bbc.co.uk/news/business-24941415

[82] FT Adviser, November 2014, 'Protection providers deny social media snooping', http://www.ftadviser.com/2014/11/10/insurance/health-and-protection/protection-providers-deny-social-media-snooping-GX8FaIt76YFF5bLsVyDjoM/article.html

Use of telematics data

Telematics data could have an inherent advantage over 'proxy' data…

Driving data collected through telematics devices or apps is fundamentally different to other types of data used by insurance providers: it is data directly related to driving behaviour, unlike the many types of proxy data discussed above. As such, it has the potential to significantly alter the basis of risk evaluation, allowing insurance providers to assess an individual's driving risk based on recorded driving behaviour over a period of time, rather than using other types of proxy data that are not directly related to driving. It could also have an incentive effect to improve driving. At present this is mainly taking place for young drivers, who historically have faced high premiums on the basis of their age and inexperience, regardless of their actual driving behaviour.

…but using the data in practice is challenging

While the potential for telematics data to enhance risk assessments is evident, achieving this may be challenging in practice. Typically, vast volumes of telematics data will be collected from individual drivers using a telematics device. The analytical process involved in deriving accurate and fair insights about their driving from this data is complex. For example, data collected about a driver's speed at a particular time may need to be contextualised in various ways, e.g. by taking into account the applicable speed limit at that location, but potentially also assessing such factors as the traffic, weather and road conditions.

The technological and IT investments required for an insurance provider to collect, process and analyse telematics data are likely to be substantial. At least partly for this reason, third-party specialist firms are often engaged to undertake such activities, though in such cases providers may still collaborate with third parties to a large extent in developing the algorithms. These third parties might also merge the telematics data with other sources, such as map data, to help contextualise it.

Stakeholder interviews suggest that the analysis and interpretation of telematics data is an area that is in evolution, with some telematics policies currently relying on fairly basic techniques and rudimentary driving scores that do not necessarily provide a full and accurate picture of driving behaviour for all drivers. This can also coincide with more basic data collection, for example, monitoring a driver's driving for a given number of miles or for a relatively short duration.

One reason why the development of sophisticated analytical models may be taking some time is that it can be difficult to build sufficiently accurate models until data from a large number of policyholders has been collected. Deloitte predicts that "*[i]nsurers will need to collect a substantial volume of such data to achieve a critical mass in order to identify potential correlations and create predictive models that produce reliable underwriting and pricing*

decisions".[83] Similarly, an insurer told us that achieving 'critical mass' could be a significant factor.

Providers might use telematics data to promote safe driving, reducing risk

Aside from helping insurers *assess* risk, an important aspect of telematics policies could be that it enables insurers to *reduce* risk to some extent, by creating incentives to change driving behaviour. Drivers who buy telematics policies may drive more safely and make fewer claims because their premium is directly related to actual driving behaviour, which drivers can control, as well as to those characteristics (e.g. age and previous claims history) outside of their control. Insurance providers may even have an active role in promoting safe driving by making detailed feedback available to drivers, e.g. through apps or online portals, which may be especially relevant for young or inexperienced drivers. Indeed, insurance providers have claimed positive effects on road safety[84] and the Department for Transport has reportedly commissioned a study exploring the effect of telematics on young drivers and road safety.[85]

Third parties often undertake analysis

Insurance providers may either opt to carry out telematics analysis in-house, or to engage the services of a specialist third-party company, of which there are several.[86] Given the substantial IT set-up costs that may be involved, the expertise that such third-party companies can offer and the tangential nature of telematics data processing to core insurance activities (at least for the moment), it appears that the majority of insurance providers opt to outsource elements of data collection, processing and analysis.

At present, there is relatively little standardisation and transparency

As different insurance providers are pursuing the development of telematics data collection and analysis separately, there is no common approach or standard either with regard to the type of raw data collected or with regard to the algorithms used to produce driving scores. Providers may wish to limit transparency over their telematics data collection and analysis activities, since expertise and intellectual property in this area could be a source of competitive advantage. Providers may also have incentives to prevent policyholders from using their telematics data when requesting quotes from alternative providers, since providers who invest in

[83] Deloitte Center for Financial Services, April 2014, 'Overcoming speed bumps on the road to telematics', http://dupress.com/articles/telematics-in-auto-insurance/

[84] See e.g. The Guardian, April 2012, 'Car insurance: satellite boxes 'make young drivers safer', http://www.theguardian.com/money/2012/apr/05/car-insurance-premiums-telematics-satellite-box

[85] Road Safety GB, November 2014, 'DfT commissions study into telematics and novice drivers', http://www.roadsafetygb.org.uk/news/4022.html

[86] For example, Octo Telematics, The Floow and Wunelli.

telematics may rely on achieving certain rates of customer retention in order to recoup their investments.[87]

A consequence of these various factors is that there is currently no portability of telematics data or driving scores for consumers who wish to obtain quotes from other providers based on evidence of their driving behaviour, collected by their current provider. (This contrasts, for example, with no claims discounts, which are typically transferable.) In theory, the lack of portability may lessen the incentives for consumers to switch, potentially dampening competition. However, we note that any such effect is currently limited by the low penetration of telematics policies at present, though it might emerge as a future concern.

Another issue from the consumer perspective is that, as a result of the substantial technical complexity in using telematics data, it is inherently difficult for insurance providers to give full transparency to their customers with respect to how their data is being used to assess driving performance (notwithstanding the fact that, as mentioned above, they may not wish to grant such transparency). While consumers who opt in to telematics policies should have a high-level understanding of the fact that their driving speed, braking patterns and so on are being taken into account in order to produce driving scores, the analytical processes that achieve this in practice effectively take place 'behind the scenes'.

Deloitte has speculated that, as telematics continue to gain traction, some consumers or stakeholders may call for greater transparency, in order to demonstrate that the data is being contextualised correctly and interpreted fairly.[88] Various stakeholders have also speculated that some form of common standards for telematics data collection and use in risk assessment may eventually emerge in the sector. Thus, the future use of telematics data for risk evaluation may evolve towards greater transparency and standardisation, but this is an area of uncertainty at the time of writing.

'Micro segments' and possible implications

If insurers increasingly enrich point-of-quote data and successfully experiment with new types of data, including telematics data, they may build ever more accurate risk models. Indeed, the ABI reports that "*[i]nsurer behaviour is also changing as they increasingly "segment" the markets available to them to pursue customers who*

[87] This also relates to a more general point about the ownership of the data.

[88] Deloitte Center for Financial Services, April 2014, 'Overcoming speed bumps on the road to telematics', http://dupress.com/articles/telematics-in-auto-insurance/

offer the best risk profile for their business model, using technology such as predictive underwriting".[89]

Enhanced risk classification could reduce risk spreading

This trend, where risk classification increasingly could be fine-tuned on 'micro segments' of consumers, might reduce the degree of risk sharing across individuals, with a "*shift from being a "big underwriting pool of risk" to more targeted and strategically set prices*".[90] As well as making risk pools (or segments) smaller, enhanced risk classification can make each pool of consumers more homogenous by putting individuals of similar risk together, thus requiring less subsidisation from low-risk to high-risk individuals.

Where risk classification is increasingly based on 'micro segments', this could conflict with any general conception of insurance as a means of spreading risks among a large pool of consumers, or even as a "*means to promote social solidarity*", with low-risk individuals subsidising high-risk individuals.[91] In an extreme hypothetical scenario where the main risk factors used by all insurers are given by an individual's telematics data, risk is assessed at the individual level and the notion of segments or pools of consumers no longer applies.

Concerns may arise if reduced risk spreading leaves consumers without cover

A specific issue is the possibility that fine-tuning risk classification could result in particularly high premiums for certain consumer types and even create consumer segments that are left without affordable cover, which has been acknowledged as a concern by Paul Evans, CEO of AXA UK and chairman of the Association of British Insurers (ABI).[92] Hypothetically, this might be a significant social concern if the consumer segments affected include vulnerable consumers, or if premiums could be influenced by individual characteristics (e.g. ethnicity or religion) in a way that is

[89] ABI, 2013, 'Identifying the Challenges of a Changing World', https://www.abi.org.uk/Insurance-and-savings/Topics-and-issues/~/media/0D97E1A140F84636BFE2A938C194EFDA.ashx

[90] Computing, August 2014, 'Insurers to demand more data via telematics to fine-tune insurance prices, says AXA CIO', http://www.computing.co.uk/ctg/news/2362319/insurers-to-demand-more-data-via-telematics-to-fine-tune-insurance-prices-says-axa-cio

[91] For example "*Some view insurance as a means of spreading risks throughout an entire population*". Swedloff, 2014, 'Risk Classification's Big Data (R)Evolution', Connecticut Insurance Law Journal, Vol. 21, http://ssrn.com/abstract=2566594

[92] "*The industry will have to take great care to ensure we're not creating, because of big data, sectors of society that can't buy insurance.*"
FT article, February 2015, 'Insurers warned to use 'big data' responsibly, http://www.ft.com/cms/s/0/08f6049c-a7cd-11e4-8e78-00144feab7de.html

deemed unfair or unlawful, but which may not easily be observable due to the complexity of algorithms used in risk classification.[93]

2.4.2 Information checks and fraud prevention

Examples were provided in Section 2.3.2 of various shared databases, such as the CUE database and the Cifas National Database, commonly accessed by insurance providers. By collecting data from these sources at the point of quote or point of sale, an insurance provider may be able to verify the individual's identity and to confirm some of the individual's information. The provider may also be able to establish whether any information links the individual or the vehicle to previous cases of fraudulent activity. Similarly, insurance providers can perform checks at the point of claim in order to verify information given by the policyholder and identify discrepancies or suspicious circumstances.

Optimising fraud prevention could entail substantial benefits for an insurance provider. As mentioned in Section 2.2.1, fraud remains a key issue in the sector. The ABI estimates that the general insurance industry invests £200 million per year in fraud detection and that fraud adds £50 to the average general insurance premium in the UK;[94] thus there is a clear commercial benefit from reducing fraud, while consumers also stand to benefit as long as some proportion of any savings is passed on through lower prices.

Various data-sharing arrangements exist

The use of such databases is typically mentioned in firms' policy documents, with varying degrees of detail provided. It is clear that a large number of fraud investigation groups and arrangements exist to allow insurance providers to collect and share relevant information. For example, one insurance provider lists 13 different fraud protection agencies and databases that it uses.[95] Different sources of fraud-related data may vary in terms of scope, type and potential usefulness of information held for the motor insurance sector. We note that the current data-sharing protocols in relation

[93] For a more detailed discussion, see Swedloff, 2014, 'Risk Classification's Big Data (R)Evolution', Connecticut Insurance Law Journal, Vol. 21, http://ssrn.com/abstract=2566594

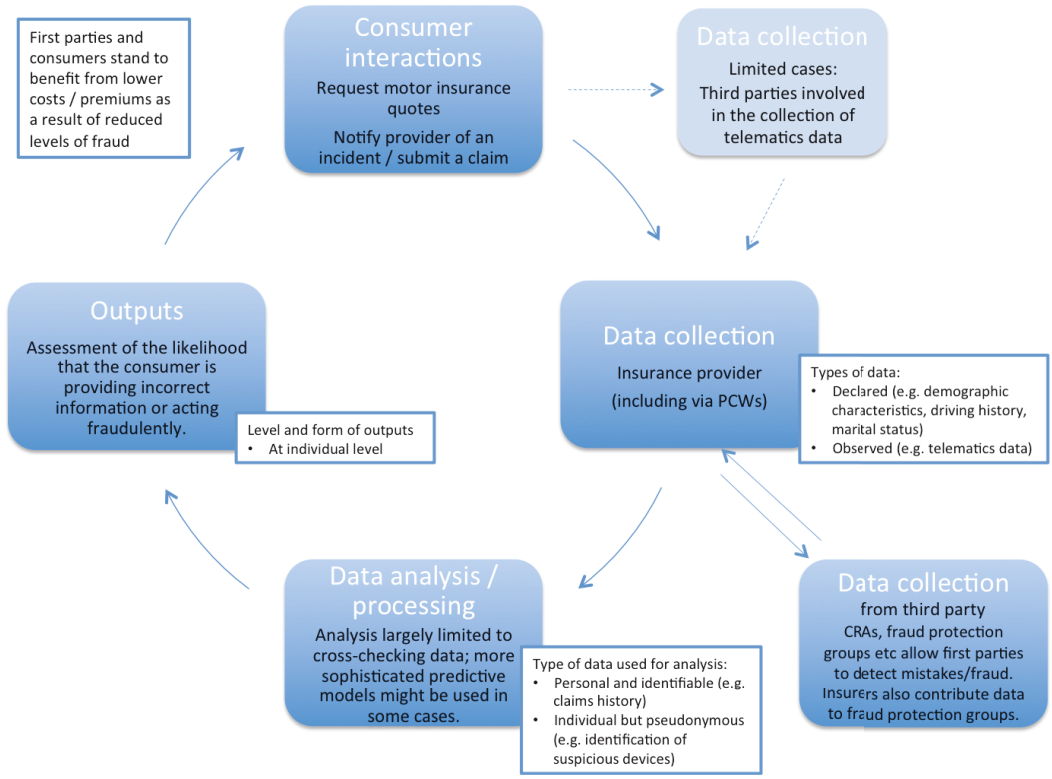
[94] See https://www.abi.org.uk/Insurance-and-savings/Topics-and-issues/Fraud

[95] https://www.coveainsurance.co.uk/about-covea-insurance/privacy.aspx
NB some fraud protection agencies are also credit reference agencies.

to fraud prevention have previously been scrutinised in light of possible data protection concerns.[96]

The diagram below summarises the role of data as part of identity verification and fraud detection activities.

Figure 6: Data flows associated with information checks and fraud prevention



Source: DotEcon and Analysys Mason

Innovations may help to fight fraud

Evidence from our interviews confirmed that innovations are taking place in the area of fraud prevention using consumer data. There are various relevant examples:

- the DVLA recently launched a service called MyLicence, which provides data to insurance providers in real time, based on the individual's driving licence number;
- the Insurance Fraud Bureau is reportedly exploring plans to act as a hub to standardise and centralise the sharing of data;[97]

[96] Pinsent Masons, August 2013, 'Insurers developing data-sharing protocols after industry body raises data protection concerns', http://www.out-law.com/en/articles/2013/august/insurers-developing-data-sharing-protocols-after-industry-body-raises-data-protection-concerns/

[97] Ibid.

- a database of no claims discount entitlement has been mooted[98] and is seemingly being developed,[99] though similar private sector solutions may already be available from third-party data suppliers;[100]
- social media data is being used to combat fraud by some insurance providers;
- telematics data can help to establish empirical evidence of the circumstances of an accident; and
- other innovative in-house and third-party fraud prevention techniques have emerged, including forms of predictive analytics or other so-called 'big data' techniques.

A more in-depth look at selected key developments in this area is provided below.

The DVLA's MyLicence service

MyLicence allows cross-checking of driver information and should deliver benefits to consumers

The DVLA and the Motor Insurer's Bureau jointly launched a new system in 2014 which allows individuals, insurance firms and car hire firms to access individual driving licence information online through the gov.uk website using an individual's licence number, national insurance number and postcode. Before this was launched, firms were unable to check licence details (e.g. type of licence, date issued, any driving convictions or points) at the point of quoting for policies, and so had to 'price in' the risk that the information disclosed by the licensee was incorrect, either due to mistakes or due to deliberate fraudulent behaviour.[101] Stakeholders told us that such behaviour is quite commonplace. By allowing checks to be made, including with regard to any motoring convictions, the system should allow fraud or misrepresentation to

[98] ABI, IFB, IIB, BIBA, Gocompare.com, Comparethemarket.com, Confused.com and Moneysupermarket.com, 'Helping to Reduce Insurance Fraud when Customers Apply for Products, A Good Practice Guide', https://www.reason-global.com/userfiles/file/fraud%20good%20practice%20guide%20from%20BIBA.pdf

[99] See for example Moneysupermarket.com, December 2014, 'Will MyLicence end 'proof of no claims discount' nightmare?', http://www.moneysupermarket.com/car-insurance/blog/will-mylicence-end-proof-of-no-claims-discount-nightmare/

[100] For example LexisNexis, see Actuarial Post, 'LexisNexis NCD module reaches 80% market engagement', http://www.actuarialpost.co.uk/article/lexisnexis-ncd-module-reaches-80---market-engagement-7229.htm

[101] The Guardian, January 2014, 'Car insurance premiums may fall as driving licence records go online', http://www.theguardian.com/uk-news/2014/jan/09/car-insurance-premiums-licence-records-online

be prevented. The ABI has estimated that premiums would be reduced by around £15 on average as a result.[102]

According to Gocompare, MyLicence should also facilitate the claims process and there are plans to eventually bring MyLicence together with another widely used database (Claims and Underwriting Exchange, 'CUE'), perhaps on an open portal that consumers could access as easily as insurers.[103]

Stakeholders agreed that this was a positive development, though at the time of writing it is understood that take-up of the service is still expanding and many providers have not yet gone live with MyLicence. We note that a recent market research study found that only 63% of consumers thought it was fair for insurers to access DVLA data.[104]

Use of the CUE database

Information provided by individuals about their claims and accident history is commonly verified by insurance providers against information held the Claims and Underwriting Exchange (CUE) database, which was established in 1994 and contains information about motor, home and personal injury incidents that may or may not give rise to a claim. The information obtained from the database is not disclosed to the consumer unless the consumer specifically requests it.

According to Gocompare, "*there has been some concern within the insurance industry that the Cue database is now being used for purposes that were never intended*".[105] Specifically, there has reportedly been a move from using the database only at point of claim to also using it at point of sale, or even at the point of quote, seemingly affecting premiums in some cases.

Some firms (both PCWs and insurance providers) have privacy policies that provide a clear account of the way the CUE database is used, stating that details may be entered when an incident is reported (even if a claim is not ultimately made) and that the database may be accessed to check an individual's record at the

[102] See https://www.abi.org.uk/Insurance-and-savings/Topics-and-issues/Insurance-industry-access-to-driver-data

[103] See http://www.gocompare.com/car-insurance/iiadd-database/

[104] Consumer Intelligence research, as reported in Your Wealth, 12 November 2014, 'Third of UK insurance customers unaware insurers use social media to check claims', http://www.yourwealth.co.uk/features/third-of-uk-insurance-customers-unaware-insurers-use-social-media-to-check-claims

[105] See http://www.gocompare.com/car-insurance/cue-database/

point of quote. Other privacy policies mention the database without providing details of usage.

Alleged point-of-quote usage has caused controversy

There has been some media coverage focusing on the fact that incidents may be logged in the database even when a claim is not ultimately made, and that allegedly insurers may take these past incidents into account when setting premiums. Reports suggest that consumers are largely unaware of the way the database may be used, with anecdotes suggesting that customers have been penalised for speaking to their insurance provider about minor incidents, even without making a claim. For example, the BBC reported that "*[m]otorists who call their insurer only to inquire about making or reporting a minor claim can find the information is recorded and used against them when it comes to renewing cover*".[106] Similarly, Gocompare claims that "*a reported minor 'incident' could result in a higher premium or quote – not just from your current insurer, but from all those using the database*".[107]

Evidence from our interviews was inconclusive regarding the extent to which this is widespread practice.

Use of social media information

Social media data is increasingly used to fight fraud

Insurance providers are known to have increasingly considered the use of social media data to help detect and prevent fraud. For example, in 2013 Direct Line Group completed a claims transformation programme in its motor and home insurance business, "*including the use of social networking techniques to combat fraud*".[108] Tesco's car insurance policy states: "*We may research, collect and use data about you from publicly available sources including social media and networking sites. We may use this data for the purposes of fraud detection and prevention*".[109]

[106] See for example BBC, August 2013, 'Why reporting insurance 'incidents' can cost you dearly', http://www.bbc.co.uk/news/business-23903966, This is Money, October 2013, 'I informed my insurer about an accident without making a claim but it still raised my premium', http://www.thisismoney.co.uk/money/cars/article-2467812/Honest-drivers-hit-insurance-penalties.html, and The Guardian, September 2013, 'Insurance: how a simple query could cost you a premium penalty', http://www.theguardian.com/money/2013/sep/30/insurance-query-higher-premiums

[107] See http://www.gocompare.com/car-insurance/cue-database/

[108] Direct Line Group, 2013, Annual Report & Accounts 2013, http://ara2013.directlinegroup.com/downloads/pdf/direct_line_group_2013_annual_report_and_accounts.pdf

[109] See https://www.tescobank.com/assets/sections/carins/pdf/data-processing-information.pdf

One insurer provided anecdotal evidence of its success in this area, illustrating how social media may be used in practice. In relation to a claim that involved several passengers in an alleged collision, together claiming a total sum in excess of £100,000, the insurer was able to use information from Facebook to establish that the drivers of the two vehicles, as well as some of the passengers, were known to each other. Ultimately this helped to prove that the accident was staged. Other possibilities are that social media might be used to confirm suspicions about fronting[110] (where a high-risk driver, e.g. a young driver, is in reality the main driver of the vehicle but is only stated as a named driver in order to reduce the premium), or that information on social media might contradict a claimant's account of an accident (e.g. where photographic or video evidence is found).[111]

One insurer confirmed to us that it has sourced information on social media in order to prevent fraud, but that it only considers this option in selected cases where there is a clear suspicion of fraud, recognising that these methods could be considered intrusive in the absence of this condition. Third parties may also offer social media data-related services to assist insurance providers in fighting fraud.[112]

One insurer informed us that it had made a conscious decision – in consultation with the ICO – to not divulge substantial details about its approach to fraud detection, for the very reason that doing so could make it easier for fraudsters to escape detection.

Use of telematics data

There may be benefits from the use of telematics data in relation to fraud detection and prevention, in the event that telematics policies become mainstream. The detailed driving data collected in the event of an accident may help to corroborate (or contradict) the claimant's account of events, such as in relation to whiplash claim fraud.[113] Since a black box acts as a vehicle tracker, it should deter

[110] AXA Connect, 18 January 2012, 'Fraud and Social Media', http://www.axaconnect.co.uk/fraud-and-social-media/

[111] See for example CBS, February 2015, 'Investigators combing social media to expose insurance scams', http://www.cbsnews.com/news/investigators-combing-social-media-to-expose-insurance-scams/

[112] See for example https://www.smdiscover.com/insurance-claims-adjustment-by-referencing-social-media-content/

[113] See for example Business Technology, 2 April 2014, 'Big data: helping insurance firms target serial fraud', http://business-technology.co.uk/2014/04/big-data-helping-insurance-firms-target-serial-fraud/

certain types of fraud in particular, such as fraudulent theft claims. It has been reported that this aspect of telematics drove adoption of telematics policies in Italy.[114] In the UK, telematics insurance policy documents typically allow for the use of telematics data to detect fraud.

Predictive analytics and other modern techniques

The advent of 'big data' and increasing use of predictive analytics might have further effects on the way consumer data is used in the fight against fraud. Research by Accenture suggests that most insurers in Europe and Latin America use, or plan to use, fraud modelling techniques to detect fraud.[115] Predictive models may be built using an insurance provider's historical data, identifying common patterns in known fraudulent cases. These insights can be used to predict the likelihood of fraud in future claims received and, through an automated process, to identify suspicious claims that warrant in-depth investigation early in the claims process.

Third parties may assist firms in developing such techniques. As an example, IBM claims that its technology solution helps a major US insurer "*ingest IP addresses, scanned document references, email addresses, driver's license information, criminal records, judgments or known fraudulent connection data to link multiple sources and identify relationships that may otherwise remain anonymous*".[116] However, evidence from stakeholder interviews does not suggest that the use of similar third-party technologies is particularly widespread at present in the UK.

The prevalence of the online sales channel also creates potential new challenges for identity verification and fraud detection; new techniques may emerge to deal with this problem. For example, Confused.com makes use of a third-party service that reportedly allows the identification of suspicious devices used to access the

[114] The Economist, February 2013, 'How's my driving?', http://www.economist.com/news/finance-and-economics/21572237-gizmos-track-driving-habits-are-changing-face-car-insurance-hows-my

[115] Accenture, 2013, 'How to effectively fight insurance fraud', Figure 1, http://www.accenture.com/SiteCollectionDocuments/PDF/Accenture-How-Effectively-Fight-Insurance-Fraud.pdf

[116] See http://public.dhe.ibm.com/common/ssi/ecm/bi/en/bic03034usen/BIC03034USEN.PDF

website, e.g. pertaining to 'ghost brokers'[117] or organised fraud networks, without relying on the use of cookies.[118]

2.4.3 Use of data for marketing purposes

Marketing uses of data in the sector are fairly 'standard'

As in most consumer-facing sectors, firms in the motor insurance sector may have incentives to leverage the available consumer data for marketing purposes, such as personalising services, targeting campaigns at particular segments, or targeting specific offers or advertisements at individual consumers, including offers to cross-sell different types of insurance or up-sell more expensive cover options.

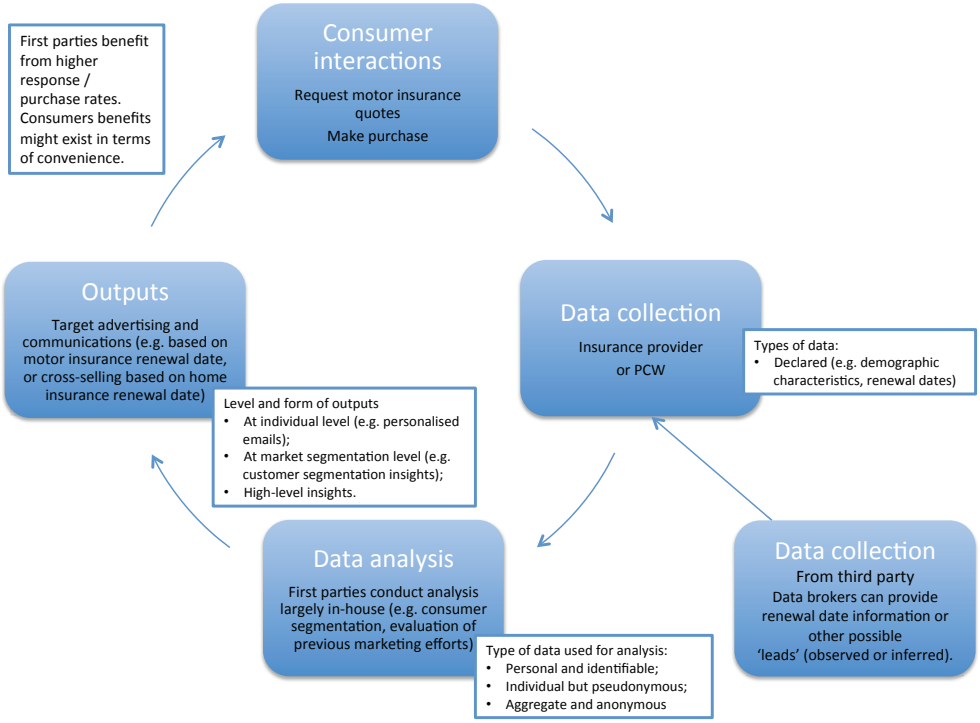
Our interviews have indicated that, relative to other sectors, the motor insurance sector employs relatively rudimentary tools and analytical techniques in this area. Nevertheless, evidence suggests that, as in the areas of risk evaluation and fraud management, there is a willingness of firms to increase and enhance the use of consumer data in a marketing context. For example, one insurer told us that it expects analysis of online behaviour to become an increasingly important area, even though at present it is fairly simplistic.

The diagram below summarises the role of data as part of marketing-related activities.

[117] These are illegal insurance advisers selling fraudulent policies to consumers. The policies may be genuine, but deliberately based on false information that invalidates them, or they may be entirely fake policies that are presented to the consumer as genuine. See https://www.abi.org.uk/Insurance-and-savings/Topics-and-issues/Fraud/Fraudulent-insurance-policies

[118] Experian, December 2014, 'Confused.com first UK aggregator to prevent insurance fraud at the point of quote', http://www.experian.co.uk/blogs/latest-thinking/confused-com-prevents-insurance-fraud-at-quote/

Figure 7: Data flows associated with marketing in the motor insurance sector



Source: DotEcon and Analysys Mason

Targeted marketing

It is commonplace for insurance providers and PCWs to use targeted marketing based on renewal date information. For example, where consumers obtain a motor insurance quote but do not make a purchase, they may be contacted with a reminder ahead of their stated renewal (or policy start) date. Where renewal date information is available for other types of insurance (primarily home insurance), direct marketing can then be used to attempt to cross-sell products to the consumer, during the window of time approaching the renewal date.

One insurer told us that it may also consider targeting its online advertising on the basis of geographic regions, based on its insight about its price-competitiveness in different areas.

Other forms of personalised marketing exist, for example based on in-house consumer segmentation analysis, though our interviews indicated that the techniques used may be less sophisticated than in some other consumer retail sectors.

Interviewees told us that quantifying the value of targeted marketing was challenging, but nevertheless were able to provide some illustrative metrics. One PCW estimated that personalised (as opposed to generic) communications could result in purchase rates being around 10% higher, while an insurer estimated that response

rates for personalised communications could be double those for standard communications.

Role of third-party firms

Third-party firms may be able to assist firms in the motor insurance sector by providing data that helps to target marketing strategies.

Third parties can provide renewal date information

A clear example is the provision of leads based on renewal date information. While insurance providers or PCWs will typically hold this information for consumers who have made a purchase or obtained a quote, they generally will not have this for other consumers. This could be particularly relevant, for example, for diversified firms (e.g. supermarkets with an insurance spin-off); such firms may hold personal details for a large customer base, only a fraction of which may be motor insurance policyholders with the firm. Such firms can purchase renewal date information from third-party data brokers (e.g. Acxiom)[119] in order to be able to offer motor insurance to the customer during the most appropriate window of time. Even when firms do not hold contact details, they may still be able to target advertisements at those consumers believed to be approaching their renewal date, as third parties can offer such a service. For example, Acxiom allows firms to target advertising on Facebook based on the consumer data held by Acxiom.[120]

Third parties may also be able to provide leads based on yet more sophisticated data collection and analysis. For example, some data brokers effectively scour and index information systematically across social media platforms, blogs and other websites. They may then be able to identify individuals that are posting relevant material or having relevant conversations (e.g. asking where to find cheap car insurance). The third parties would then allow insurance providers or PCWs to target advertising at such individuals, or they may even be able to initiate direct contact with the individual in certain circumstances.

Firms in the motor insurance sector also engage third parties for the purpose of administering and implementing marketing campaigns. This includes the provision of email addresses or postal addresses

[119] Acxiom holds information on individuals' insurance renewal month. See https://www.myacxiompartner.com/data-selection.pdf. Vehicle insurance renewal information also appears in the Federal Trade Commission's illustrative list of data elements and segments held by data brokers, along with other individual and vehicle information that is potentially relevant. See FTC, 'Data Brokers – A call for Transparency and Accountability', 2014, Appendix B.

[120] See http://dq2qu0j6xxb34.cloudfront.net/wp-content/uploads/2014/01/Facebook-Case-Study-Final-no-bleed.pdf

for the purpose of processing and sending a large volume of communications, for which the firms may not have the required capabilities in-house.

Finally, third parties will typically be engaged for the purposes of carrying out consumer or market research, as in many other commercial sectors.

Opt-out controls for marketing use of data

Firms told us that consumers are always given the option to opt out from having their personal information used for marketing purposes in the future. This is typically possible at the point of obtaining a quote (online or in a follow-up email presenting the same quotes) and also thereafter (for example, through the customer's online profile on a provider or PCW's website). The use of soft opt-ins is common.

One PCW told us that, though insurance providers may be permitted to use the data to contact the consumer directly in specific circumstances, insurers' conduct is monitored on an on-going basis (through the use of dummy accounts) in order to ensure that contact attempts are only made when the relevant circumstances apply and that the number of attempts remains within strict limits. In any case, an opt-out control is available for consumers to avoid being contacted by the providers directly.

The FCA's thematic review of PCWs highlights some potential concerns over opt-out controls offered to consumers.

Box 8: FCA research on PCWs

The FCA conducted a thematic review[121] of PCWs in the general insurance sector in 2014, accompanied by externally conducted consumer market research.[122] It identified aspects of PCW behaviour in relation to consumer data that could be of concern.

The FCA found that many consumers did not feel that PCWs were transparent about data use and "*[s]ome PCWs did not clearly explain how they would use consumers' data or have secure access controls over it*". Despite the perceived lack of transparency, PCW users generally do not read terms and conditions. The research found that accepting terms and conditions on both PCWs and insurance provider websites, without reading them, was "*something of an accepted social norm*".

Consumers could also be confused by the opt-out controls provided to them for marketing uses of data:

"*A number of PCWs had tick boxes for agreeing to the T&Cs and privacy of data policy, without providing the option for the customer to opt in/out for the use of their personal data for marketing purposes. In such a case, the consumer may not know how to opt out of being contacted for marketing, as the process for doing so was difficult to find and/or was unclear or misleading.*

Further, the framing of certain questions which offered the customer the option of opting in or out could be confusing. For example, some PCWs required customers to opt out for two different types of contact before providing the quote results. The first question related to whether they would like to be contacted by the two cheapest providers and the second question was whether they would like to be contacted for marketing purposes. To opt out, the customer would have to un-tick a box for the first question but tick one for the second.

Some participants in our consumer research thought they had opted out for contact for marketing purposes when they had actually opted in."

2.4.4 Supply of data to third parties

As discussed, insurance providers will routinely share information through a number of fraud protection initiatives that facilitate data sharing between providers. In addition to this, they will enter individuals' details into the Motor Insurance Database, as well as releasing information to authorities such as the police where they are required to do so, e.g. as part of a criminal investigation. As part of data enrichment activities, they may also share information with

[121] FCA, 2014, 'Price comparison websites in the general insurance sector', TR14/11, https://www.fca.org.uk/your-fca/documents/thematic-reviews/tr14-11

[122] Atticus (for the FCA), 2014, 'Price comparison website: consumer research', https://www.fca.org.uk/static/documents/research/price-comparison-website-consumer-research.pdf

third parties in order to be provided with additional information about the individual, or in order for the third party to conduct data analysis (e.g. on telematics data).

PCWs will share the bulk of the data they collect with insurance providers for the purposes of obtaining quotes, while they may also consider sharing data for additional purposes, as described below.

Beyond the types of arrangements mentioned above, insurance providers and PCWs told us that they generally do not share consumer data with one another.[123]

Personal information and referral fees

Referral fees for personal injury claims caused controversy

The personal information held by an insurance provider may be valuable to third parties. The clearest example of this is during the aftermath of an accident. In the words of one commentator, "*[a]ny claimant on a motor insurance policy is a valuable commodity to a whole range of firms, keen to offer legal assistance, vehicle hire or repair*".[124] In particular, the sale of information in relation to personal injury claims has been an area of controversy in the sector in the past, eventually being outlawed, though reported levels of nuisance telephone calls related to insurance claims remain high.

[123] Some data-sharing agreements between insurance providers may be covered by the Insurance Block Exemption Regulation (IBER), which exempts agreements related to joint compilations, joint tables and studies from competition law under certain circumstances (in summary, where aggregated and anonymised statistical information is exchanged, being made available on reasonable, affordable and non-discriminatory terms, for the purpose of risk calculations). We have not encountered evidence that the IBER is having a material impact on data-related activities in the UK motor insurance sector in practice. However, we note that the IBER may have facilitated the development of the telematics segment in Italy, as "*the diffusion of data collected by ANIA [the Italian insurance association] helped a number of insurance companies to penetrate this new segment and to propose these kinds of innovative offers to consumers*". (http://ec.europa.eu/competition/consultations/2014_iber_review/insurance_europe_en.pdf)

[124] Norton Rose Fulbright, June 2012, 'The regulation of the motor insurance industry', http://www.nortonrosefulbright.com/knowledge/publications/63780/the-regulation-of-the-motor-insurance-industry

Box 9: The sale of personal details to third parties in connection with personal injury claims

The subject received widespread media coverage in 2011, when former MP Jack Straw brought attention to the practice of selling personal details of individuals who had recently been involved in accidents, alleging that some insurance providers would sell these details to claims management companies and personal injury lawyers offering 'no win, no fee' services.[125] The data might then be used for aggressive marketing campaigns. In the same year, a Transport Committee report found that personal injury lawyers commonly paid referral fees to insurance providers, as well as to other firms that had access to claimants' personal details.[126] The Committee's view was that consumers were largely unaware of their details being used and monetised in this way, so it called for a more transparent regime.

The payment and receipt of referral fees in connection to personal injury cases was eventually banned from 1 April 2013.[127] However, a follow-up report by the Transport Committee found that "*[a]lthough referral fees have now been outlawed, links between insurers and such firms still exist and there are new legal mechanisms for bringing insurance firms and solicitors together under one roof*".[128] In the Committee's view, its previous call for transparency had not been answered.

Recent figures from the ICO show that consumers reported several thousand concerns about accident-related nuisance calls during the quarter October–December 2014.[129] A representative of the ICO has commented that there still seems to be a lucrative trade in the personal details of individuals who have been involved in accidents, though the ICO had not seen any evidence of insurance providers deliberately selling leads in violation of the ban.[130]

[125] The Times, June 2011, ''Dirty secret' that drives up motor insurance', http://www.thetimes.co.uk/tto/opinion/columnists/article3075361.ece

[126] Transport Committee, September 2011, 'The cost of motor insurance', §24, http://www.publications.parliament.uk/pa/cm201012/cmselect/cmtran/1466/146602.htm
This practice was acknowledged by parties such as the ABI – see e.g. ABI, September 2011, 'Tackling the compensation culture: the Legal Aid, Sentencing and Punishment of Offenders Bill', p. 12, https://www.abi.org.uk/Insurance-and-savings/Topics-and-issues/~/media/A1DCD3A1AE944458BC8ACBEAF3FB95DA.ashx

[127] As part of the Legal Aid, Sentencing and Punishment of Offenders Act, see http://www.justice.gov.uk/civil-justice-reforms/personal-injury-claims

[128] Transport Committee, 15 July 2013, 'Cost of motor insurance: whiplash', §66, http://www.publications.parliament.uk/pa/cm201314/cmselect/cmtran/117/117.pdf

[129] See https://ico.org.uk/action-weve-taken/nuisance-calls/

[130] BBC, January 2015, 'Privacy regulator's warning over accident claim calls', http://www.bbc.co.uk/news/business-30642368

Referral fees still exist in other contexts

Aside from the issue of personal injury claims, insurance providers are still able to receive referral fees from other firms involved in the claims management process, such as vehicle repairers, paint or part suppliers, and credit hire companies (which provide temporary replacement vehicles). Insurance providers told us that policyholders would typically be informed that their information is being passed on for this purpose and that they would retain the option of refusing this and seeking their own repairer, outside of the provider's approved network of suppliers.

The CMA considered the effect of referral fee payments in its PMI market investigation. The CMA found that this mechanism induced credit hire providers and other third parties to compete over the level of referral fee offered, creating a substantial revenue stream for insurance providers. The average referral fee paid for credit hire was £328 and the equivalent fee for credit repair (or write-off) was £53. The CMA's assessment was that this revenue stream was likely to reduce premiums, though referral fees may not be passed fully onto consumers.[131] Referral fees were not considered to be a source of adverse effects of competition in the PMI market.[132]

Anonymised or aggregated information

There is little concrete evidence to suggest that the sale of data from firms in the motor insurance sector to third parties is widespread practice at present.

Point-of-quote data may be valuable to third parties

Point-of-quote data volunteered by consumers may have some commercial value. Indeed, in 2014, the FCA's thematic review of PCWs found that PCWs "*may also earn income from providing data intelligence services or by selling consumers' data to third parties*"[133], though it is not specified what type of data this might include (personal, anonymised or aggregated). The privacy policies adopted by PCWs differ on this issue, in terms of the restrictions that apply to the provision of consumer data to be passed on to third parties for purposes that are not related to the policy. The FCA also noted that some PCW policy statements allowed the possibility of sharing data with third parties, even if such sharing was not actually taking place at the time of the FCA's review.

[131] CMA, 24 September 2014, 'Private motor insurance market investigation Final report', §8.5

[132] Ibid., §10.146

[133] FCA, 2014, 'Price comparison websites in the general insurance sector', TR14/11, https://www.fca.org.uk/your-fca/documents/thematic-reviews/tr14-11

At least one first party firm, MoneySuperMarket, has publicly acknowledged the value that might be generated from selling data to third parties, as outlined below.

Box 10: MoneySuperMarket's plans to sell data

MoneySuperMarket's privacy policy appears relatively permissive in that it states:

"*As part of using our Services you consent to us disclosing your personal information to the following parties:*

a. Third parties, including but not limited to companies whose products or services are included on our website with a view to providing you with an online quotation for the product and/or service requested by you. […] Some third parties with which we share your information may use it to carry out research such as analysis of market trends and customer demographics and to customise and develop the product/service which they offer to you or other individuals in the future."[134]

Indeed, MoneySuperMarket has stated its intention "*to develop a new source of revenue worth tens of millions of pounds by selling data about its 21m customers*".[135] However, the company has been keen to emphasise that it will offer aggregated trend data, rather than data about individual customers.

Firms may be sensitive to any competition implications of data sharing

In our interviews, one PCW told us that it has considered the commercial dissemination of data to third parties, but it has not pursued this to date. It recognised that the vast volumes of data that it collects and processes on a daily basis render the data very up-to-date and therefore valuable to third parties, such as insurance providers. However, one important consideration in this respect is the precedent set by the OFT's Whatif? ruling in 2011, which set strict limits on the type of data that could be shared when the data was relatively current (less than six months old) and which might restrict a PCW's ability to sell valuable data. An overview of the Whatif? investigation is provided below.

[134] See MoneySuperMarket's privacy policy, available at: http://www.moneysupermarket.com/legal/privacypolicy.asp

[135] FT, March 2014, 'Moneysupermarket to sell data on customers', http://www.ft.com/cms/s/0/dc169c48-a381-11e3-88b0-00144feab7de.html

Box 11: OFT investigation of 'Whatif?' information exchange tool

In January 2010, the OFT launched a formal investigation into certain arrangements between several insurance companies, suspecting that those companies were indirectly exchanging commercially sensitive information through third parties, potentially affecting competition.

The focus was on an information exchange product known as Whatif?, sold by Experian and widely used throughout the industry. Whatif? was "*an information product consisting of an information exchange which enables insurers to access other insurers' pricing information for any risk profile*". Detailed, disaggregated information was exchanged frequently and could allow one insurer to anticipate its rivals' future prices three weeks in advance, so it was deemed conducive to co-ordination.

In 2011, the OFT accepted commitments from seven large insurers and two IT providers, based on the principle that any pricing information exchanged through Whatif? should, if less than six months old, be anonymous (i.e. not identify the insurer by which it is provided), averaged across at least five insurers, and already live within broker sold policies.

The OFT did acknowledge that other similar market analysis tools existed in the motor insurance market. It expected stakeholders to give due consideration to such arrangements and ensure compliance with competition law.

Another example of PCW data being disseminated to third parties is Confused.com's car insurance price index,[136] in collaboration with specialist risk management firm Towers Watson. Confused.com supplies extensive data on insurance quotes to Towers Watson, though the data is anonymised both with regard to the individuals and the insurance providers involved. Towers Watson process the data, allowing aggregate price trends to be quantified and also broken down by demographic and other characteristics. It appears that this information is distributed freely rather than being available commercially, but it may still create significant value to the company, for example by driving traffic to the website.

Consumers may object to data sharing, even when anonymised

In general, where data is provided to third parties, privacy concerns may arise not only when personal information is sold, but also where the data is aggregated. An example of this possibility is the reported sale of data collected from satellite navigation devices by TomTom. The example, outlined below, is not directly related to the motor insurance sector, but it is relevant in that the data collected by TomTom is comparable to GPS data collected by telematics devices or apps.

[136] For details, see http://www.confused.com/car-insurance/price-index/about

Box 12: TomTom's sale of anonymised data to the police

An example of public outcry in response to the sale of consumer data concerns satellite navigation company TomTom. The company made controversial decision to sell anonymised driving data collected from users of its devices to the police in the Netherlands.[137] The company issued an apology to its customers after it was revealed that the data had been used to help set speed traps. The example highlights that, even when data cannot be traced to individuals, there may be strong public opposition if the data is used for purposes that consumers deem to be non-essential and potentially detrimental to their own interests, and that consumers providing the data may not be aware of its possible uses. While TomTom is not directly involved in the motor insurance sector, the example is of some relevance, since data collected by telematics devices, for example, could seemingly be used for a similar purpose in theory.

Telematics data may be valuable to third parties

Though there is no concrete evidence of third-party uses of telematics data from desk research, the ABI has recognised that telematics data will be an attractive resource for third parties, including for marketing and research purposes.[138] In its guidance, the ABI recommends that as a matter of good practice, consumers should be required to actively opt in for having their data shared with any non-essential third parties. According to recent reports, the German Auto Industry Association is asserting its view "*that the [telematics] data first and foremost belongs to the driver and should not be sold on to advertisers and other third parties – to be adopted by both the EU and United Nations*".[139]

Terms and conditions regarding third-party use vary

The terms and conditions used by insurance providers differ. In general, there are strict restrictions on the provision of the data to any third parties, other than any third-party analytics supplier. Nevertheless, the restrictions might theoretically still allow certain uses of the data that monetise its value in ways that consumers could object to, similar to the TomTom case (see above).

[137] FT, April 2011, 'TomTom sorry for selling driver data to police', http://www.ft.com/cms/s/2/3f80e432-7199-11e0-9b7a-00144feabdc0.html

[138] ABI, 2013, 'Selling Telematics Motor Insurance Policies, A Good Practice Guide', p.12, https://www.abi.org.uk/~/media/Files/Documents/Publications/Public/Migrated/Telematics/Selling%20telematics%20motor%20insurance%20policies%20-%20ABI%20good%20practice%20guide.ashx

[139] Pinsent Masons, January 2015, 'Telematics insurance, market disruption and control of data', http://www.out-law.com/en/articles/2015/january/telematics-insurance-market-disruption-and-control-of-data-/

Box 13: Restrictions on third-party use of telematics data

According to Direct Line's terms and conditions for its DrivePlus telematics policy, the third-party analytics provider may use the data for purposes that include analysis of road safety issues. The permitted purposes are:

- *"Road and vehicle usage including for road safety issues, real time traffic flow, environmental impacts such as idle time at junctions, journey times, distances and speeds, and the analysis of junctions and the risk they represent;*
- *Driving behaviour analysis and profiling including determining what constitutes safe and dangerous driving and the typical behaviours of various segments of the UK population;*
- *Analysis of the causes of, and forces involved in, collisions and other road incidents; and,*
- *Researching and refining techniques for analysing motor vehicle telematics data."*[140]

According to Admiral's privacy and security statement in relation to its LittleBox telematics policy, the third-party technology company may use the data for policy-related purposes, but also for broader purposes: "*[g]eneral research and analysis, mapping purposes, researching and refining techniques for analysing motor Telematics data and the supply of traffic information. In all such circumstances the information will be used anonymously and will not identify any individual, vehicle user, or the policyholder*".[141]

Other policies appear somewhat more restrictive. Zenith's BrightBox telematics policy restricts the uses of telematics data to policy and policyholder-related purposes and to helping the insurer gain a better understanding of driving behaviours. It also provides the following clarification that might allay consumer concerns about third-party use:

"*Please note that whilst the information collected on driving speed will be used in determining the premium to be paid under the policy and to identify unacceptable driving behaviour as defined in the telematics conditions of this policy it will not be used to support a speeding prosecution in any way. The insurer or its service providers may however be required by law to disclose information about your driving behaviour to the authorities, for example in answer to any enquiry by our regulatory body or to a court of law if we are issued with a court order.*"[142]

[140] See http://faqs.directline.com/terms/disclaimers/black-box-terms-conditions

[141] See http://www.admiral.com/car-insurance/your-policy/your-privacy-and-security.php

[142] See http://documents.markerstudygroup.com/media/1011/ZENBBT1114-Zenith-Brightbox-policy-booklet-201411.pdf

2.4.5 Additional uses of telematics data

The use of telematics data in the context of risk evaluation and fraud prevention has been discussed, but there are further possible uses that insurance providers may make of this data, which could be desirable to the customer as value-added services. Accenture predicts that "*insurers will need to offer additional sustainable value to incentivise the mass market* ".[143]

Currently, some available value-added services include:

- accident response, whereby the policyholder and/or emergency services are contacted in the event of an accident being detected;
- streamlined claims process, e.g. where it is the insurance provider that contacts the policyholder to collect details about an accident, rather than the other way round;
- theft recovery, where the location of the car is tracked in liaison with the police; and
- the ability for the policymaker to locate the vehicle using a smartphone app.

[143] Accenture, 2014, 'The Digital Insurer: Insurance Telematics', http://www.accenture.com/Microsites/insights/Documents/pdfs/Accenture-Telematics-Low-Res-Version-Final.pdf

3 The clothing retail sector

3.1 Summary

Use of data and potential value

Retailers, including clothing retailers, have for years collected information about their customers through loyalty programmes, surveys, catalogue subscription services and conversations with in-store sales assistants. However, in an increasingly digital age, the retail industry is able to benefit from a shift to online sales channels that allow retailers to collect more, and more systematic, data about their customers.

Customer data collected both online and offline are becoming increasingly important to retailers for improving services offered to their customers, increasing sales and reducing costs. This data allows retailers to:

- increase the conversion of consumer interest into sales and increased average order values by providing personalised search results, product recommendations and targeted advertising. By providing such services to consumers, search costs are reduced and if recommendations are accurate, customers will return to buy more;
- reduce the number of returns. The average return rate in the UK for online clothing is approximately 25%.[144] There are strong incentives to reduce returns, which are costly for the retailer (due to both delivery costs and, where retailers offer free returns, return postage which appears to be standard practice in the UK); and
- inform business strategy by understanding more about its consumers and about the performance of particular items or brands. This data can be used to help manage stock levels, inform purchasing decisions and even determine space allocation in-store where retailers have physical presence.

How is data collected?

Retailers collect data about their customers in three main ways:

- directly (for example personal demographic data provided at the point of registration);
- observed data through use of cookies or trackers that may be deployed on the retailer's website; and

[144] See Fits.me, 'Whitepaper - Garment Returns and Apparel Ecommerce: Avoidable Causes, The Impact On Business And The Simple Solution', http://communication.fits.me/acton/attachment/8391/f-0012/1/-/-/-/-/Whitepaper_A4_garment_returns_UK.pdf

- inferred data by, for example, reviewing past purchase history and making inferences about consumer brand or style preferences.

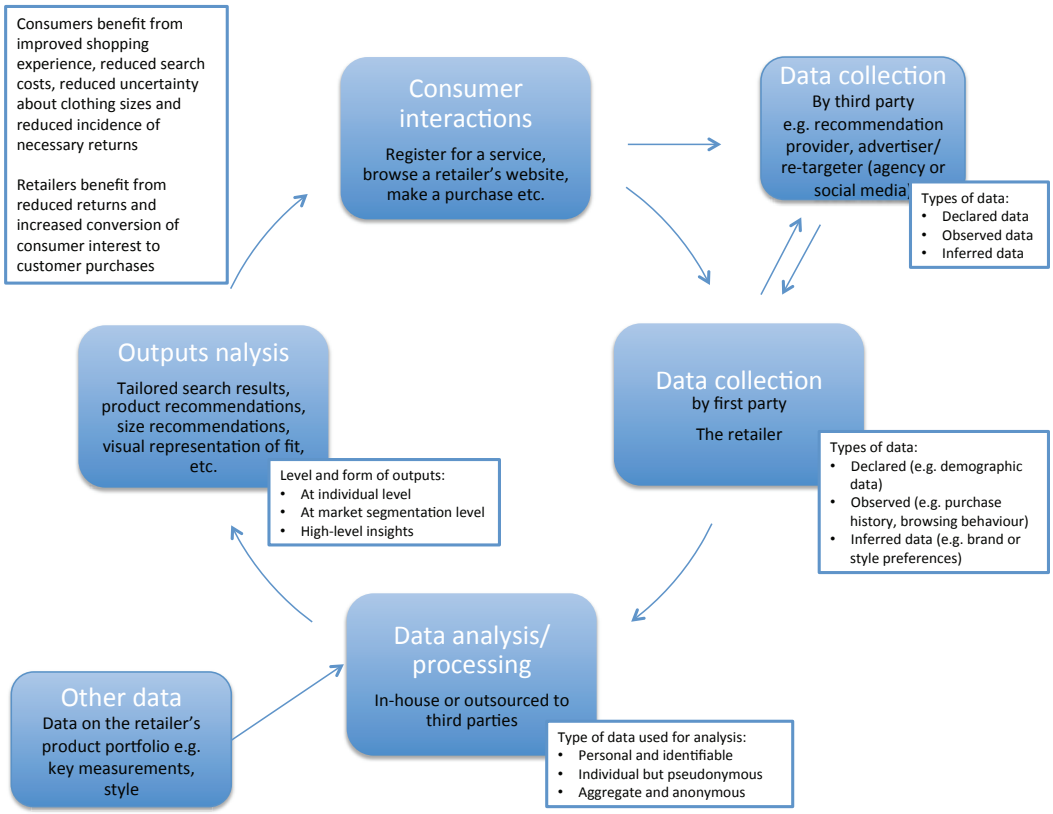
In some cases, data is collected by third parties who are providing a service that runs on the retailer's website, for example for the provision of product recommendations, or size and style recommendations. These services may collect observed data about the customer's behaviour through cookies or trackers. However, in some cases they collect information directly from the customer (such as key measurements or body shape) and can also collect data from the retailer including previous transaction history (depending on the specific agreements in place).

In each of these cases, the data collected may be supplemented by additional data, for example about the retailer's products, so the data collected about the customer can be matched with relevant items in the retailer's product portfolio.

Further, where targeted or re-targeted advertising is used, third-party advertising agencies – including social media sites that serve adverts (such as Facebook) – may also collect some observed data about users by tracking their behaviour online. This allows them to determine relevant adverts to show the individual (for example based on items viewed on a retailer's website).

The processes by which the clothing retail sector collects and uses data are summarised by the diagram below:

Figure 8: Data flows involved in the collection and use of consumer data in the clothing sector



Source: DotEcon and Analysys Mason

Recent developments and barriers to effective use of data

An increasing number of specialist service providers have entered the market to provide support to retailers to improve the customer experience and provide further ways of driving conversions, average order values and decreasing the number of returns. These services include:

- third-party personalisation/recommendation engines that are embedded in the retailer's website. These are often used where the retailer does not have sufficient in-house expertise or capabilities to provide its own recommendation engine. These services primarily aim to help increase conversions;
- visualisation tools allowing the customer to see items of clothing on a virtual model of similar size and shape, with the aim of driving conversions and reducing returns by providing customers with a better idea of what the garment will look like;
- numerical based size recommendation tools that use measurements provided by the customer to provide recommendations on size and a 'fit score' aimed at reducing fit related returns; and
- size and style recommendations based on information provided by the user about body shape and user

preferences for style and brands which help to increase conversions and reduce returns.

Evidence from desk research and interviews suggest that where these services are being used, they are successful in driving conversions, increasing order value and reducing return rates, showing that there are commercial benefits for retailers. Whilst these new techniques would appear to have wide applicability across the sector, we have found that at present such services are available on only a sub-set of retailers' websites, and even where offered not all customers are using them. Therefore, it seems as though the potential value of these services may not yet have been realised.

However, there are also some potential barriers to the success of these services, particularly those that are reliant on exact measurements and that are focused on reducing returns:

- requiring a large number of measurements from users will increase accuracy but may discourage people from using the service, or may lead to inaccuracies due to difficulties faced by consumers when measuring themselves;
- information from retailers and manufacturers about the products including sizing and measurement data are often poor, and even where measurements 'off plan' are available, imperfections in the manufacturing process may mean that such data is not entirely accurate;[145]
- in the UK, many retailers offer free carriage for returns, meaning that there is still an incentive for consumers to 'over-order' and just return the items that they do not want or that do not fit.

3.2 Sector overview

In this section we present a review of the clothing retail sector[146] in the UK, including an overview of the scale of the sector, the firms involved and the competitive dynamics along with any recent trends.

[145] For example, one provider of virtual fitting room and visualisation services told us that today the quality of data, including on garments is often inadequate because of the tolerances and variability in the manufacturing process.

[146] For the purposes of this research we have focused on the retail sector only, thus excluding wholesalers and manufacturers.

3.2.1 Scale of the sector

Expenditure in the UK clothing and footwear sector was over £50 billion in 2014.[147] Purchases in this sector are increasingly being made online; according to the Office for National Statistics (ONS) clothing is currently the most popular good or service bought online in the UK and in 2014, almost half of adults bought clothing online.[148]

Online fashion sales of £10.7 billion in 2014 and expected to rise to £19 billion by 2019

According to research by Mintel, sales of fashion online accounted for approximately 17% of total spending on clothing and footwear in 2014, up from 13% in 2011. Having recently seen a rapid growth in year on year online fashion sales, the annual growth rate has slowed slightly, yet sales are still strong and is forecasted to reach £19 billion by 2019.[149]

The importance of online shopping is particularly great for younger adults, with 9 out of 10 adults aged between 25 to 34 buying online and almost two thirds of those buying clothes online.[150]

Increasing use of tablets and smartphones supporting growth of the industry

With the increasing availability and use of smartphone and tablets, more people are also using the internet 'on the go' to make online purchases. For example, there was an increased uptake in the use of portable computers (laptop or tablet) from 32% of all adults in 2013 to 43% in 2014 with 16-25 year old being the biggest users.[151] Online sales, especially for younger adults, are increasingly coming from mobile and tablet devices; while 86% of those who purchase clothes online still use desktops and laptops to do so, tablets have become increasingly popular both for browsing and buying. In the 12 months to August 2014, 20% of online shoppers purchasing

[147] Retail Economics, Retail Sector Analysis, http://www.retaileconomics.co.uk/download/Retail%20Sector%20Analysis.pdf

[148] Office for National Statistics, August 2014, Statistical Bulletin, 'Internet Access – Households and Individuals 2014', http://www.ons.gov.uk/ons/dcp171778_373584.pdf

[149] Mintel, 12 September 2014, 'TOP OF THE ONLINE SHOPS: ONLINE FASHION CLICKS WITH 70% OF BRITS'. Press release http://www.mintel.com/press-centre/retail-press-centre/top-of-the-online-shops-online-fashion-clicks-with-70-of-brits

[150] Office for National Statistics, August 2014, Statistical Bulletin, 'Internet Access – Households and Individuals 2014', http://www.ons.gov.uk/ons/dcp171778_373584.pdf

[151] Office for National Statistics, August 2014, Statistical Bulletin, 'Internet Access – Households and Individuals 2014', http://www.ons.gov.uk/ons/dcp171778_373584.pdf

clothing did so via a tablet. This is higher than the number of consumers (13%) using mobile devices to shop online.[152]

3.2.2 Firms participating in the sector

Large number of retailers with different business models and distribution channels

The clothing retail sector is not concentrated; it is made up of a large number of heterogeneous retailers. In addition to the large number of players in this sector, there is also a considerable diversity of retailer types in the clothing sector, including a large number of clothing only retailers providing for the mass market, luxury fashion retailers, department stores and supermarkets that sell clothes and shoes as one part of a wider offering of products. Retailers and brands also differ in terms of their chosen distribution method.

Online vs offline

Traditionally, clothing retail was primarily confined to the high street, but many of these retailers now also operate online. There are now only a modest number of high-street stores that do not sell online (e.g. Primark). In contrast, there are also online-only clothing retailers who do not have bricks and mortar stores, such as ASOS, Boohoo and Net-a-Porter. However, according to research by Mintel, most of the growth in online clothing retail sales has come through existing high-street stores, as only 5% of clothing sales are through online-only clothing retailers.[153]

Clothing only versus generalist retailer

In addition to the online/offline distinction, some clothing retailers, such as Primark, H&M, New Look and several under the umbrella of the Arcadia Group, only sell clothing directly to customers through their own stores and online channels (we define these as 'clothing only retailers'). Other more diversified retailers sell a wider range of products over and above their clothing items, such as Marks & Spencer (we define these as 'generalist retailers'). Furthermore, some clothing brands supplement their distribution channels by offering their products for sale through stores bringing together clothing from many brands; we define these retailers as 'multi-brand retailers'.

Own brand versus multi-brand

In addition, a further differentiating factor is whether the retailer makes and sells its own products versus those that are simply stockists of other brands' clothing.

---

[152] Mintel, 12 September 2014, 'TOP OF THE ONLINE SHOPS: ONLINE FASHION CLICKS WITH 70% OF BRITS'. Press release http://www.mintel.com/press-centre/retail-press-centre/top-of-the-online-shops-online-fashion-clicks-with-70-of-brits

[153] Telegraph article, 17 December 2014, 'Omnichannel retail trends are setting the shopping style', http://www.telegraph.co.uk/sponsored/technology/4g-mobile/connected-retail/11297081/omnichannel-retail-trends.html

We provide some examples of these different types of retailers in Table 1 below.

Table 1: Categorising some familiar clothing retailers

	Clothing only retailer (own brand)	Clothing only retailer (multi-brand)	Generalist retailer also selling clothes (own brand)	Generalist retailer also selling clothes (multi-brand)
Offline and online presence	Arcadia group (Burton, Dorothy Perkins) New look GAP H&M Warehouse	Topshop Urban Outfitters Moss Bros	Marks & Spencer Tesco Sainsburys Asda TK Maxx	House of Fraser Debenhams John Lewis
Online presence only	Little Mistress Boohoo (Shop Direct)	ASOS Net-a-Porter Figleaves.com		Very.co.uk Isme.com

Source: Analysys Mason and DotEcon

There are some brands that may not fall into just one category defined above. For example, there are some clothes brands that sell their clothes through stores such as House of Fraser or John Lewis (which often act as a one-stop shop for branded clothing retailers), but who also have their own high-street stores and online propositions. Theses brands include those such as Hobbs, LK Bennett and Kurt Geiger.

Mail order business is shifting to online sales

Whilst there has also been a shift away from some of the more traditional clothing retail models such as mail order and catalogue based ordering systems, these still exist but are supplemented by digital retailing. For example, the Shop Direct Group, which was initially launched as the Littlewoods catalogue retail subscription service in the 1970s has more recently embraced online retailing. They want "*to invest in medium- to long-term innovation, as well as delivering more customer-unique email marketing and online product recommendations.*"[154]

[154] Essential Retail, 2 July 2014, 'Shop Direct CEO: seize chance to know your customer better than ever', http://www.essentialretail.com/news/ecommerce/article/53b3b77f34af1-shop-direct-ceo-seize-chance-to-know-your-customer-better-than-ever

Third party firms have also emerged to support retailers provide enhanced services

In addition to clothing retailers, a number of third-party firms (for example, Monetate, Peerius, Baynote) have entered the clothing retail market to help retailers provide a personalised shopping experience for customers based on data collected about the customer. For example, they can help the retailer provide personalised email-shots, tailored search results and on-site product recommendations based on observed and inferred customer data. These third parties provide the technology and analytical capabilities to retailers that do not have the in-house resources or capabilities to provide such services to their customers.

New business models

Furthermore, as consumers are spending more time and generating ever-increasing amounts of content on social media platforms, these provide an additional way to contact and interact with consumers to provide another platform for targeted advertising. For example, retailers can combine their customer databases (email addresses, phone numbers or device identifiers) with a social media site's database to identify and target advertising at consumers that have previously registered or interacted with the retailer.

The large number of retailers including niche brands can make it difficult for customers to search for clothing that they like, particularly on smaller screened mobile devices. For this reason, a number of third-party aggregator services/platforms have emerged, usually provided as a mobile app, that bring together a wide range of fashion retailers/brands on a single platform on which consumers can browse and shop. For example Mallzee, an app founded in 2012, presents users with clothing from over 100 retailers and allows users to like or dislike items (by swiping left or right), and make purchases through the app. Over time the app learns from the choices made by the user and presents/recommends clothes that the consumer is more likely to have a preference for. However, given that these services are still relatively new, and given the fast moving nature of the fashion industry, it is not yet clear how successful these new business models will be.

3.2.3 Competitive dynamics

Competition in the retail clothing market is based on a number of factors including price and choice. However, customer service and catering to the needs and preferences of customers is also important for retailers.

Long tail demand in online clothing retail

Unlike stores on the high street, online clothing retailers are not confined by physical retail space and this creates a much wider scope for the provision of a greater choice of clothing lines and brands, including niche products which meet the demand of a

smaller selection of consumers. For example, ASOS's website presents consumers with over 75,000 products (including clothing, footwear and jewellery) from over 800 brands ranging from every day brands to the niche designers.[155]

Virtuous circle in the online clothing retail market

The greater range of clothing online makes online shopping an appealing proposition for a greater number of consumers. However, the sheer volume of choice online means those consumers may actually find it difficult to find items that suit their individual tastes. Therefore, the adoption of online channels relies on making the sales journey easy and frictionless for the users. In this respect, powerful search functions and personalised recommendations are important tools. There is a virtuous circle because as consumers engage further with online channels, they contribute additional data that can be used to further increase online sales, average order values and minimise returns.

The extent to which a retailer sells own-brand clothing may influence exactly what value the retailer hopes to extract from the consumer data it collects; for example, it might conceivably provide additional data to help identify trends and improve size and fit of own brand clothing.

Blurring the offline/online distinction

For companies with both an online and offline presence, knowing more about its customers regarding their preferences, and purchasing behaviour may also help influence decisions taken in store including space allocation to particular brands for example. There is an increasing level of integration between the online and offline distribution channels and this could be another area we see retailers trying to compete on an improved customer interaction and service provision in future. For example, an in-store sales representative might be able to assist a customer who is seeking a product that is out of stock by using a tablet to order the product online for delivery to the customer's home, allowing the customer to pay in-store. Some firms may offer the option to buy online but pick up in-store, or to buy in-store but allow a refund process to be initiated online.

We understand that the integration of online and offline behaviour is still relatively underdeveloped in the sector as a whole and the integration of data collection across multiple channels and devices is seen as an important area of development in the sector.[156]

[155] ASOS annual report and accounts 2014, http://ASOSplc.com/~/media/Files/A/ASOS/results-archive/ASOS_Report_2014%20FINAL.pdf

[156] Deloitte, report for Ebay, February 2014, 'The omnichannel Opportunity, Unlocking the power of the connected consumer', p. 28, http://www2.deloitte.com/content/dam/Deloitte/uk/Documents/consumer-business/unlocking-the-power-of-the-connected-consumer.pdf.

3.3 Collection of consumer data

Online sales channels allow retailers to collect more data about customers than they previously could

Retailers, including clothing retailers, have for years collected data about their customers through loyalty programmes, surveys, catalogue subscription services and conversations with in-store sales assistants. However, in an increasingly digital age, the retail industry is able to benefit from a shift to online sales channels that allow retailers to collect more data about customers than they previously did.

Consumer data is collected both online and offline, and is increasingly important to retailers to provide an improved service to their customers.[157] Data is collected in a number of different ways in the clothing retail sector. The customer may provide the data explicitly (for example when registering for a service), but in some cases data collection may occur without explicit declaration. For example, behavioural data may be collected about the customer by tracking their purchase history or monitoring their movements when browsing a retailer's website. Sometimes third-party service providers will also be involved in the collection of data. Therefore, it is helpful to distinguish between that is volunteered by the customer to retailers and to third-party service providers and the collection of data via other methods such as cookies.

3.3.1 Data collected directly from the consumer

Consumers will typically declare basic personal information when registering

When registering for a service or wishing to make a purchase online, the consumer will have to volunteer some personal data. Depending on the service the consumer is registering for, this data will either be provided directly to the retailer, or directly to the third party providing an additional service (for example a size recommendation service).

Data collected by the retailer

Basic personal information is collected…

Customers may provide some personal data to a retailer when registering for an in store loyalty card, or when registering with the retailer online (for example when creating an account to make a

[157] Helen Dickinson, Director General of the British Retail Consortium expressed the view that the retail industry is in a *"new wave of transformational change. One that places customer data, with its "profound, strategic and qualitative" ability to affect change in the way we will sell, as a key differentiator."* http://www.spiked-online.com/event_readings/helen-dickinson-retail-week.pdf

purchase). At this point, the customer will typically be required to provide basic personal information, which enables the identity of a person to be known (name, address, payment details, etc.) and demographic information such as gender and date of birth (for customer segmentation purposes).

…and supplemented by transaction data

Once registered with the retailer, further customer data is gathered to develop the customer profile. Typically this is much more easily done online, as everything is linked directly to the user's registered account. For example, online retailers will bring together purchase history of registered users including size and style information, and in some cases data about brands and price range of items bought. Retailers will also record returns and exchange data and where the option is available the reason for a return (e.g. 'wrong size', or 'didn't look right'). Much of this information will be visible to the customer through its registered profile on the retailer's website.

Data may also be collected to enhance the customer profile over time by allowing the retailer to infer information about the customer's tastes and shopping behaviour – for example, explicit logging of the brands of products that an individual customer purchases over time to infer information on the customer's favourite brands. This information may be added to survey responses covering the same topic. Another example is the logging of dates and times of purchase, which over time allow spend patterns to become apparent.

Although online retailing provides greater scope for collection of data, customers will still tend to shop both online and offline.[158] Whilst collection of consumer data online is well developed, retailers are looking for ways to combine this data with the data collected in store to capture consumer's purchase behaviours across both settings.

[158] This view is also expressed by the UK Competition and Markets Authority (CMA)'s Philip Marsden (made in a personal capacity): *"As we all readily experience, some shoppers start searching in the real world, then go online to comparison shop, and then make the purchase; some go the other way; and some consumers start online, dip into a shop and pop back online again. Some even go online while in-store, comparison shop with other bricks or clicks outlets, then show these rival offers at check-out, which stores can then match or better, or lose that custom. To the consumer then, online and offline options are merging; and firms are experiencing this too, moving towards providing service in both channels to meet growing customer expectation."* https://www.gov.uk/government/speeches/philip-marsden-speaks-about-competition-enforcement-in-online-markets. Similarly, the Interactive Media in Retail Group (IMRG) agrees *"It's not either online or offline. We shop as we need to and as time dictates. There are times when we like to go to the shops and times when we prefer to shop online."* http://www.telegraph.co.uk/sponsored/technology/4g-mobile/connected-retail/11297081/omnichannel-retail-trends.html

Some data may be collected to link online and in-store transactions

Some retailers may be able to use payment card information to link purchases made by a particular customer. For example, a multi-brand retailer with both an online presence and nationwide stores told us that it collects information about purchases made with a particular credit or debit card in store. This can be used to link purchases made by a particular individual and can be used to link purchases made both in store and online where the same card is used. However, we heard from a number of stakeholders we interviewed that, at present, retailers are still not fully able to match customers and their behaviours between online and offline channels. For example, a consumer may conduct their research online, finding the cheapest and most suitable garment, but make the purchase in-store, in which case the online retailer may not be aware that the purchase has been made.

Whilst not widespread across the sector, there are examples of retailers trying to further integrate the online and offline customer experience by ensuring that in store retail staff have access to the customer's profile, in order to update it with data being collected by digital devices in-store. Box 14 provides one example of how a retailer is trying to develop customer profiles based on both online and in store behaviour.

Box 14: Data collection in store to improve the overall customer experience

In 2013/14, luxury clothing brand Burberry rolled out a new platform to its major stores including the flagship branch in London. The platform, referred to as 'Customer 1-2-1' is a customer permission-based iPad application tool that in store sales associates can use to view and maintain customer profiles, including *"a visual wardrobe, global transaction history online and offline, and recorded product and fit preferences."*[159]

This is part of Burberry's increasing focus on the use of technology in its stores, and as part of its 'Customer 360' program, Burberry invites shoppers to share their fashion phobias, buying history and preferences online when registering for the service. Combining information on an individual's purchase history, preferences and even Twitter posts with data on fashion trends, sales assistants in any Burberry store can then use the app to help make recommendations based on predictive analytics powered by SAP.[160]

Since September 2012, Burberry has also started to use Radio Frequency Identification (RFID) technology within some of its flagship stores. Initially used for stock control, in some stores the RFID tags are used to "*enhance the customer experience*". For example, Burberry can trigger interactive videos in store based on what garments the customer is trying on,[161] for example, "*Mirrors turn instantly to screens with runway footage and exclusive video*"[162]

Whilst there has been some media attention given to the possibility that this would be used to build customer profiles by tracking what garments its customers have tried on,[163] Burberry has specified that it does not currently link the information from the RFID tags to a customer or customer transaction although in the future it may link the RFID tags to the customer database. Burberry specified that this would not be done without the prior consent of its customers.[164]

[159] Burberry Strategy 'Accelerate retail-led growth', http://www.burberryplc.com/about_burberry/our_strategy/accelerate-retail-led-growth

[160] SAP article, 14 February 2014, 'Data Helps Burberry Engage Customers Wherever They Are', http://blogs.sap.com/innovation/industries/data-helps-burberry-engage-customers-wherever-01244657

[161] Burberry privacy policy, https://uk.burberry.com/legal-cookies/privacy-policy/rfid/

[162] Burberry press release, 13 September 2012, 'Burberry World Live arrives in London', http://www.burberryplc.com/media_centre/press_releases/2012/burberry-world-live-arrives-in-london

[163] Forbes article, 28 October 2013, 'How Fashion Retailer Burberry Keeps Customers Coming Back For More', http://www.forbes.com/sites/sap/2013/10/28/how-fashion-retailer-burberry-keeps-customers-coming-back-for-more/

[164] Burberry privacy policy, https://uk.burberry.com/legal-cookies/privacy-policy/rfid/

Data collected by other parties

In addition to retailers themselves, other specialist service providers (such as size and fit recommendation providers, personalised recommendation providers etc.) also collect data from the consumers. Such data may be collected directly from the consumer by asking them to fill out a simple form upon registering with the service and indirectly through the use of trackers or cookies that collect observed/behavioural data.

Data on size, style and fit preferences

Some retailers offer size, fit and style recommendations on their website. These services are typically provided by specialised service providers, through a 'plug in' or 'widget' inserted in the retailer's website. This 'plug in' or 'widget' usually appears as a separate pop-up or page on the retailers' website which asks the consumer to enter data on size and fit.

Although the exact data collected by the third-party service provider will differ depending on the extent to which they rely on exact body measurements versus general body shape and fit preferences, the main groups of data that are processed to provide recommendations include:

- size and shape (this can either be key body measurements such as waist, hip, chest measurements, or a simple qualitative self-description about body shape e.g. 'shoulder wider than hips' guided by illustrations);
- fit and style preferences (loose fit/tight fit, below the knee/above the knee etc.) as well as information on their hair tone, skin colour;
- existing items of clothing (brand preferences or particular items of clothing they have bought in the past that they especially like);
- personal data such as name, email address and key demographic data such as age and gender;
- items a consumer has purchased from them in the past, returns data, browsing behaviour etc. (collected by third-party cookies and provided by the retailer);[165]
- data collected from the retailer/manufacturer on the items of clothing. This may include key measurements, and a

[165] For example, True Fit advises that this additional data *"includes, but is not limited to, instances in which you affirmatively authorize third parties to provide us with information. For example, retail partners may provide us with information about items you have purchased from them in the past so that we can provide better fit recommendations for you"*. See True Fit's privacy policy, http://www.truefit.com/privacy-policy

product description. However, several of the recommendation services told us that this information is often poor or inaccurate;

- additional data input manually by the service provider. This includes viewing the clothing items up close and adding additional 'tags' in the database to provide more information on the style, the shape, the fit, the colour etc. Those providing visualisation services will also photograph the items worn by their models;
- general data about fashion trends may also be collected, which is necessary to understand why consumers may be returning a particular item of clothing;
- third-party non-personal data, including geolocation and weather data, so that recommendations could be seasonal for example.

Data on brand and store preferences

Some newer service providers (such as Lyst, Mallzee or Grabble) offer consumers a service which aggregates a multitude of clothing retailers and brands and provide consumers with a selection of clothing recommendations across retailers based on the consumers stated and/or observed preferences. All three of these services are provided using mobile apps and Lyst also provides this service on its website (however, over a third of the traffic to Lyst came from mobile devices in 2014).[166] Consumers can create feeds and place items they like into certain categories. Consumers are presented with suggestions of clothing and brands that they are likely to desire based on their preferences, based on behavioural data or data collected directly through registration. For example, when registering for these services, consumers are required to provide some personally identifiable data, and also data on stated preferences. The more data the consumer hands over to the platform the better the recommendation model becomes. The types of data collected by these services include:

- name;
- email;
- favourite products/brands/stores explicitly made at registration;
- social data;[167]

[166] Wired Retail article, 24 November 2014, "Half our 70 staff are data scientists' reveals Lyst cofounder', http://www.wired.co.uk/news/archive/2014-11/24/chris-morton-lyst

[167] Social data is captured when a user of the app logs in to the app with their Facebook account. Facebook then transfers a range of data to the app that enables it to connect users of the app to their Facebook friends and also search for Facebook friends on the app. For example see Mallzee's privacy policy, https://mallzee.com/privacy-policy.html

- browsing data;
- purchasing data; and
- wardrobe data i.e. information on existing clothing and styles.

3.3.2 Information about what consumers are saying about the brand

Retailers will sometimes conduct surveys

Some third parties conduct surveys for retailers and provide this data to the retailer to help improve their understanding of how their business is working and how customers feel about their overall experience.

The data collected is mainly the customers' own active replies to surveys, often taking place online, via email or SMS, at key points in the process (for example after making a purchase, or after completing the returns process). However, one survey company told us that they also do some passive data collection, mainly by recording some aspects of users' behaviour and characteristics such as the types of device used. It can also incorporate personal-level information from its clients' Client Relationship Management (CRM) systems.

'Social listening services' may also be used

In addition to surveys, retailers are increasingly turning to social media to engage with their customers. This may be through setting up their own 'fan pages' on Facebook for example. This allows the retailer to interact with customers who 'like' their page, allowing them to post messages to raise awareness of new products or sales. It also provides another point of contact for customers to interact with the retailer, sharing positive and or negative experiences.

Some specialised infomediaries enable a broader collection of social data, by collecting data from a very wide set of online sources and platforms including social media, blogs, forums, websites, comments and reviews for example. Whilst, much of the data collected and provided to the client is openly available on the internet it is personal information, including names, and any other information that may be publically included on a user's social media profile. There can also be a significant amount of metadata attached to the information collected. As an example, tweet-related metadata can include the timer and location at which the tweet was made for example. If someone is tweeting about a particular brand, then the brand owner will be notified and be able to see who has sent this tweet, any information linked to their personal profile and meta data, which may include geolocation data depending on the user's settings.

3.3.3 Behavioural data

Although the customer may provide key information to the retailer or to third parties providing services to the retailer, data is also collected from the users by tracking their behaviour by use of cookies, web beacons, tags or clear graphics interchange format files (GIFs). These are used to track customer behaviour on a given website, to remember certain preferences and also collect data essential for the correct operation of the site.

In some cases the cookies will be placed by the retailer themselves, whereas in other cases these will be placed by third parties who are providing additional services to the retailer such as product recommendation engines, size and fit recommendation engines, or data analytics.

Consumers may well be less aware that this data is being collected unless they read through the cookies policies or privacy policies of the company involved, however, the EC cookie laws introduced in the ePrivacy Directive may mean that customers are made more aware of how cookies are being used.[168] In most cases, customers are also provided with information about how they can prevent this data being collected via cookies in a 'How to Manage Cookies' section of the privacy and/or cookies policies.[169]

3.4 Use of consumer data

The key uses of consumer data in the clothing retail sector are to increase conversions and average order value, and to reduce the number of returns, which are costly for online retailers. There are a number of ways in which retailers (and specialist service providers) use data from customers to do this.

The main ways in which retailers use data to increase sales and conversions is by:

- making the online sales journey easy and frictionless for the user by personalising search results, on-site product recommendations and email recommendations; and

[168] In the UK, the Information Commissioners Office also provides clear guidance to organisations on how this EU cookie law must be followed and what it means for informing customers and receiving consent.

[169] For example see M&S's guide to disabling cookies, http://help.marksandspencer.com/support/company-website/disable-cookies-how-to

- providing targeted advertising and re-targeted advertising.

The main ways in which retailers aim to reduce returns is by:

- refining recommendation engines by incorporating data on the reasons for return;
- providing numerical based size and fit recommendations (typically involving third-party specialist service providers);
- providing visualisation tools and style and fit recommendation services (typically involving third-party specialist service providers).

Whilst we have categorised the use of data based on their primary purpose, the exact value proposition may differ between the different recommendation engines and specialist service providers; some focus on both reduced returns and increased conversions, so there is some degree of overlap.

Retailers may also make use of consumer data to inform strategic decisions. We discuss each of these main uses of consumer data in turn below.

3.4.1 Increase sales and conversions

Personalised search and product recommendations on-site

Personalisation is a key part of the business model of online clothing retailers

Many retailers, including clothing retailers, believe that providing a customer-centric, personalised shopping experience for their customers is critical to the success of their business.[170] The consensus in the clothing industry appears to be that personalised shopping recommendations are essential in today's online environment, and the collection of data is crucial in providing appropriate suggestions to consumers.[171] One might see this as using technology to deliver an analogous function to that an in-store sales assistant knowing a regular customer might have provided previously.

[170] In an online survey of more than 1,100 digital and ecommerce professionals working for brands and agencies (including retail), carried out in February 2013, 94% of companies agreed that personalisation *"is critical to current and future success."* E-Consultancy, 'The Realities of Online Personalisation in association with Monetate', https://econsultancy.com/reports/the-realities-of-online-personalisation-report

[171] Silverbean article, 18 March 2014, 'E-Commerce Expectations – What do UK customers want from online fashion retailers?', http://www.silverbean.com/blog/e-commerce-expectations-uk-customers-want-online-fashion-retailers/

As consumers interact with retailers through online platforms, this creates greater opportunities for retailers to collect data about individuals and build a picture of them to provide a personalised shopping experience.

Personalised search results, customised homepages and product recommendations on-site

A personalised shopping experience on-site can include product recommendations at different points of the customer's shopping journey e.g. on the home page, on the side of search results and at the point of check-out. Furthermore, a retailer may present products on the landing page and search results, which are consistent with the customer's current and previous browsing and purchasing behaviour. For example, a new petite clothing range may be displayed prominently on the home page because the customer has previously bought petite clothes.

In order to provide these personalised search results at the right time, real-time browsing data, collected by cookies or tags on the retailers' website, is taken and combined with more granular information, such as information on where the user is clicking on the website to develop an understanding of what each individual is trying to do as they navigate through the retailer's site, and looking at the time of day, the position in the purchasing cycle (for example browsing, added to shopping cart or at the 'checkout') to provide some context to the consumer's decision making.[172]

This observed/behavioural data may be combined with data that the retailer holds about the customer including declared data (such as gender and age) and inferred data from purchase and return history e.g. retailers can infer consumers' preferences – such as favourite brands, styles and colours through past purchases. The combination of real-time data and past purchases and returns ensures that recommendations are best suited to the customer's needs and preferences.

Personal and anonymous data can both be used to provide recommendations

If a retailer knows the identity of the shopper, and is able to make recommendations on this basis, it is akin to using personal (or at least pseudonymous) data. If the firm does not know the identity of the shopper, but is able extract some information about the shopper (i.e. their characteristics), it can then relate the shopper to known consumer segments (based on anonymous segment data) in order to make recommendations.

Some retailers will have the capabilities to provide personalised recommendations without the assistance of third parties. For example, Net-A-Porter and ASOS, two online only clothing retailers, put a strong emphasis on in-house IT teams and stylists to provide recommendations that are relevant to each customer.

[172] For example, see http://www.peerius.com/assets/PEE1285-SMART-recs3.pdf

Box 15: Net-A-Porter uses consumer data to provide a personalised service to customers

Net-A-Porter (a major international online luxury fashion retailer) believes that providing a personalised service to its customers must be data led and it gathers information from all the customer 'touchpoints' including data on the customer's purchase history, what and where they have been browsing on the site and may infer certain details about the customer. For example, a female customer who has bought an item from 'Mr Porter' may be considered to be in a relationship.[173]

Such data collected by Net-A-Porter is used to suggest items to customers in a more intelligent way by for example only presenting clothing in the consumers' size and by providing a 'Wear It With' service. This service allows the site to "*recommend other items from the 650 premium brands from around the world that may match an item in a customer's basket to create a whole outfit*."[174] Net-A-Porter has a 300 strong in-house IT team for customer relationship management.

[173] ComputerWeekly.com, 15 August 2014, 'Interview: Hugh Fahy, CIO, Net-A-Porter', http://www.computerweekly.com/news/2240226917/CIO-Interview-Hugh-Fahy-CIO-Net-A-Porter

[174] Ibid.

Box 16: ASOS and personalised recommendations

ASOS is a global online only fashion retailer and it has a rapidly growing active customer base of over 9.1 million.[175] This puts ASOS in a strong position to collect vast amount of data about its customers and create advanced segmentation models to target those customers with relevant merchandising and offers. ASOS has "*[d]edicated teams to collect and analyse data from customer spend and site usage habits to establish important trends*."[176]

ASOS mines the data it collects about its consumers (purchasing behaviour, favourite brands, spend patterns and payment method, as well as basic personal information and lifestyle data).[177] It uses this data to segment its customers into groups based on purchasing habits and identify the frequent shoppers and those who bought specific brands. ASOS can use this to gain useful insights into the purchasing behaviour of groups of individuals.

ASOS also uses the customer data it collects by building its own bespoke recommendation system, which it is able to do because it "*know[s] the ASOS customer better than anybody else*." ASOS believes that the amalgamation of machine learning techniques with staff that choose fashion, design it, etc. allows ASOS to build the best recommender systems that they can.[178]

Many other retailers also use in-house teams where they have the scale and capabilities to do so. For example, we interviewed a major high-street retailer with nationwide stores and an online presence, which said that it conducts about 95% of their data analysis in-house. However, not all retailers have the capabilities or resources to provide such services in-house.[179] A recent survey of online businesses found that whilst personalisation is seen as vital for business performance, the extent of tactics and testing and the ability of retailers to provide personalisation is limited because most companies are still held back by technology. A compounding factor is that data sources are often disparate given that consumers use

[175] This figure is based on customers having shopped with ASOS in the last 12 months up to 31 December 2014, http://www.ASOSplc.com/investors.aspx

[176] ASOS Annual Report and Accounts 2014, p 17, http://ASOSplc.com/~/media/Files/A/ASOS/results-archive/ASOS_Report_2014%20FINAL.pdf

[177] Marketing Magazine article, 1 December 2008, 'ASOS targets the individual', http://www.marketingmagazine.co.uk/article/865320/ASOS-targets-individual

[178] Presentation by David Williams Director of customer intelligence at ASOS, https://www.youtube.com/watch?v=Aci_XFqzcws.

[179] For example, this retailer did mention that it outsourced one data analysis project to a third-party firm that has significant expertise in real-time data analysis, noting that it outsources because the retailer itself does not have sufficient capabilities to undertake this analysis.

multiple channels to interact with retailers and therefore combining and analysing this data requires significant IT resources.[180]

Retailers without in-house data analysis capabilities can employ third-party service providers to provide the technology and analytics capabilities to retailers to enable them to create a personalised shopping experience on their website. Companies that provide these services include Peerius, Monetate, Baynote and RichRelevance (amongst others).

One third-party infomediary we interviewed explained how it uses customer data to provide recommendations. It collects data on consumer behaviour and preferences via a tag on the retailers' web pages, which sends information to its servers. The tag creates an anonymous ID for the retailer to follow what the user does on that site and see where conversions take place. The infomediary collects this data across websites, sessions and devices (PC, tablet, smartphone) in order to attempt to draw a unified profile of individual customers e.g. not only captures the customer's journey across the client's website and app, but also the referring website and referring search terms.

In some cases, third-party recommendation engine providers will enrich this observed data with data provided by the client and external sources providing more information. For example, one of the infomediaries we interviewed informed us that they enrich data collected with data from clients' customer relationship management (CRM) systems and non-personal data from third parties, such as weather information. However, not all infomediaries providing personalisation services will have access to data from the CRM such as customers' shopping history (both browsing and buying). For example, one online retailer told us that they do not share any customer data with their third-party recommendation engine provider. Where customer data isn't provided the recommendation engine providers may instead "[focus] *on the individual by examining their unique behaviour on the site*",[181] making product recommendations at the cart page/check out based on the mix of products included in the consumer's basket, based on a variety of factors, including pre-configured clothing collections and price ranges.[182] The latter approach adopted by companies such as Peerius "*doesn't slavishly follow what*

[180] E-Consultancy, 'The Realities of Online Personalisation in association with Monetate', https://econsultancy.com/reports/the-realities-of-online-personalisation-report

[181] Peerius, Arcadia Group Case Study, http://www.peerius.com/clients/arcadia-group-limited/

[182] Ibid.

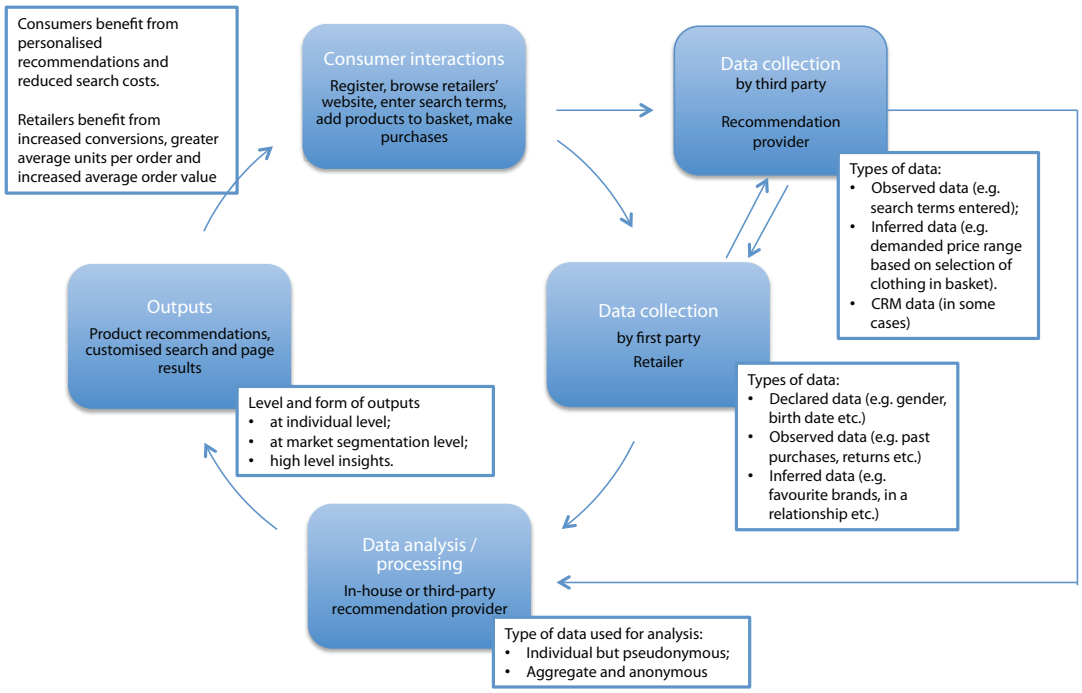
the customer has previously looked at and bought, or what others have bought in similar circumstances. Instead it focuses on the individual by examining their unique behaviour on the site."[183] This may be an indication that the extra value associated with including data on past purchases may increase the complexity of the model with limited incremental benefit.

The recommendation provider will use the data collected as an input to a recommendation engine that applies proprietary algorithms, or customised algorithms specified by the client's requirements and key performance indicators (KPIs), to offer recommendations. These service providers are often able to improve their algorithms through experience with multiple clients (although they are not able to enrich personalised profiles with data from multiple retailers). The algorithms used will give different weights to different types of data to ensure that the recommendations are still relevant. For example, we understand that the most recent purchase and browsing history are given more weight when providing recommendations. One large multi-brand retailer told us that their online recommendation engine only takes into account past purchases made within the last six months, and one style recommendation service provider described how they only consider fairly recent style choices, as individuals' decisions and preferences are likely to change over time, especially in the fashion industry where trends play a big role in influencing consumer choice.

Figure 9 below illustrates the types of data collected and the data flows between the various stakeholders in the data value chain i.e. consumers and data collectors (retailers and third-party recommendation providers). The outputs of the analysis are fed back to the consumer in the form of personalised recommendations and customised web pages.

[183] Ibid.

Figure 9: Data flows associated with providing personalised service search and product recommendations



Source: DotEcon and Analysys Mason

The customer benefits from receiving relevant recommendations and reduced search costs. These customer benefits can lead to greater engagement with online sales channels thus increasing sales (both units and value) and conversions, which is a clear benefit to retailers. The box below provides some specific examples of reported improvements retailers have achieved when using recommendation services provided by third-party retailers. We could not obtain similar data for recommendation services provided by retailers themselves because this data is commercially sensitive.

Box 17: Specific examples of success rates as reported by service providers

One third-party provider of recommendation services reports that retailers using its service can expect to gain a sales uplift of up to 20% and an increased Average Order Value of up to 50%.[184]

Arcadia group, who have been using Peerius technology since 2012 have reported an average order value increase of 67%, an increase of 66% for average units per order, and 7% of Arcadia's online sales were driven by Peerius alone.[185]

Women's active clothing retailer 'Sweaty Betty' has also reported increased activity as a result of the recommendations service: 6% of total online sales in the month following the introduction of recommendations came from engagement with such recommendations. Furthermore, for orders being made that included recommended products, there was a reported 75% increase in average order value and a 58% increase in conversion. [186]

Fee structure

The contractual relationship between retailers and their recommendation service providers is not widely advertised. However, reviewing publically available information from some of these service providers suggests that different pricing models are adopted. We understand that some provide their services on a cost per action basis (based on the conversion of recommendations to actual sales) whereas others base it on a flat fee.[187]

Consumers are informed of data collection for personalisation in retailers' privacy and cookies policies

Customers will by definition be aware of the collection of personal data they provide retailers directly by registering for their services online. The use of cookies and other tracking devices used to collect behavioural data for the purposes of creating personalised webpages and recommendations is normally clearly noted in the privacy policies and cookies notices of online retailers. For example, one retailer's privacy policy explains that it "*use[s] browsing behaviour data to create personal product recommendations*"[188] and another specifically listing all the cookies used by the site and their

[184] See http://www.peerius.com/smart-solutions/smart-recs/#modal

[185] Arcadia Group Press Release, 'ARCADIA FASHIONS A 67% JUMP IN ORDER VALUE WITH PERSONALISED PRODUCT RECOMMENDATIONS', https://www.arcadiagroup.co.uk/press-relations/press-releases-1/personalised-product-recommendations

[186] RichRelevance, Sweaty Betty Case Study, http://info.richrelevance.com/rs/richrelevance2/images/Sweaty%20Betty%20RichRecs%20Case%20Study%20UK.pdf?mkt_tok=3RkMMJWWfF9wsRons6rIZKXonjHpfsX56usoXKa%2FlMI%2F0ER3fOvrPUfGjI4ASMVrI%2BSLDwEYGJlv6SgFS7TCMatny7glUhU%3D

[187] For further information on the contractual relationships see Analysys Mason (for Ofcom), 2014, 'Online data economy value chain, Annexes' p. 26.

[188] Sweaty Betty's privacy policy, http://www.sweatybetty.com/security-and-privacy/

functions, and expiry time.[189] Further, most privacy policies make it clear that information is provided to reputable third-party service providers so that they can *"report on customer behaviour and/or preferences."*[190] Given the benefits to the consumer it is perhaps unsurprising that a major retailer with an online and offline presence informed us that they have not received any complaints from consumers about the use of data to improve the level of personalisation provided on their website and/or emails.

Targeted advertising

As with many online services, targeted advertising is a key part of the online clothing retailers' marketing strategy. The main methods used to provide targeted/tailored adverts to consumers are:

- targeted email shots that display personalised content/product recommendations;
- targeting/re-targeting to display relevant advert banners to customers when browsing other websites (including social media).

In addition to targeted advertising to existing customers, data brokers can assist retailers to find prospective customers with particular characteristics and target those with relevant advertising.

Targeted email shots

Data collected by retailers is used to target advertising and merchandising to consumers via personalised emails. Consumers are segmented into small groups based on similar characteristics e.g. gender, clothing brands they like or previously bought, willingness to pay and so on. Emails may include personalised content such as:

- clothing from brands that might appeal to the consumer based on what he/she previously bought and/or based on and what other people that bought that brand also bought;
- clothing that falls within a certain price bracket that might appeal to the customer based on past purchasing decisions and browsing behaviour;[191]

[189] See http://www.boohoo.com/restofworld/cookies/page/cookies and

[190] Sweaty Betty's privacy policy, http://www.sweatybetty.com/security-and-privacy/

[191] For example, see details of ASOS' tailored email marketing strategy in Marketing Magazine article, 1 December 2008, 'ASOS targets the individual', http://www.marketingmagazine.co.uk/article/865320/ASOS-targets-individual

- items the retailer considers the customer may be interested in based on other characteristics such as age, gender demographic etc.;
- advertisements or reminders for items of clothing the customer has previously added to their basket but did not subsequently order (referred to as 'cart abandonment re-targeting');
- retailers that sell other products in addition to clothing may also seek to cross-sell products that similar customers have shown interest in previously.

Delivering relevant adverts to the customer and re-targeting

In addition to emails, retailers may target customers through tailored advertisement banners that can be shown to customers when viewing their site and other websites online. This will often display items that the customer has expressed interest in by viewing the product or by adding the product to their basket/cart but not completing the purchase (this is termed 'cart abandonment re-targeting').

Often this requires the involvement of third parties using data collected through cookies to provide insights on the items that the customer has shown an interest in. For example, sites which are using retargeting technology will have a JavaScript tag included on the page, which adds a cookie to the user's browser when they visit that website. This cookie allows that customer to be tracked around the web, and see what products they are viewing. This then allows targeted display adverts to be delivered to the customer on subsequent sites they visit.[192]

Firms involved and pricing models

Criteo holds the largest share of the market for re-targeters across retail sites in the UK (41%), with other major service providers including Avail (RichRelevance) and Stuq.[193] A number of different pricing models may be served by these service providers including costs based on per thousand impressions served, cost per single click, cost per action or simply a monthly fixed fee. We understand

[192] Econsultancy, March 2014, 'Display Retargeting – Buyer's Guide', p. 8, http://www.criteo.com/media/1030/research-3.pdf

[193] Ibid. Figure 2. However note this excludes *"Google Remarketing, who are a leading provider of remarketing services across the Google Display Network (GDN) and using Google Analytics. For legal and privacy reasons, Google did not provide tagging or market share data to Pikslme, therefore they were not included in the market share charts."* p. 14.

that the majority of vendors work on a cost per impressions serviced.[194]

Cross platform re-targeting

Targeted and re-targeted adverts provided to consumers via email or website banners are valuable for the retailer primarily in minimising cart abandonments and increasing conversions. However, targeted email advertising and banners may not be entirely effective. Retailers understand that consumers use several platforms at different times during the day – for example, a consumer may check emails on a work PC in the morning, but browse social media on a phone at lunchtime or in the evening. Retailers may not effectively engage with consumers by collecting data from one channel. To target consumers, single channels such as email may be ineffective, both because some consumers may not open emails received and because they require reminding of particular promotions at a different time in the day when he/she may have more time to engage. Therefore, an effective marketing campaign aims to reach out to consumers on multiple channels and some re-targeting providers are offering cross-device re-targeting services.[195]

Re-targeting through social media

In order to expand the reach of their targeted advertising retailers are also turning to social media to target consumers. As with the re-targeting, the consumer does not have to be registered or have bought an item with the retailer in order for the retailer to target them via social media sites. Social media networks such as Facebook enable retailers to target consumers that have been browsing the retailer's website using data collected by the social media site's tracking pixel.[196] This "*reports back general information about the browsing session, a hashed version of the Facebook ID, and the URL being viewed. Optionally, an advertiser can send Custom Data back to Facebook that contained additional information about the browsing session.*"[197] This data is then used to serve a targeted advert at a specific user on Facebook.

---

[194] For further information on the contractual relationships, data flows, funds and sources of value see Analysys Mason (for Ofcom), 2014, 'Online data economy value chain, Annexes' pp. 21 – 23, http://stakeholders.ofcom.org.uk/binaries/research/online-data-value/Annexes_to_Analysys_Mason_report_online_customer_data.pdf

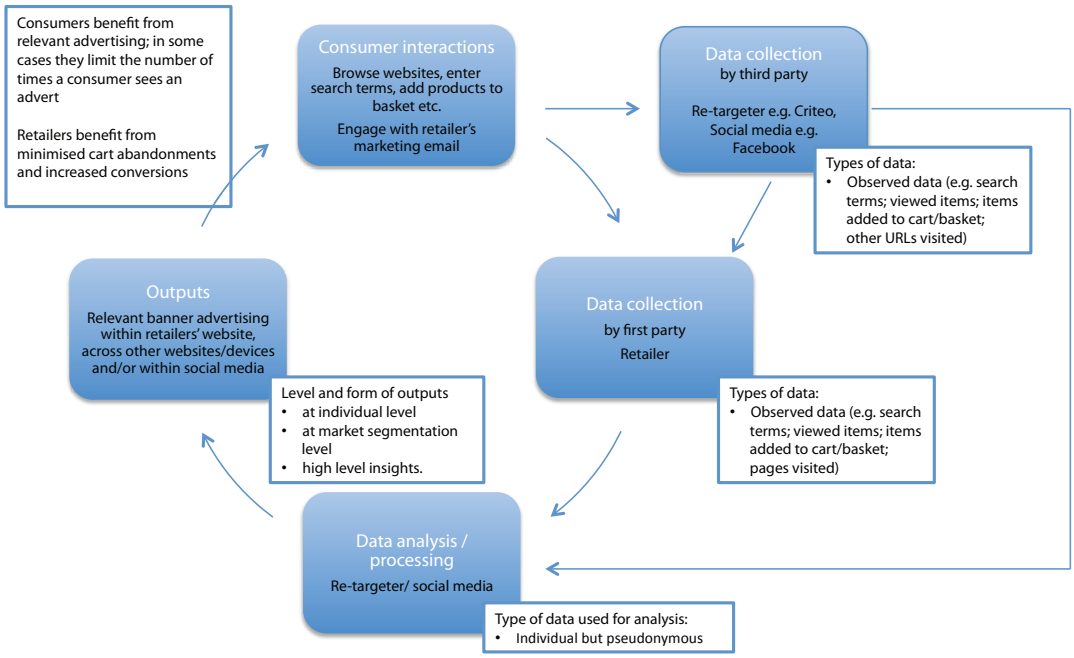
[195] For example, Struq markets a cross device retargeting service, which it claims could drive an extra 30% of sales. See http://www.struq.com/home/features/

[196] Retailers can install a 'custom audience pixel' (a piece of javascript) on every page of its website to reach the people who have been browsing their website with a targeted Facebook advert. See https://developers.facebook.com/docs/marketing-api/custom-audience-website/faq/v2.2

[197] See https://developers.facebook.com/docs/marketing-api/custom-audience-website/faq/v2.2

Figure 10 below illustrates the data flows associated with re-targeted advertising provided across websites and within social media sites.

Figure 10: Data flows associated with re-targeting via cookies



Source: Analysys Mason and DotEcon

The consumer benefits by receiving relevant advertising and in some cases the cookies will limit the number of times a consumer sees an advert.

Customers can see information about re-targeting cookies in retailer's policies

Customers can find some information about the cookies used to provide targeted advertising on other websites, and these are specified in the privacy policies and/or cookie policies of the online retailers. In some cases, these explicitly mention the use of third-party cookies where the retailer uses a third party to provide the targeted advertising services. Box 18 below provides some extracts from various privacy policies.

Note that in the case of Facebook's advertisements, no personal information is reported back to the advertiser/retailer. Furthermore, advertisements can only be targeted to groups of those browsing the retailer's site and only "*once they have reached a critical mass by which it is impossible to learn the individual identity an [sic] any person visiting a website.*"[198]

[198] https://developers.facebook.com/docs/marketing-api/custom-audience-website/faq/v2.2

Box 18: Extracts of privacy policies of online clothing retailers - targeted advertising

ASOS – *"Targeting Cookies or Advertising Cookies - These cookies collect information about your browsing habits in order to make advertising relevant to you and your interests. They remember the websites you have visited and that information is shared with other parties such as advertisers. For example we use 3rd party companies such as Criteo to provide you with more personalised adverts when visiting other websites."*[199]

Net-A-Porter – *"Targeting cookies or advertising cookies - These cookies are used to deliver adverts relevant to you. In addition, they limit the number of times you see an advertisement as well as helping us measure the effectiveness of our advertising campaigns."*[200]

Arcadia Group – *"third party cookies from business Partners such as Struq and Criteo to provide you with personalised adverts when you visit other selected websites."*[201]

Sweaty Betty – *"Browsing behaviour is also used to create relevant banner advertising with product recommendations that we believe best relate to the items you viewed on our website. These banners are then served across other websites that you may visit, typically news sites, video sites and blogs. This process is called behavioural advertising and although the adverts are individually tailored to each visitor, all cookie data is anonymised and stored temporarily. Additionally, all banners of this type feature an "i" icon that provides more information from each advertising network and instructions on how to opt out."*[202]

Other forms of targeted advertising on social media

Facebook also provides targeted advertising services based on customer data the retailer has collected. For example, using Facebook's Custom Audiences service,[203] retailers can define their target audience based on customer data they have already collected, including email addresses, phone numbers and other identifiers such as Facebook User IDs, app user IDs, Apple's Advertising Identifier (IDFA), or Android's Advertising ID for example.

[199] ASOS's privacy policy, http://www.ASOS.com/infopages/pgeprivacy.aspx

[200] Net-A-Porter's Privacy and Cookie Policy, http://www.net-a-porter.com/gb/en/

[201] For example see: http://www.burton.co.uk/webapp/wcs/stores/servlet/CatalogNavigationSearchResultCmd?catalogId=33052&viewAllFlag=false&categoryId=281999&langId=-1&storeId=12551&interstitial=true#fragment-8

[202] Sweaty Betty's privacy policy, http://www.sweatybetty.com/security-and-privacy/

[203] *"Custom audiences allow advertisers to target their ads to a specific set of people with whom they have already established a relationship on/off Facebook."* See https://developers.facebook.com/docs/marketing-api/custom-audience-targeting/v2.2

Targeted advertising to prospective customers using social media

Whilst the above method is based on targeting specific customers using a type of personal identifier, Facebook also offers a service called 'Lookalike Audiences'. This service allows advertisers to extrapolate their targeted advertising campaign to reach more people who 'look like' their established customers that have been identified through the Custom Audiences service.[204]

As described in Box 19 below, there are controls on the exchange of data between the advertisers/retailers and Facebook and the way in which the retailer collects data for advertising purposes. The customer data provided by the retailer/advertiser e.g. email, phone number etc. will be hashed and only used by Facebook to match the data to a Facebook profile. According to Facebook this data is not shared with third parties and is deleted promptly after the match process is complete.

Box 19: Advertisers must comply with Facebook Terms of Service when using their ad targeting services

In using the Custom Audiences feature, advertisers do have to comply with the Custom Audiences Terms of Service, and Facebook will not give access or information about the Custom Audience to any third-party or advertiser without permission. Some highlights from the Custom Audiences Terms of Service include:

- *"You represent and warrant that you (or your data provider) have provided appropriate notice to and secured any necessary consent from the data subjects whose data will be hashed to create the Hashed Data, including as needed to be in compliance with all applicable laws, regulations and industry guidelines. If you have not collected the data directly from the data subject, you confirm, without limiting anything in these terms, that you have all necessary rights and permissions to use the data. If you are using a Facebook identifier to create a custom audience, you must have obtained the identifier directly from the data subject in compliance with these terms."*
- *"To the extent a data subject exercises such an opt-out after you have used data relating to that data subject to create a custom audience, you agree to remove that data subject from the custom audience"*
- *"The Hashed Data you provide to us will only be used for the matching process, will not be shared with third parties or other advertisers and will be deleted promptly after the match process is complete"*[205]

[204] Facebook describes this in more detail: *"A lookalike audience uses several kinds of user set as a "seed" and an audience is built of similar users…The seed can be of these types: An existing custom audience, Campaign/Ad Set conversions, Conversion data based on conversion pixels, Page like users."* See https://developers.facebook.com/docs/marketing-api/lookalike-audience-targeting/v2.2

[205] For the full Custom Audiences Terms of Service see: https://www.facebook.com/ads/manage/customaudiences/tos.php

For targeted advertising on Facebook, retailers are charged for adverts on a 'cost per click' or a 'cost per impression' and the advertiser/retailer can simply set a daily or a lifetime budget that will never be exceeded.[206] Therefore, the flows of funds are simply from the advertiser/retailer to Facebook.

Targeted advertising services provided by data brokers

Other third-party data brokers, provide services similar to those provided by Facebook, allowing retailers to improve their advertising by understanding more about its existing customers and expand the reach of its advertising to prospective customers. One data broker informed us that if a retailer sends it data on its customers (e.g. customers who are loyal to the retailer, spend significant amounts etc.) the data broker may be able to (a) identify these customers in its own database (b) analyse many other characteristics about this group of which the retailer may be unaware, thereby building a model and (c) based on this, identify many other prospects from a larger universe in the data broker's own database. These prospective consumers can then be targeted via email,[207] on other websites using cookies placed by the data broker or targeted via advertising on social media (linking data with Facebook Customer Audiences).[208] However, data shared with Facebook is encrypted.[209]

Other examples include Experian who assisted a luxury US retailer to reach target audiences through Facebook after having identified customers that were not engaging with emails. Experian segmented the luxury retailers' email database into inactives (where the recipient receives the email but may not open or visit the website) and bouncebacks (where the recipient does not receive email because, for example, they have changed their address). By segmenting consumers into these categories, the retailer was able

[206] https://www.facebook.com/help/214319341922580

[207] See Acxiom's Liberty London case study 'CREATING NEW CUSTOMERS THROUGH PRECISE TARGETING' : http://www.acxiom.co.uk/resources/liberty-london/

[208] *"A profiling exercise of a client's database would help identify key audiences and enable Acxiom to find lookalikes of these audiences within Acxiom's consumer lifestyle database (over 40 million consumers and 1,000 variables). These audiences can then be uploaded onto Facebook and shown relevant ads."* See http://dq2qu0j6xxb34.cloudfront.net/wp-content/uploads/2014/01/Facebook-Case-Study-Final-no-bleed.pdf

[209] As explained by Acxiom "*Facebook is able to match users without actual data ever being shared. When the data is uploaded onto Facebook, each record is hashed i.e. the email address or phone that is supplied for the matching is turned into a short fingerprint that can't be decrypted.*" See http://dq2qu0j6xxb34.cloudfront.net/wp-content/uploads/2014/01/Facebook-Case-Study-Final-no-bleed.pdf

to focus its Facebook marketing spend on targeting consumers that were not engaging via the traditional email channel.[210]

3.4.2 Use of data to reduce returns

Being unable to try on garments to check the size and fit before buying can deter consumers from buying clothes online. Therefore, to encourage consumers to buy online many, if not all, UK clothing retailers provide returns policies that are convenient and free-of-charge (providing returns labels, convenient collection points and so on) and immediate refunds. However, this can be a costly strategy, as the retailer then pays the cost of carriage. Furthermore, it can have implications for stock levels and cash flow because providing free returns can encourage consumers to buy multiple sizes of the same item and return those that are not the right size or fit. In the UK, on average consumers will return 1 in 4 clothing items bought online.[211]

Therefore, retailers have an incentive to manage these costs by reducing the likelihood of returns. Retailers may be able to reduce returns by feeding returns data into their personalised products recommendation engines. For example, ASOS provides a tick-box form with options such as 'Ordered more than one size', 'Doesn't fit properly', 'Doesn't suit me' etc.[212] which is then fed into its recommendation model. However, providing more suitable recommendations does not eliminate the fact that consumers may still have a preference to better understand how the item will fit before buying. For example, without size standardisation across the sector consumers may not be confident that a particular size from one brand will fit the same as a different brand's item in the same size. Furthermore, according to a recent survey 45% of all shoppers who regularly or always buy clothes online find the inability to try on the garment to check the fit "*the most disappointing aspect of shopping online*".[213]

---

[210] Experian case study, 'Experian's combined Facebook and e-mail marketing activity drives an ROI increase of 350% for a luxury retailer', http://www.experian.co.uk/assets/marketing-services/case-studies/case-study-luxury-retailer.pdf

[211] Ibid.

[212] Asos's returns note, http://asos.custhelp.com/ci/fattach/get/9346330/0/filename/United+Kingdom+-+Returns+Note.pdf

[213] Fits.me, February 2014, 'How fit is online fashion? Consumer attitudes towards clothes shopping online', research report, http://communication.fits.me/acton/attachment/8391/f-0003/1/-/-/-/-/Fits_me_How_Fit_Is_Online_Fashion_report_UK.pdf

In response to this issue, some clothing retailers are providing standardised size recommendation services, online visualisations/fitting rooms and style and fit recommendation services on their websites. Below, we provide a number of examples of services that use consumer data to provide services that reduce returns. However, we acknowledge that the exact value proposition may differ between the different service providers. For example, numerical based size and fit recommendation services will be focused primarily on reducing size based returns, whereas visualisation tools seek both reduced returns and increased conversions.

Numerical based size and fit recommendation

Specialist service providers such as Virtusize and True Fit use numerical models to provide consumers with size and fit recommendations. These outputs are based on information explicitly provided by the user on brands, styles, or particular clothes that they already own that fit well. For example, Virtusize allows customers to enter measurements of an existing item of clothing that they know fits well. Similarly True Fit asks the user some simple questions on body shape, fit preferences and questions on any brands or styles that they customer knows fit perfectly and uses this data to provide consumers with a fit rating for each garment indicating how well the item will fit that specific consumer. It may also recommend the best size for a specific consumer in a specific style (Box 20).

Style and fit recommendation services (Box 21), have tried to differentiate themselves from these pure size recommendation engines by focusing on helping consumers choose a garment to match their style and preferences as well as simply just the fit.

Visualisation tools and virtual fitting rooms

Services such as Metail or fits.me (Box 22) are focused on helping consumers choose garments of the rights size with the help of a visualisation tool akin to a 'virtual fitting room', which allows customers, upon entering key body measurements, to see how a garment of a given size might look on them.

Box 20: Numerical based fit score services example – True Fit

For first time buyers or guest shoppers, consumers can create a True Fit profile by answering a few questions and then see their personal fit rating and size for all styles brands and stores.[214]

For example, on the House of Fraser website, customers can get personalised size and fit recommendations as they shop by providing a few details about themselves, providing details of a few previous they already own and their favourite brands.[215]

The customer will receive a personal fit rating expressed on a five point scale indicating how well the item will fit that specific user. TrueFit will also recommend the best size for a specific user in a specific style.

True Fit can also provide repeat buyers with personal size recommendations based on purchase history.

In order to provide this service**,** "*True Fit leverages billions of rich data points to show consumers how well the clothes and shoes they view onscreen will fit them in real life...Consumers also receive constantly improving recommendations as True Fit's machine learning engine learns more about what truly fits and flatters each user.*"[216]

Box 21: Style and fit recommendations example - Dressipi

In addition to assisting users with size guides, Dressipi also seeks to be a personal style guide and provides recommendations based on style and fit preferences rather than simply matching measurements of the customer to measurements of the garments.

Using a combination of individual algorithms and style advisers, Dressipi can provide the customer with tailored product recommendations at a number of different stages of their shopping journey, for example when browsing and searching or when entering items into the shopping cart.

They also provide a style guide helping the customers understand what styles are best for them and their body shape. Dressipi will then link this information to specific items of the retailer's stock to help make it easy for the customer to buy these items.[217]

Dressipi's clients include the likes of M&S, Arcadia Group and Shop Direct, who use Dressipi's product mainly as part of their web or mobile e-commerce propositions.

[214] http://www.truefit.com/products/for_retailers

[215] http://www.houseoffraser.co.uk/True+Fit/TrueFit,default,pg.html

[216] See http://www.truefit.com/about-us/the_true_story

[217] For example, see http://partners.dressipi.com/style-adviser.html

Box 22: Virtual fitting rooms example - fits.me

Fits.me has partnered with a number of online clothing retailers to provide a virtual fitting room for their website. When viewing clothing items on the retailer's website, customers have the option to "*Find the Size in 3D*".

This takes customer to a 'virtual fitting room' powered by Fits.me providing shoppers with a photo-accurate visualisation of fit. Users enter their body measurements (or these can be estimated based on your age, height, weight and overall body shape) and then they are shown how the item of clothing chosen will fit their body (based on a photograph of a robotic mannequin)

"The shopper sees exactly how the garment they are thinking of buying will look on their exact body size and shape when the item is delivered. The shopper can choose to see how other sizes will look on them, giving a looser or tighter fit, and will even warn the shopper where the fit may be unsatisfactory, such as in the sleeve length or collar size." [218]

The exact output of these various service providers is different, i.e. visualisation of garment fit, recommendation on size and/or style and therefore the exact types of data they collect vary from exact body measurements or measurements of existing clothes, to broader declared data such as body shape and fit preferences. However, the flows of data between the customer, the retailer and the service provider are similar (as described below).

In most cases, these specialist service provider tools are included in the retailer's website in the form of a 'plug-in' or 'widget' that appears to be an integral part of the retailer's website. To the extent that the service provider collects and processes data from consumers directly, their brand is normally displayed somewhere in the widget, however, how visible this is varies across retailers and in some cases the service provider's brand is only provided in small print. Therefore, whilst consumers are interacting with these service providers directly they may not be aware that they are providing data to a third party.

Customers will typically be able to take advantage of the service on a retailer's website without having to register with the service provider, and in some cases, a consumer can revisit the retailer's site without saving their fit profile and the site will recall the basic consumer data entered (through the use of cookies). If the customer does choose to register with the service provider they may be required to provide some more personal information (including more detailed measurements, name, age, contact details, gender) which can then help them get more accurate

[218] For example, see http://communication.fits.me/acton/attachment/8391/f-000e/1/-/-/-/-/PG_case_study_A4%2B3_UK_2.pdf

recommendations for future purchases. Once they have registered with the service provider, they will be able to log in to their fit profile when using other retailer's websites who also use the service.[219]

In any case, the service provider combines information entered by users with their actual behaviour (such as purchase behaviour and browsing activity) to build a more accurate and up-to-date picture of the customer and their preferences. For example, a number of size and fit recommendation services told us that they have found consumers are not very good at measuring themselves, so there is a benefit associated with complementing stated sizes with evidence about the sizes they actually purchase. This additional information is collected by the third party using cookies and in some cases includes information provided from the retailer about the customer.

Making size and fit recommendations also requires the matching of consumer data to the right size in the brand of garment being bought. Therefore product data including key measurements, product description, style and shape are also required to provide the outputs. However, tolerances in the manufacturing process may limit the usefulness of product data. Based on interviews with these service providers, it appears that these tolerances are on the whole pretty high, which means that getting precise information on a garment is a barriers to the accuracy of quantitative fit engines.

Our understanding from interviewees is that all of the data collected is used as an input to the recommendation engines and complex algorithms are used to process the data and provide recommendations to the customer. The algorithms are updated over time and statistically validated using A/B testing and more sophisticated machine learning.[220] Where style is also an important part of the recommendation services, one such service provider informed us that, fashion experts are also involved to ensure that the data-driven approach is actually providing suitable recommendations that will work for the individual. Collaborative

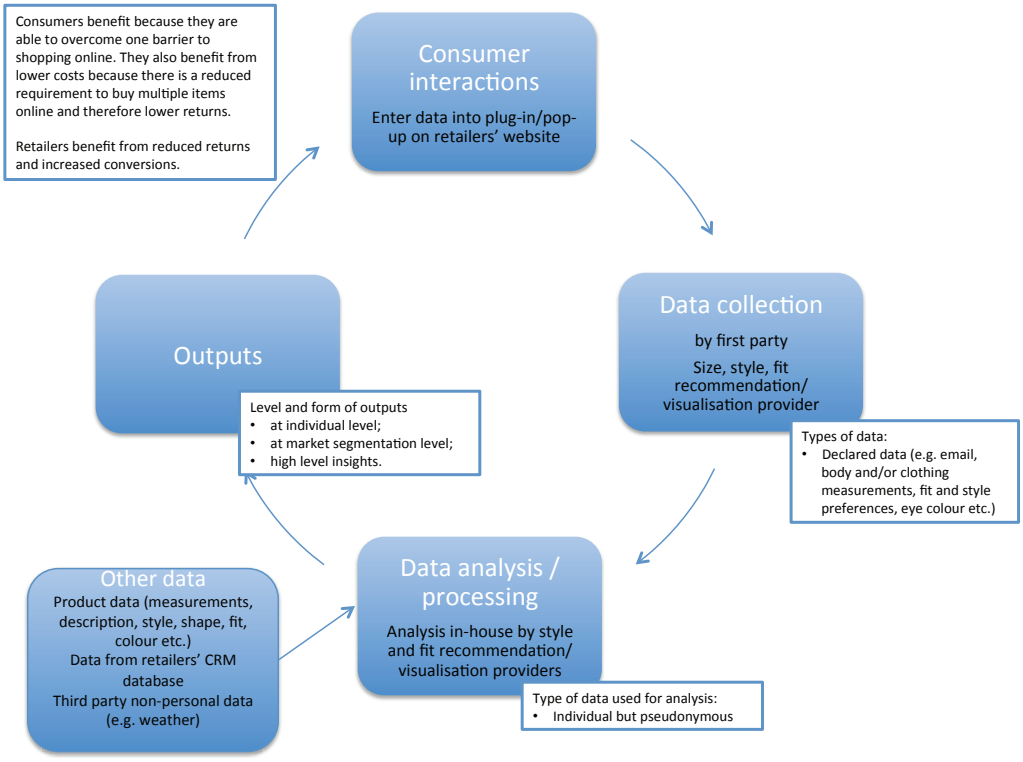
[219] Additionally, some of these providers also have a consumer-facing website of their own, where customers can create a profile and provide more detailed information which is then used to generate recommendations on clothes from a range of retailers. For the size, style and fit recommendation providers we spoke to during the interviews, at present this is not as well a developed proposition as the services provided directly on retailer's websites, but represents an interesting area for further development.

[220] For example, A/B testing can be a test in which one group of customers are the control group (group A for example) and are not presented with recommendations whilst the second group (B) are presented with recommendations. The key metrics such as conversion rates, average order value, returns etc. are measured for both of these groups and compared. If group B has a higher conversion rate then the recommendations are said to produce an uplift in conversion rates.

filtering techniques may also be used to support recommendation. For example, if consumers A and B have both bought brand X in the same size successfully, and consumer A has also bought brand Y in a given size successfully, then consumer B can be offered that size if they express interest in brand Y.

As such there is a large amount of data required and flows of data between different parties to be able to provide recommendations to customers. Figure 11 below illustrates the types of data collected and the data flows between the various stakeholders in the data value chain. In this case, the first party is the specialist service provider as they are the party collecting the data from the customer.

Figure 11: Data flows associated with size, style, fit recommendation and visualisation services



Source: Analysys Mason and DotEcon

The consumer benefits from such services because they help overcome barriers to shopping online, reduce the cost of returns (although small to the consumer) and reduce the requirement for the consumer to buy multiple sizes of the same garment. These benefits ultimately lead to lower return rates, which have a significant cost impact on online retailers. Furthermore, the use of visualisation engines can also increase conversion rates as consumers are better engaged with the online shopping channel.

The box below provides some specific examples of reported improvements retailers have achieved when using these services.

Box 23: Specific examples of success rates as reported by service providers

Metail reports that by using its personal fit services, its clients can increase conversions (up to 12% higher); reduce returns (11% lower); increase Average Order Value by 6%; increase engagement (76% adoption rate).[221]

Retailers using **fits.me** have also reportedly seen improvements in conversions and reduced returns. For example, Hawes and Curtis report an increase of 57% in conversions from first-time buyers who use the virtual fitting room and return rates down by 35%; Thomas Pink saw conversions improve by 21% when persuading people to check the fit. Pretty Green saw a remarkable 77% reduction in fit related returns.[222]

Dressipi quotes that retailers using its services can expect to see a 5% increase in conversion, 30% increase in Average order value and 10% reduction in returns.[223]

Payment-by-performance

We understand that there are a number of different models used to charge retailers for the service being provided, which depend on the preferences of the retailer, and one option is pricing on a performance-based model. Performance-related payments are linked to improvements in key performance indicators and detailed A/B testing is employed to determine the success of the recommendation service. However, we understand that in some cases retailers prefer a flat fee in order to better control costs due to cost control etc.) This fee will be based on the number of garments that the size and fit retailer will need to process. The fixed fee model will often be employed by the visualisation services as there are significant upfront costs associated with photographing the garments and adding the information to their systems. One online fitting room told us that it costs about £10 to process each garment, although they are hoping to bring this down to £5.

Consumers are informed of data collection and use in third-party privacy policies and some retailers' policies

It is the privacy/cookies policy of the service providers that describes the data collected, differentiating between data volunteered and that automatically collected through cookies. Information is also provided on the use of this data, whom it is shared with and how the customer can see the data that is held on them. However, the level of detail included in the retailer's own privacy policy varies, with some retailers only mentioning the use of cookies by third parties or including only a small print link direct to the third-party service provider's privacy policy. In other cases, the service provider's privacy policy is included in full or referred to in the retailer's privacy policy. Therefore, coupled with the fact that these services are not always clearly branded as third-party services,

[221] See http://metail.com/retailers/

[222] See fits.me case studies : http://fits.me/resources/case-studies/

[223] See http://partners.dressipi.com/index.html

consumers may not be fully aware of the extent to which their data is being collected, used and shared with these parties.

We understand that in most cases, data collected by the service provider about individuals is not shared with its clients (retailers) and/or other third parties where this would mean the individual was identifiable. Likewise, no information collected about customers for one retailer is shared with other retailers. Instead, anonymised and aggregated data may be shared to provide customer insights related to fit, style and purchase preferences of customers might be provided to retailers and their manufacturers.

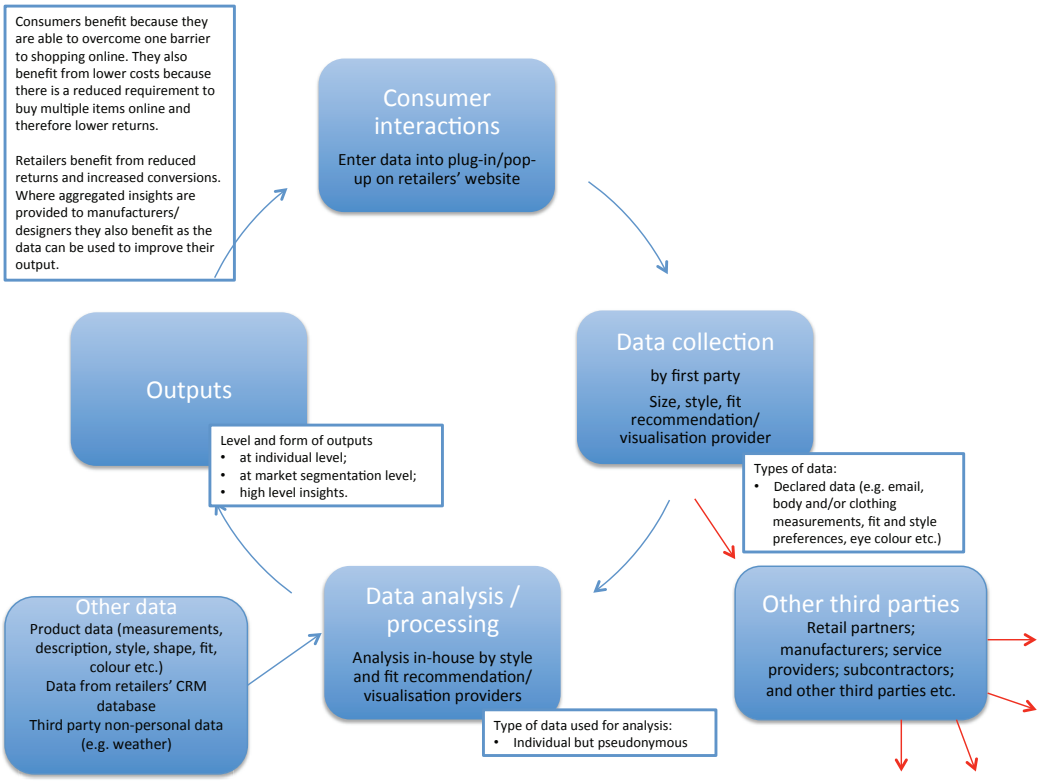
However, one fit provider's privacy policy states that whilst it does not sell customer data to third parties it may share a copy of the customer's profile, which may include the customer's email address (email address is only provided if the customer wants to save their profile). In this case the customer's profile information may be shared with: retail partners; service providers, subcontractors and agents who perform services on the fit providers behalf; manufacturers; and "*to third parties in certain business relationships, such as, but not limited to, co-branded offerings … If you provide information in connection with a co-branded service or feature, that information may be shared between True Fit and the third party…the third party's treatment of your information will be subject to the third party's privacy policy*."[224] Creating a profile with the fit provider may also mean that consumers receive marketing and advertisements from the fit provider or third parties.[225]

Figure 12 below illustrates the case where data collected by the fit provider may be shared with other third parties.

[224] See True Fit's privacy policy : http://www.truefit.com/privacy-policy

[225] See True Fit's privacy policy : http://www.truefit.com/privacy-policy

Figure 12: Data flows associated with size, style, fit recommendation and visualisation services including where data leaves the core system



Source: Analysys Mason and DotEcon

3.4.3 Use of data to inform business decisions

High-level insights to inform business strategy

Along with all the information collected by retailers about their customers and their behaviour, many third-party service providers also offer to provide the retailer with high-level insights about customers. For example, one specialist service provider of style and fit recommendations for online retailers told us that it provides its clients with aggregated and anonymised statistics about customers and their buying behaviour which can help retailer identify why certain garments sell well and why others fail to sell. Similarly, third-party aggregators such as Lyst, Mallzee and Grabble also offer partner retailers aggregated and anonymous data about customers and their behaviour to provide indications of what customers are buying and when. For example: red dresses have twice the return rate on mobile; clothes bought on a Monday have the highest return rate; and consumers in a specific region e.g. London that buy brand Y may also buy another brand X, whilst consumers in another region e.g. Manchester that buy brand Y also buy brand Z, but not

X. [226] Therefore, retailers have a rich set of data they can use to influence their business decisions.

We spoke to a large multi-brand retailer with online and offline stores who said that the data collected can help to improve its understanding of common characteristics amongst its customers as well as what products or brands are popular or underperforming. This data may then be used to inform strategic business decisions including re-stocking decisions and space planning in store (i.e. how much space to allocate a particular brand in store).

Geographic segmentation data to inform regional business decisions

One data broker informed us that it can use geographically segmented data to assist retailers with key business decisions. The data broker combines its vast amount of generic data (e.g. Census data, which is openly available to all at various levels of geography such as Output Area, Postal Sector, Local Authority District) with its own segmentation data on a geographical basis. The output of the analysis is provided to the retailer on an aggregate and anonymous basis. Retailer can then look at catchments around their outlets to determine the demographic characteristics of the area which helps, for example, to provide the retailer with insight on existing store performance, merchandise planning, store segmentation and local area marketing. It also helps retail clients to assess opportunities for opening new stores in areas that provide the right kind of customer base for their particular offering or service.

In some cases, the insight provided by the data broker can be enhanced if it has access to information about the client's own customer base, including a customer's location, declared data and observed data on spending patterns. This can be used to provide the retailer with insights about the particular demographics they are pulling from within the catchment area and where key concentrations of customers are geographically. However, the data broker was clear that they can only do this where the relevant permissions are in place and where customers are aware that data may be used for these purposes. Where the retailer provides data, this may be in one of three forms:

- Personally identifiable information – including an address and postcode (name not required) so that the data broker can enrich the data with its own segment database and demographics at the household level. The infomediary can anonymise the data and aggregate it to whatever geographical level is appropriate for that client for analytical purposes. The original personal data is then purged, as

[226] Wired Retail article, 24 November 2014, 'Half our 70 staff are data scientists' reveals Lyst cofounder', http://www.wired.co.uk/news/archive/2014-11/24/chris-morton-lyst

once the infomediary has aggregated it by geography it has no further need for the personal data.

- Postcode Level - the retailer might just pass the infomediary a customer postcode with a set of attributes/spends (on average a postcode defines 15 households). The data broker would then use the postcode to aggregate the information to higher geographies.
- Higher Geographic Aggregations - the retailer might aggregate the data to the required geography themselves and only provide the infomediary with a file at, for example, Output Area/Postal Sector level. In this case it is already aggregated and anonymised when it gets to the infomediary.

Brand perception and customer feedback

Data about customer perception of a brand, or feedback about a particular marketing strategy can also be used to inform business decisions. For example, one provider of social listening services informed us that the data it provides to clients could be used for a variety of purposes including:

- gauging public reaction to marketing campaigns, and identifying if the general sentiment is positive or negative. This would allow the brand to assess the success of their campaign or make changes to their marketing approach to ensure that they were reaching the correct target audience;
- realising very early on if there were any PR 'issues' allowing the brand to react as early as possible;
- deal with consumer complaints before they make it through to consumer call centres. We were told that this not only saves costs associated with call centres, but also ensures that disgruntled customers who would not normally make a formal complaint can be approached and the situation dealt with to try and maintain their custom and prevent too much negative chatter on social media for example;
- using geolocation data attached to comments made about the brand online to see if these comments (positive and negative) could be linked to a particular retail store for example.

More importantly, these methods can be used to assist with supply chain management where it can be identified that people are complaining about lack of stock of particular items in a certain geographic area, or for identifying emerging and popular trends to

integrate these into the demand forecasting process to stock inventories more effectively.[227]

Manufacturers of clothing can also benefit from consumer data collected on retailers' websites. For example, we found evidence that one size and fit recommendation provider, discussed in Section 3.4.2 above, can share its consumer data with partner retailers and their manufacturers to "*provide insights relating to the fit, style and purchase preferences relating to their products*".[228] Clearly this information has the potential to influence the design and production decisions of manufacturers. However, as discussed below, despite the perceived value associated with this, many consider there is greater scope for use of this data in future.

3.4.4 Future uses of data in clothing retail

One provider of style and fit recommendations told us that they believe there is a far greater potential for the use of data in the clothing retail sector and the fashion industry more generally and they felt that today there is a very low level of sophistication with regards to data use, particularly in the design stage. Designs are often based on inspiration and a drawing. While creativity is a very important part of the garment design process, it could be enhanced through data so that the fashion ecosystem can mitigate wastage (e.g. high level of returns and unsold inventory). A key question going forward is to what extent today's ready-to-wear industry can be replaced with more personalised components in some segments; infomediaries with a good knowledge of consumers' tastes could play a central role in that world.

Another provider of virtual fitting room and visualisation services shared similar views. They considered that the data provided by manufacturers on their clothing is often inadequate due to variability in the manufacturing process. However, they believed that data flows in the other direction i.e. measurements from the customer to the manufacturer could be used to support customised or made to order items.

The increasing provision of in-store Wi-Fi across retail stores and shopping centres could also support the collection of additional

[227] For example see http://www.logistics-gs.uni-bremen.de/fileadmin/Upload/StudentenPDFs/Flyer_Samaneh_Beheshti-Kashi.pdf

[228] True Fit's privacy policy states that they *"may share a copy of your True Fit profile, which profile information may include your email address (if provided) with: … (iii) manufacturers of products sold on our Partner Sites to provide insights relating to the fit, style and purchase preferences relating to their products…"* See http://www.truefit.com/privacy-policy

data.[229] One provider of in-store Wi-Fi explains that retailers could use their Wi-Fi to collect data on "*new versus returning users, the number of users logged onto the Wi-Fi service, in-store footfall and customer behaviour, mobile browsing behaviour and user demographics...Using tools such as in-store location-based marketing and mobile messaging, retailers can use this data to deliver targeted and personalised real-time communications to customers.*" [230] For example, we understand that multiproduct retailer ASDA is currently using this data to "*promote multichannel offers and take advantage of other retail innovation opportunities*".[231] In some cases the Wi-Fi provider may also use the data collected.[232]

Whilst in-store Wi-Fi is available in a number of stores, not all retailers are currently using this as a means for collecting customer data. For example, one large multi-brand retailer informed us that whilst they are investigating whether they could collect useful data from customer devices connected to their in store Wi-Fi, they are currently not doing so. However, the retailer confirmed that it could use this data to track mobile devices to better understand footfall by time and store and therefore get a better understanding of the busiest trading times and also help to plan the layout of the store more effectively.

There are some questions about how customers would be informed about any tracking activity that may occur, however, we understand that Wi-Fi providers such as EE, place strict requirements on retailers in terms of the way in which they can use the user data collected via the Wi-Fi and the information that is given to users regarding the collection and use of this data. For example, EE only permits retailers to use the user data in a way that is "*expressly permitted by the Wi-Fi Privacy Policy and User Consents*" and the Privacy Policy must be included on the Wi-Fi landing page (which a user must accept before it can use the Wi-Fi services). Furthermore, users

[229] Examples of retailers in the UK which already have or are looking to introduce in-store WiFi include: Argos, ASDA, Debenhams, John Lewis, Ted Baker, many shopping centres including Westfield, Kingsgate shopping centre etc.

[230] Connected-Retail EE, http://ee.co.uk/business/large/why-ee/total-enterprise-mobility/connected-retail

[231] Ibid.

[232] For example, *"The Customer [retailer] acknowledges and agrees that EE may use the User Data for its own commercial purposes including without limitation for marketing EE products and services to Users, for analytics purposes and to provide aggregated and anonymised data about WiFi usage to third parties."* See EE, 'PUBLIC WIFI FROM EE SOLUTION TERMS', Section 6 'Data protection', p. 5 https://ee.co.uk/content/dam/ee-help/files.ee.co.uk/business/terms-and-conditions/solutions/b2blegal3571%20Public%20WiFi%20from%20EE%20solution%20terms%20v1.3%2019.01.15.pdf

must be given the ability to opt out of marketing communications each time it receives such marketing.[233]

[233] See EE 'PUBLIC WIFI FROM EE SOLUTION TERMS' section 5 'Landing Page' and 6 'Data protection', https://ee.co.uk/content/dam/ee-help/files.ee.co.uk/business/terms-and-conditions/solutions/b2blegal3571%20Public%20WiFi%20from%20EE%20solution%20terms%20v1.3%2019.01.15.pdf

4 The games apps sector

4.1 Summary

Use of data and potential value

Games applications are published for use on mobile devices (smartphones and tablets) or through social network websites, such as Facebook. In both cases, games application developers collect and process a large amount of data about the activity of users in their games.

These developers earn revenue from their games in a number of different ways:

- charging for initial download of their games;
- advertising within their games, where adverts ('ads') display either content provided by third parties such as ad networks or other games within a developer's own games portfolio;
- offering an ad-supported free version of the game and making an ad-free version available at a cost; and
- making games available to download for free and generate revenues through user purchases made within the games ('in-app purchases').

This final model of monetisation[234], the 'freemium' model, is the predominant model of revenue generation for games apps. While this model is sometimes used in combination with other monetisation methods (for example, generating revenue through both adverts and in-app purchases in a single game), a substantial share of revenues from these types of games results from in-app purchases made by games users. For example, 90% of the revenue generated from 'games' category apps on one of the two main app stores, the Apple App Store, is generated using the freemium model.[235] Examples of popular in-app purchases include buying extra lives or playing time (for example, in *Candy Crush*), virtual currency or goods that can be used within a game, or customisation features within the game (for example, avatars within *Crossy Roads*).

In line with this main method of monetisation, the predominant use of consumer data amongst game developers, whether accessed through mobile apps or social media, is to drive in-app purchases and revenues by optimising the gaming experience. This

[234] Throughout the report, monetisation refers to the means by which app publishers earn revenue from their apps.

[235] Distimo, December 2013, '2013 Year in Review', http://www.distimo.com/publications

optimisation focuses on all the main revenue drivers: number of users, time spent in the game ('engagement'), length of period using the game ('retention') and in-app purchasing behaviour.

Consumer data is also used for targeted advertising purposes, but this often appears to be of secondary importance to developers, as benefits associated with encouraging use and driving in-app purchases are perceived to be greater that those related to the use of ads. Further, it appears that the methods of targeting advertisements to games users are quite basic.

How is data collected?

Games developers harvest data from games apps on mobile devices and from social network games. For example, this typically includes collection of behavioural data linked to a host of 'events' occurring within a game (e.g. where the game is being played, how long a user plays for, high scores, whether in-app purchases are being made[236] and, if so, at what points in the game).

In the case of games played through mobile applications, games developers collect data that is pseudonymous, where the 'identifier' it has for a user, linking that user to specific game activity, is a device number, advertising identifier or an IP address as opposed to a known individual. Games developers also have access to analytics tools, allowing them to augment data collected with aggregate data on their users such as gender, age range of players and in which countries the game is popular. Such tools include 'Google Analytics' for Android apps[237], 'Flurry' (part of Yahoo) used mostly by iOS developers and 'Facebook Analytics for Apps' for apps used on iOS, Android and Facebook.

In addition to pseudonymous and aggregate information, games played through social network sites also result in the collection of information linked to an individual's social network profile. In some cases, developers of mobile apps also collect some personal data if the user logs in with social network credentials where this is supported, or if users register directly with the developer. Users may choose to log in with these services so they can play a single game while switching between devices.

For the developers we interviewed, most undertook analysis of pseudonymous or aggregate data for game improvement purposes

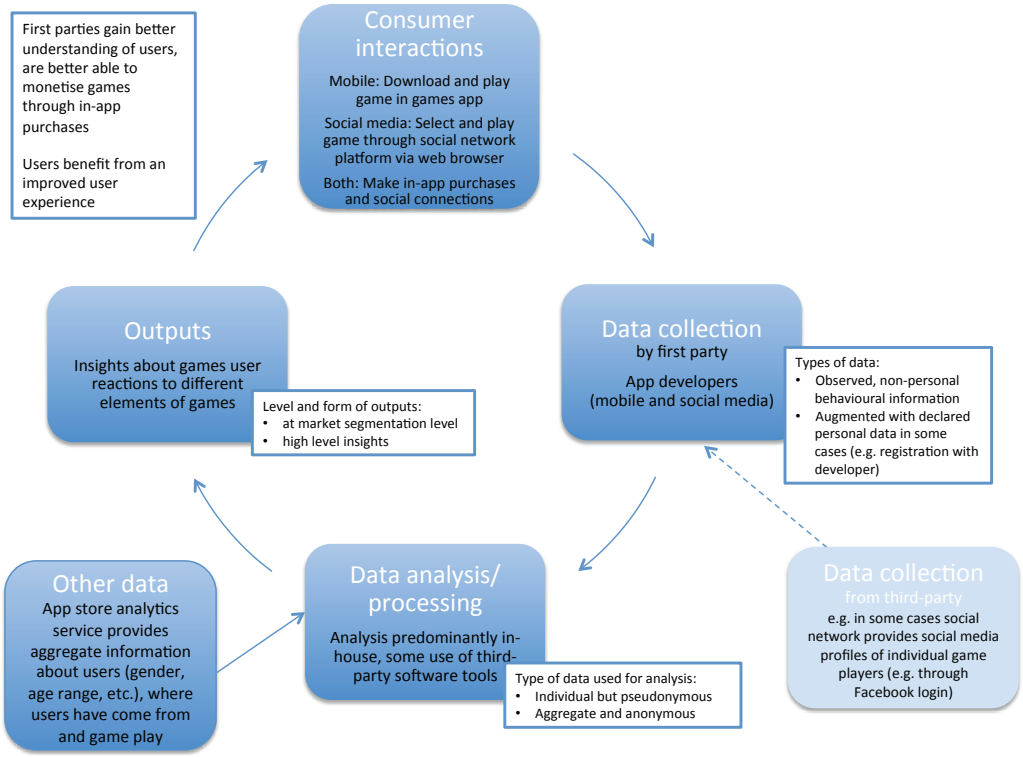
[236] We note that where in-game purchases are made the developer will know through the collection of behavioural data. However, payment data (such as credit card information) is not provided directly to the app developer. This is controlled by the platform (that is, from the app store the app was downloaded from (e.g. Apple's App Store or Google's Play store) and or the social media platform the game was played on (e.g. Facebook)).

[237] Google, 'Link Google Analytics and Google Play', https://support.google.com/analytics/answer/2956981?hl=en

in-house, although third-party analytics tools are available to facilitate the analysis.

The processes by which the games apps collect and use data are summarised in Figure 13 below.

Figure 13: Data flows involved in the collection and use of consumer data in the games apps sector



Source: DotEcon and Analysys Mason

Much of the focus of games developers is on the analysis of game users' behaviour in order to increase their engagement and retention. Understanding how to create a compelling game for users – which may drive both take-up and in-app purchases – is a key source of competitive advantage for developers, who are likely to keep this information confidential.

Games developers we interviewed considered that pseudonymous usage and demographic data, enriched with high-level insights into their overall user base, were sufficient for the purposes of analysis of user behaviour. Indeed, one games developer told us that while they receive personal data from Facebook relating to users accessing their games through this platform, they do not use this, as it is not considered to improve the value of their user analysis.

4.2 Sector overview

In this section we present a review of the games apps sector, including an overview of the scale of the sector, the firms involved and the competitive dynamics along with any recent trends.

An application or 'app' is a self-contained software program designed to fulfil a particular purpose or enable a user to perform a task. Games apps are "*apps that provide single or multiplayer interactive activities of skill or chance for entertainment purposes*".[238] These include action, educational, family, puzzle, word and adventure games.

This case study focuses on data collection and use in games applications, including games that can be accessed through smartphone or tables ('mobile games apps') and those accessed and played over social network sites such as Facebook ('social network games apps'). We have purposely excluded gambling apps from this case study and concerns relevant only to children's games apps have also been omitted.[239]

Games apps can be accessed on a smartphone or tablet…

A mobile app is an app that can be accessed and used on a portable device such as a smartphone or a tablet. To obtain mobile games apps, a would-be user would need to enter an app store (through an application on its device) and then select, download and in some cases pay for the app. Once the app is installed, the game can then be accessed directly on the device's main screen.

Some developers may choose to include a social network login feature in their app. This provides users with an option to log in to an app with their social network credentials. This may add functionality to the app, for example, allowing social interaction, sharing progress on a social network site or seamless gameplay across devices.

…or through a social network.

A social network games app is installed and used through a social network, accessed through a web browser. These can be accessed through a store or interface on the social network site, for example, users can find games on Facebook's 'App Center'. Social network games are primarily played on Facebook; we note that around 53%

[238] Apple, 'Developer Guidelines: App Store Product Page', https://developer.apple.com/app-store/product-page/

[239] This has been covered by the OFT report and consultation for their principles for children's games, 30th January 2014, 'The OFT's Principles for online and app-based games', https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/288360/oft1519.pdf

of UK adults are Facebook users[240] and that, reportedly, up to one quarter of Facebook users may be playing games regularly on the platform.[241] In 2010, 83% of social gamers were playing on the Facebook platform,[242] and few other social networks offer games to users.[243] Indeed, one major developer we interviewed said that, in Europe, Facebook was the only social network platform it uses to publish its social network games. In this case study, when discussing games apps played through a social network site, we focus primarily on Facebook and its role in the consumer data chain.

In addition to differences in how games are accessed, there can be differences in the features of a game accessed as a mobile app and those accessed through a social network. For example, social network games apps often have features that differentiate them from mobile games apps (which do not have a social network login), such as in-game social connections, real-time game playing with friends and gifting virtual currency.[244] Therefore, despite their superficial similarities, we will need to draw a distinction in certain cases between mobile gaming apps and social gaming apps, especially as the latter have the ability to connect with data held in social networks.

4.2.1 Scale of the sector

Global app market

Since the launch of the App Store by Apple with 500 apps in 2008, the sector's growth has transformed the mobile industry. The increasing take-up of smartphones and tablets has enabled the emergence of a new market for the development and sale of apps

[240] Ofcom reported in 2014 that 83% of adults 'go online', of which 66% have a social network profile, of which 96% have a Facebook profile. Ofcom, 29th April 2014, 'Adults' Media Use and Attitudes Report 2014' http://stakeholders.ofcom.org.uk/market-data-research/other/research-publications/adults/adults-media-lit-14/

[241] Mashable, 20th March 2014, 'Facebook: 375 Million Users Play Games Each Month', http://mashable.com/2014/03/19/facebook-games-stats/

[242] At the time, 24% of social gamers played on MySpace, 7% played on Bebo (which no longer exists) and 5% played on Friendster. See Popcap Games, 2010, '2010 Social Gaming Research – Information Solutions Group', http://www.infosolutionsgroup.com/2010_PopCap_Social_Gaming_Research_Results.pdf

[243] According to GlobalWebIndex the top 8 social networks in the UK are Facebook, Youtube, Twitter, Google+, LinkedIn, Pinterest, Instagram and Tumblr, http://wearesocial.net/blog/2015/02/uks-top-social-networks/

[244] GCO Games Convention Online 2010, S Bjork, 2010, 'Principles and patterns of social games: Where's the difference compared to other games?', http://www.slideshare.net/staffanb/principles-and-patterns-of-social-games

that provide a wide range of services for users. In June 2014, the Apple App Store and Google Play were the two most popular major mobile app stores with both having around 1.2 million active apps available for download[245] and the Apple App Store reporting cumulative downloads of 85 billion since its launch.[246]

Global mobile app revenue reached over $20 billion in 2013, with predicted revenue reaching $70 billion by 2017.[247] Apple recently reported evidence of this continued growth, stating that Apple App Store service revenue totalled $5 billion for January to March 2015, equivalent to $20 billion annually for Apple's platform alone.[248]

Games apps are by far the most popular app category, both in terms of number of downloads and revenue, on both Google Play and the Apple App Store.[249] In 2013, games generated 80% of app store revenue,[250] and all of the top 10 revenue-generating apps in the Apple App Store were games.[251] The mobile games apps sector is expected to become the most valuable component of the wider global videogame industry in 2015, overtaking games for

[245] See http://www.statista.com/statistics/276623/number-of-apps-available-in-leading-app-stores/

[246] Tech Crunch, 20 October, 2014, 'App Store Downloads Top 85 Billion, Revenue Up 36 Percent Year-Over-Year', http://techcrunch.com/2014/10/20/app-store-downloads-top-85-billion/

[247] Digi-Capital, April 2014, 'Mobile apps to hit >$70B revenue driven by explosion of diversity', http://www.digi-capital.com/news/2014/04/mobile-apps-to-hit-70b-revenue-driven-by-explosion-of-diversity/#.VTjCV4upr8s%20with%20different%20estimates

[248] Nasdaq, 27 April 2015, 'Q1 2015 Results Earnings Conference Call', See http://www.nasdaq.com/aspx/call-transcript.aspx?StoryId=3107596&Title=apple-s-aapl-ceo-tim-cook-on-q2-2015-results-earnings-call-transcript

[249] App Annie, 29th January 2015, 'App Annie Index: 2014 Retrospective', page 36 and 38, http://filearchive.cnews.ru/img/cnews/2015/01/29/app_annie_index_2014_retrospective_en.PDF

[250] Newzoo, July 2014, 'Newzoo Global Games Market Webinar Presentation', slide 9, http://www.newzoo.com/keynotes/newzoo-global-games-market-webinar-presentation/

[251] Distimo, December 2013, '2013 Year in Review', http://www.distimo.com/publications

consoles;[252] it represents over a quarter of the global videogame industry by revenue and is the fastest growing component.[253]

Focus on games apps in the UK

Ofcom statistics show that the percentage of adults playing games on mobile phones and on smartphones has been increasing rapidly in recent years.[254] In 2014, the value of the UK mobile games apps sector was estimated at £548m in consumer spending, making up a substantial part of the wider UK videogames market (valued at £3.9bn).[255]

Robust statistics for social network games are not available, but we note that a large number of UK games companies are involved in the development of social games.[256] Social games refer to those played on social networks, but may also cover some mobile games apps where the user can use a social network login (such as Facebook Login) to interact with other users.

In 2013, 19% of UK consumers played games on smartphones and 10% played games on tablets.[257] Games apps appeal to a relatively broad demographic; the growth in downloads of mobile games in the UK has been driven by consumers in the 25-34, 45-54 and 55-64 age ranges.[258] In 2011, the average age of a social network gamer in the UK was around 35.5 years old and 58% of social network gamers were women.[259]

Gamers not only download many games but also tend to play them frequently; UK gamers using Android devices spend an average of

[252] GamesIndustry.biz, 22nd October, 2014, 'Mobile to become gaming's biggest market by 2015', http://www.gamesindustry.biz/articles/2014-10-22-report-mobile-to-become-gamings-biggest-market-by-2015

[253] Newzoo, 15th May 2014, 'Global Games Market will Reach $102.9 Billion in 2017' http://www.newzoo.com/insights/global-games-market-will-reach-102-9-billion-2017-2/

[254] Ofcom, 2014, 'Adults' Media Use and Attitudes Report 2014', Figure 7 http://stakeholders.ofcom.org.uk/binaries/research/media-literacy/adults-2014/2014_Adults_report.pdf

[255] Ukie, 'The games industry in numbers', http://ukie.org.uk/research

[256] TIGA, in Ukie, 'UK video games fact sheet', http://ukie.org.uk/sites/default/files/cms/UK%20Games%20Industry%20Fact%20Sheet%2014%20April%202015_0.pdf

[257] Mashable, 20th March 2014, 'Facebook: 375 Million Users Play Games Each Month', http://mashable.com/2014/03/19/facebook-games-stats/

[258] Ibid.

[259] Information Solutions Group, 2011, '2011 Social gaming Research', http://www.infosolutionsgroup.com/pdfs/2011_PopCap_Social_Gaming_Research_Results.pdf

32.4 minutes daily playing games[260] and around 67% of social network gamers play social network games at least once a day.[261]

Future growth for mobile games apps is likely, but it might slow somewhat

There is a general expectation that the games apps sector will continue growing globally, with Newzoo estimating that smartphones and tablet use will drive a compound annual growth rate of over 20% over the next few years.[262] However, there are also indications that this trend may be beginning to plateau in the UK. It appears that the frequency of downloads among UK app users is decreasing over time, though to an extent this might arguably reflect increasing quality of apps, which would in turn reduce the need for users to download additional apps frequently.[263] This is putting pressure on the mobile apps sector,[264] and the task of gaining a share of the decreasing number of downloads whilst generating revenue effectively may become increasingly difficult. Gartner - an IT research company - has estimated that in 2018 less than 0.01% of consumer mobile apps will be considered a financial success by their developers.[265]

Social network game growth may be held back by increasing use of mobile devices

Increasing smartphone and tablet use may have an impact on social network games, as these are typically played on desktop or laptop computers. Reflecting this, we note that Facebook's Payments revenue (mostly generated by games) has increased markedly from $557m in 2011 to $886m in 2013[266], but Facebook has since noted that this trend appears to be reversing, predicting that use of social network games may decline in line with the decline in use of

[260] The Guardian, 19th September 2014, 'Android games players averaging 37 minutes a day, with Americans keenest', http://www.theguardian.com/technology/2014/sep/19/android-games-minutes-americans-flurry

[261] Information Solutions Group, 2011, '2011 Social gaming Research', http://www.infosolutionsgroup.com/pdfs/2011_PopCap_Social_Gaming_Research_Results.pdf

[262] Newzoo, 29th October 2014, 'Global Mobile Games, Revenues to Reach $25 Billion in 2014', http://www.newzoo.com/insights/global-mobile-games-revenues-top-25-billion-2014/

[263] Financial Times, 17th August, 2014, 'Smartphone owners' appetite for new apps wanes', http://www.ft.com/cms/s/0/4c3e5708-2628-11e4-9bca-00144feabdc0.html#axzz3TQQtXeZz

[264] Financial Times, 15th August, 2014, 'Valuations grow for apps that offer less and less', http://www.ft.com/cms/s/0/f178898e-2482-11e4-ae78-00144feabdc0.html

[265] Techcrunch, 13th January 2014, 'App monetization to get tougher still, with Gartner predicting 94.5% of downloads will be free by 2017', http://techcrunch.com/2014/01/13/making-apps-pay-gets-harder/

[266] Facebook, 'Annual Report 2013', http://files.shareholder.com/downloads/AMDA-NJ5DZ/69183990x0x741493/EDBA9462-3E5E-4711-B0B4-1DFE9B541222/FB_AR_33501_FINAL.pdf

computers in preference to mobile devices. In our interviews, one developer stated their games are more popular on mobile devices than through Facebook while another developer expressed the view that fewer customers were playing games over Facebook and they were looking at new ways to encourage social gameplay.

4.2.2 Firms participating in the sector

The games apps sector is complex with a range of firms playing different roles, including app developers, app publishers, app stores, mobile operating systems and other third-party service providers.

Games app developers v game publishers

Games app *developers* create the game software, whether or not they then publish the game on an app store or social network. Games app *publishers* are those firms that publish the apps and may accompany this with marketing or promotional activities, whether or not they developed software themselves. Whilst developing software is a distinct activity from publishing and promoting a game, most large game publishers also have internal games app development teams or own game development companies, and app developers may easily be able to publish themselves (unlike, for example, console game developers). Therefore, many games app developers are also game publishers and vice versa. Given that these parties are typically vertically integrated, for simplicity in this report we refer mainly to games developers.

App developers v more general games developers

It is also possible to distinguish between firms that develop and/or publish only mobile or social network games apps, and those firms that have historically been involved in the videogames sector (e.g. making games for PCs or consoles) and have moved into the development of mobile and social network games. For example, some app developers such as King solely produce games apps, whereas long-established games firms such as Electronic Arts have expanded into the mobile sector, with a subsidiary specialising in mobile games apps (EA Mobile). App developers may also offer certain games on a cross-platform basis, i.e. offering a mobile version as well as versions that can be played on PCs or consoles (though functionality might differ).

In contrast to the console games sector, where these require agreements with the console-maker, manufacture of a physical product and distribution to retailers, it is relatively easy for a small developer to enter the sector and make a games app widely available to consumers. Indeed, this has significantly influenced the current composition of the games apps sector, which has a few large firms existing alongside a long tail of small developers.

The most popular games app developers in the UK in terms of downloads include both established cross-platform firms and mobile app specialists.[267] Many social network game developers[268] also publish across platforms, providing their games over both social networks and mobile devices. Facebook is the primary social network platform on which users play games and thus the focus of our research.[269]

Operating systems

Operating systems for devices can offer the means for app developers to access user data. Operating systems define the capabilities that apps can have, in particular what data held on the underlying device, including personal data, the app can access through the Application Programming Interfaces (APIs)[270] it provides. In the case of social network games, the social network will also typically provide the APIs. These APIs can act as a constraint on the data developers can collect.

Google's Android and Apple's iOS are the most popular operating systems for developers with 71% of developers targeting Android and 54% targeting iOS. Windows is fast emerging as a third significant platform with 30% of app developers now developing apps for Windows.[271] Other platforms include Blackberry OS, Linux and Tizen. This is expanding with new platforms, such as wearable devices like the Pebble, which also has its own selection of apps.[272]

App stores as platforms

App stores act as intermediaries, setting the terms for app developers and publishers for their apps to be made available on app store platforms, and also distributing apps to those that wish to download (and in some cases pay for) them. As part of this

[267] Based on number of UK game downloads from the Apple App Store and Google Play in 2014, the top 10 publishers are: King, Electronic Arts, Gameloft, Rovio, Storm8, Ketchapp Studio, Disney, Zynga, XPEC and Glu. See page 31 of App Annie, 29th January 2015, 'App Annie Index: 2014 Retrospective', http://filearchive.cnews.ru/img/cnews/2015/01/29/app_annie_index_2014_retrospective_en.PDF

[268] The most popular developers of social network games include Zynga, MegaZebra, Wooga, 5 Minutes, Plinga, Playdom, Kabam and Crowdstar.

[269] Other social networks have attempted to create similar capabilities on their networks in the past, for example, Google+ Games. Hi5 has a sole game, called 'Pets', and users of the instant messaging service Tencent QQ are able to download QQ Games and play with their instant messenger contacts.

[270] An API dictates how software functions and interacts with the hardware of the device.

[271] Developer Economics, February 2015, 'Developer Economics Q1 2015: State of the Developer Nation', vmob.me/DE1Q15

[272] TechCrunch, 13th January, 2014, 'App Monetization to get tougher still, with Gartner predicting 94.5% od downloads will be free by 2017', http://techcrunch.com/2014/01/13/making-apps-pay-gets-harder/

intermediating role, an app store provides the marketplace and payment mechanisms (including for in-app purchases) needed for conducting transactions between developers and consumers.[273]

App stores also have a role in promoting best practice for developers and set guidelines and terms and conditions for the collection of personal data. In some cases there are review processes that aim to prevent breaches and avoid publishing apps that do not meet standards. For example, Facebook reviews apps to be published on their platform[274] and apps using Facebook login that request sensitive items of information[275].

In some cases app stores are vertically integrated with their respective operating system. For example, in the case of Apple, the app store is vertically integrated with the 'iOS' operating system, and for Android, Google Play is the vertically integrated app store. Whilst Apple app store is the only app store for Apple devices (unless they have been 'jail broken'[276]), Android users have a range of options over and above Google Play, including third-party app stores such as the Amazon Appstore, Archos Appslib and SlideME. There are also app stores for Blackberry devices (BlackBerry App World), Palm devices (Palm App Catalog), Nokia devices (Nokia OVi Store) and Windows phones (Windows Phone Store).[277]

More generally, Google, Apple and Windows are simultaneously operating system and device manufacturers, app store owners, app developers and app publishers.

Figure 14 below provides a simple illustration of the firms involved and how they interact:

- the app developer writes the code to create the game software;

[273] For example, transactions made on Apple apps are run through a commissionaire model, where purchase and transactions always take place via the relevant iTunes entity. For Europe this is iTunes S.A.R.L in Luxembourg. There is no transaction between consumer and developer (this applies to any in-app purchases).

[274] Facebook, 'Developer Guidelines: App Review', https://developers.facebook.com/docs/apps/review

[275] Facebook, 'Developer Guidelines: Login', https://developers.facebook.com/docs/facebook-login/permissions/v2.3

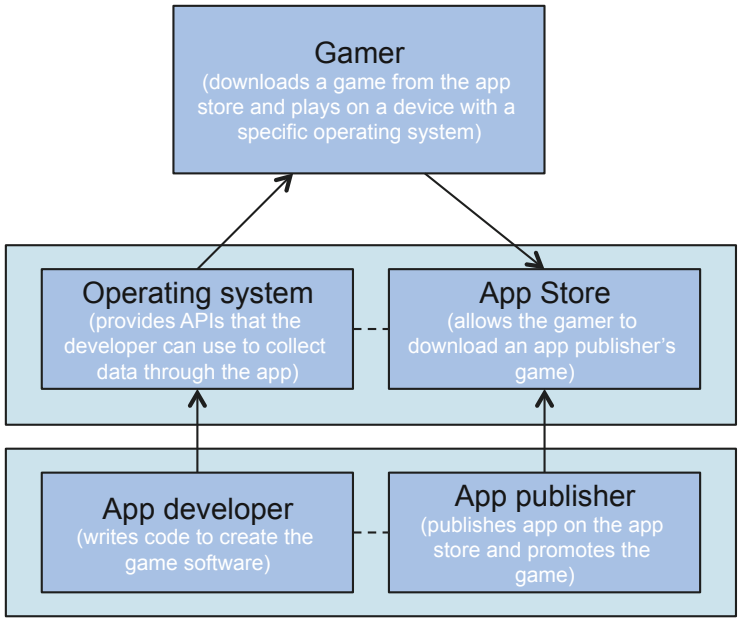
[276] Jailbreaking is the process of removing hardware restrictions on iOS devices.

[277] In July 2014, Google Play offered 1.3 million apps and Apple app store offered 1.2 million. This far exceeds the 300,000 offered over Windows Phone Store and the 240,000 offered over Amazon Appstore, the third and forth largest app stores respectively. See http://www.statista.com/statistics/276623/number-of-apps-available-in-leading-app-stores/

- the app publisher publishes the app onto an app store and promotes the game;
- the app store provides the means by which the gamer can download the game, carries out transactions and sets some terms for the app developer and publisher;
- the gamer will play the game on a device with a particular operating system; and
- the operating system (or social network platform) offers the means for app developers to access user data by providing APIs.

The developer and the publisher may be vertically integrated and the operating system and the app store may be vertically integrated.

Figure 14: Basic illustration of the firms involved in the games apps sector



Source: DotEcon and Analysys Mason

Third party services

In addition to these core parties in the games apps sector, some app developers also use third-party services.[278] These include app ad networks, user analytics and app store analytics services. In some cases, third-party developer tools (i.e. code to enact certain actions) are also available to help with social network integration, push notifications, and performance tools.[279]

[278] Developer Economics, February 2015, 'Developer Economics Q1 2015: State of the Developer Nation', vmob.me/DE1Q15

[279] Ibid.

Third-party app analytics and ad networks are the most popular third-party services provided to app developers.[280] A brief description of these services is provided in the boxes below. We discuss the involvement of these firms in more detail in the sub-sections below.

Box 24: Examples of third-party services used by app publishers and developers

Mobile ad networks

Ad networks provide ads from advertisers to apps when an app makes a request for one. These ads may be targeted; in such circumstances, the ad network will need to collect information from the app or information linked to the unique device identifier known as an advertising ID, to determine which advert to serve. Spending on mobile advertising reached $13.1billion in 2013 and is expected to expand (however, this figure include both in app and in browser advertising).[281]

The largest independent ad network for mobile devices, InMobi, revealed that around 40% of ad time in their network comes from gaming.[282] [283]

These statistics do not apply to social network games apps. Games published on Facebook may serve advertising through third-party ad networks that have agreed to Facebook's Terms and Conditions.[284] Information relating to the extent to which games developers use third-party advertising, however, is limited.

[280] Ibid.

[281] Gartner, 21st January, 2014, 'Gartner says mobile advertising will reach $18 billion in 2014', http://www.gartner.com/newsroom/id/2653121

[282] inMobi, 25th August, 2014, 'InMobi on how its native first emotionally lead approach can increase CPMs to $20', http://www.inmobi.com/company/news/inmobi-on-how-its-native-first-emotionally-lead-approach-can-increase-cpms/

[283] Other major mobile advertising networks include iAd, NativeX and Appsfire, YeahMobi, StartApp, AdColony, Millenial Media/Nexage.

[284] Facebook provides a list of ad providers that have agreed to Facebook's Terms and Conditions. See https://developers.facebook.com/docs/adproviders

App analytics providers

Mobile app analytics firms provide metrics for the various channels users come from, their activity within an app, how many users are monetised (that is, generate revenues) and some aggregated and anonymous demographic data about users.

While the majority of analytics consumers use Flurry, Umeng and Google Analytics, a large market of smaller firms exists.[285] Analytics usage is differentiated by platform, 47% of Android apps using analytics are supported by Google Analytics, whereas Flurry is the most popular analytics provided for iOS apps taking 52% of those using analytics. Facebook Analytics For Apps is also available to developers publishing games across iOS, Android and on Facebook.[286]

4.2.3 Key business models

App developers work with a wide range of business models. The main routes for monetisation include:

- in-app purchases;
- paid-for app downloads;
- third-party advertising;[287]
- cross-promotion of a developer's own profitable games; and
- corporate promotion, where brands release games to engage a target audience or promote a wider franchise.

The 'freemium' model with in-app purchases is a popular way of generating revenue in games apps

In-app purchases appear to be a significant driver of revenues in the games apps sector, based on the so-called 'freemium' model. The freemium model offers apps as free downloads, but allows developers to make money through in-app purchases. Examples of possible in-app purchases include buying extra lives or playing time (for example, in King's *Candy Crush*), virtual currency or goods that can be used in-game (for example, in EA Mobile's *The Sims*) or customisation features within the game (for example, avatars within Hipster Whale's *Crossy Roads*).

[285] Other major analytics firms include Localytics, inMobi Analytics, Upsight, Mixpanel, Testflight and Digital Analytix.

[286] Facebook, 'Developer Guidelines', https://developers.facebook.com/docs/analytics/overview

[287] In a 2015 survey by App Annie 70% of companies said they aimed to earn revenue through the app store (through downloads and in-app purchases), almost half aimed to earn revenue through in-app advertising. See App Annie and IDC, 31st March 2015, 'Mobile App Advertising and Monetization Trends', http://blog.appannie.com/mobile-app-advertising-and-monetization-trends-2013-2018/

The freemium model has grown to become the most prevalent with around 90% of the revenue from 'games' category apps on the Apple App Store generated using the freemium model.[288] In the UK, it has been reported that "*[t]he free apps with in-app purchase model is king, representing 76% of the revenue share of the UK for apps*".[289] Social network games are also generally available for free and follow the same freemium model.

However, some games require the user to pay to download

While the most popular apps are free to download, some app developers and publishers monetise apps by charging for downloads. However, in 2014 only one of the 30 top-grossing iOS apps was a paid-for app.[290] Developers may in some cases offer a single paid version of the app, or offer a free or 'lite' version of the app (that contains advertising and/or has limited features) and a superior, paid-for version. The latter business model, albeit relatively rare, has been used for some very successful apps, such as Rovio's *Angry Birds*, which can be downloaded as a 'lite' version with advertising (*Angry Birds Free*), or a paid-for version.[291] When free and paid for versions are both available, in-app purchases may still be used in both versions (as is the case for *Angry Birds*).

Third-party advertising in-app is also used, although not necessarily a major source of revenue for app developers

Adverts may be served in-app in the form of banner ads, video ads, full screen ads that come up during use of an app ('interstitials'), product placement or white label games (which contain virtual billboards that can be in-filled by advertisers) and native advertising. In the case of games played through Facebook, developers can also present adverts 'in-app' using ad providers that follow Facebook's advertising guidelines.[292]

In the UK 32% of mobile app revenue comes from in-app advertising.[293] However, we understand that for games apps, advertising tends to generate only a relatively small income

[288] Distomo, December 2013, '2013 Year in Review', http://www.distimo.com/publications

[289] Distomo/App Annie, in Ukie, 'UK video games fact sheet', http://ukie.org.uk/sites/default/files/cms/UK%20Games%20Industry%20Fact%20Sheet%2014%20April%202015_0.pdf

[290] Business Insider, 20th May 2014, 'The 15 Highest-Grossing iPhone And iPad Games', http://www.businessinsider.com/highest-grossing-iphone-and-ipad-games-2014-5

[291] TaylorWessing, May 2013, 'Making apps pay', http://www.taylorwessing.com/download/article_making_apps_pay.html#.VTYAAa1VhHx

[292] Facebook provides a list of ad providers that have agreed to Facebook's terms and conditions. See https://developers.facebook.com/docs/adproviders

[293] App Annie and IDC, 31st March 2015, 'Mobile App Advertising and Monetization Trends', http://blog.appannie.com/mobile-app-advertising-and-monetization-trends-2013-2018/

stream.[294] For example, one app developer we spoke to reported that while third-party advertising was used, this only represented about 10% of their total revenue stream.

A 2014 survey of developers indicated that, while advertising was more popular and cost-effective than paid-for download monetisation, the most popular and cost-effective model was freemium with in-app purchases.[295] King, a successful developer of both mobile and social games, decided in 2013 to eliminate all elements of advertising from all of its games, relying on in-app purchases to generate revenue.[296] Similarly, another app developer told us that a strategic decision was made to drop third-party advertising altogether and only advertise its own apps. As described in more detail in Section 4.4.1 below, this strategy is often popular as it drives the number of people accessing a developer's apps, creating a larger market for monetisation through in-app purchases.

For both social network and mobile games apps, monetisation may rely on a relatively small proportion of users who spend small amounts of money over a very large user base. For example, in relation to social games it has been estimated that only around 1-5% of players make in-app purchases; this was corroborated by one developer who stated that this was around 4% for their games in 2013. Within this small subgroup, it is estimated that just 15% of paying users account for around 50% of freemium revenues.[297]

4.3 Collection of consumer data

In this section, we focus on consumer data collection activities in the sector, considering:

- the type of information collected;
- the methods by which it is collected or sourced; and

[294] We note, however, that advertising may play a role in 'versioning', where it is used in a 'lite' version of the app in order to reduce user enjoyment and incentivise take-up of the paid version of the app.

[295] Gamasutra, 15th April 2014, 'In-app purchases really are most effective for mobile game monetization', http://www.gamasutra.com/view/news/215546/Inapp_purchases_really_are_most_effective_for_mobile_game_monetization.php

[296] Techcrunch, 12 June 2013, 'King Quits Advertising Since It Earns So Much On Candy Crush Purchases', http://techcrunch.com/2013/06/12/king-quits-advertising-since-it-earns-so-much-on-candy-crush-purchases/

[297] Gabling Commission, January 2015, 'Social Gaming', http://www.gamblingcommission.gov.uk/pdf/Social-gaming---January-2015.pdf

- the firms involved in this process.

Given the differences between mobile games apps and social network games apps in this area, we discuss each in turn.

4.3.1 Mobile apps

Types of data collected and firms involved

In principle, app developers can access and collect a large range of data from the device being used to run the app. For example, this may include:[298]

- device information including the device model, operating system and device identifiers, such as the Android_ID, to consistently identify app users. It is also possible to collect data about the identity of the device (for example the name of the phone – 'Joe Bloggs' iPhone'); [299]
- data stored on the device by the user including contacts, calendars, phone call logs, photos, SMS history, browsing history etc.;
- connection information including information from Wi-Fi, Bluetooth or NFC;
- data from the different sensors and components of a device including camera, microphone, location data (GPS and network-based);
- usage information such as browsing and behavioural data on gameplay, including access times, gameplay activity or interaction with other players.

Note that these different types of data may often be interlinked. For example, in addition to visual data, an image file may have additional data stored alongside it such as creation date and location within its EXIF data. App developers are encouraged to

[298] App permissions on Google Play, See https://support.google.com/googleplay/answer/6014972?hl=en-GB and Data Protection Working Party for the European Commission, 27th February 2013, 'Opinion 02/2013 on apps on smart devices, 00461/13/EN WP 202', http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2013/wp202_en.pdf

[299] Data Protection Working Party for the European Commission, 27th February 2013, 'Opinion 02/2013 on apps on smart devices, 00461/13/EN WP 202'.

[299] Others include IMEI (International Mobile Equipment Identity), IMSI (International Mobile Subscriber Identity) and MAC address (Media Access Control). Lookout mobile security, 'Mobile app advertising guidelines', https://www.lookout.com/resources/reports/mobile-ad-guidelines

strip out unnecessary metadata,[300] and data obtained may be stored, processed and transferred in its original form or may be 'hashed' to secure the data.

Although, in principle, the app developer could collect a vast amount of device data and personal data, in practice the amount of data collected by an app may be more restrained. As we describe further below, access to certain data is limited by the OS provider. App developers may also be incentivised to limit the amount of data they collect given the need to provide explicit notification to customers about the data that will be collected (explained in detail in the sub-section below).

Our research and interviews with app developers indicates that overall, the collection of consumer data from mobile games apps is relatively limited. The most common type of data collected is the data on gameplay and usage, linked to a unique ID such as the device identifier. Most data collected about the device (such as the device identifier) and in-game performance is collected directly by the app developer. This may include, for example, number of levels played in a session, number of 'fails' on a particular level, users playing one game mode more than another or preferences for different levels/characters. This is pseudonymous data linked to a unique user ID but is not personally identifiable.

However, in some cases, personal data may also be collected where there is an opportunity for the user to register directly with the app developer for competitions or when creating a user account to enable them to play the same game across different devices. For example, King allow its users to create an ID and will collect any "*information that you provide us with when you fill in forms when using our Services, or when you create an account with us*".[301] This may include personally identifiable data such as name, address, telephone number or email address.

Some app developers also provide users the option of logging in to the game with a social network login (e.g. Facebook or Google+ Login).[302] Box 25 below provides examples of the data that can be collected by developers when users login using 'Facebook Login'.

[300] Information Commissioner's Office, 'Privacy in mobile apps: Guidance for app developers', ico.org.uk/media/for-organisations/documents/1596/privacy-in-mobile-apps-dp-guidance.pdf

[301] King, 'Privacy Policy', https://king.com/#!/privacyPolicy

[302] Note that, for example, Google may use this data as part of its mobile ad serving technology to serve personalised ads (you can opt out) but the app developer will probably never be exposed to the underlying data. However, Facebook adopt a different policy.

Box 25: Examples of data that can be collected when there is a Facebook Login

Facebook Login

If a user logs in, the app developer can collect 'public profile' information which is provided by default when Facebook Login is used. 'Public Profile' information includes data such as the name of the user, their Facebook link, profile picture, gender, location and time zone. App developers can also collect data about 'App Friends' - a list of the user's contacts who also use the app. Profile information and 'App Friends' are the most common items of data collected by developers. App developers can also collect email address without a review by Facebook.

Beyond this, there are a wide variety of Permissions that a developer can request (where a Permission is a string passed with a login request or an API). Extended profile permissions may include information written in their 'About Me' section of their profile, their birthday, education history, events, games activity, hometown, likes, location, photos, posts, relationships, religion, politics, statuses and groups that they have joined.[303] However, app developers will no longer be able to access data about a user's activities and interests (based on their likes of activity and interest related pages) from the 23 June 2015.[304]

The app developers that we spoke to noted that they only use the minimum data necessary to allow users to interact with their friends in the app and do not use personal information for any other purpose.

Examples of the data collected by popular app developers through their apps are given below.

Box 26: Examples of data collected

Thumbstar

Thumbstar are a UK based mobile games app developer, known for their popular *Smash it! Adventures* app. The app requires access to "*mobile device identifiers including MAC address and IP address*." Other device information collected includes "*the name associated with [the] mobile device, device type, mobile telephone number, country and any other information [the user chooses] to provide, such as [their] user name, character name, geo-location or email address*." Thumbstar is also capable of accessing the user's contacts so that the user may invite contacts to play with them. For analytics purposes, Thumbstar collects "*usage statistics about [the user's] interaction with the Thumbstar Service*".[305]

[303] Facebook, 'Developer Guidelines: Login', https://developers.facebook.com/docs/facebook-login/permissions/v2.3#adding

[304] Facebook, 'Developer Guidelines: Permissions', https://developers.facebook.com/docs/facebook-login/permissions/v2.3#adding

[305] Thumbstar, 'Privacy Policy', http://www.thumbstar.com/privacy-policy

Halfbrick and *Fruit Ninja*

Fruit Ninja is a popular action mobile games app, published by Halfbrick Studios, which is available for Android, iPhone, iPad and Windows Phone. The free *Fruit Ninja* app requests permissions for associating accounts on the device, phone status and identity and location. Information about the device and various device and user identifiers are collected.[306] When a user downloads a game, a random, hashed ID is generated for that user which is stored in the shared storage on their device. Information is checked/linked to that hashed ID any time the game starts and/or a new Halfbrick game is installed on their device.

Halfbrick Studios may also collect information when a user signs in to games with Facebook, Google+ or another provider. Halfbrick Studios will "*use the identity and friends from that provider so that [the user] can play the same game on different devices, pick up where [the user] left-off, and interact with [their] friends*".[307] When a user signs into the app through a social media account, Halfbrick Studios collects information such as the user's profile picture and friend list. However, Halfbrick do not routinely collect any information such as names, email addresses or demographic information from customers using their gaming apps on mobiles or tablets.

In addition to data collected by the app developer, some data is also collected by third-party services. For example, third-party analytics companies may collect gameplay data on behalf of the developer, and developers use app store analytics tools and the services of analytics firms to link their games with demographic data to provide high level segmentation data. For example, Facebook Analytics for Apps and Google Analytics can provide such aggregated app store data. Third-party analytics providers such as App Annie and Flurry can also aggregate app store data to provide performance information to the app developer on information such as downloads, revenues, total active users and user demographics.

Whilst most of the data collected by third parties and shared with app developers is either pseudonymous or aggregated and anonymous, app-stores do also collect personal data. Typically a user will be required to register with an app store to create an account and provide personal information including name, payment card details and billing address. This payment data is used to process transaction and payment data including when downloading paid-for apps, or when making in-app purchases. However, this data is captured only by the app store and is not passed directly to the app developer or publisher.

[306] These include the model of the device, operating system, screen resolution, web browser, IP address, MAC address, advertising identifier, language, cookies and other generated unique identifiers.

[307] Halfbrick Studios, 'Privacy Policy', http://halfbrick.com/pp

Mechanisms for data collection

App developers must interact with the operating system of the mobile device in order to be able to collect data stored on the device and gain access to components such as the camera or microphone. The operating system provider is responsible for the Application Programming Interfaces (APIs), which enable an app developer to interact with the OS and as such are able to determine the means and extent of access to consumer data on the device.[308] To access this data, the app must make subroutine calls to the operating system through an API, which controls the release of information according to privacy controls in place at the operating system level. Therefore, the app is effectively 'sand boxed' and able only to access information that the operating system permits.[309] Whilst app developers are normally the main data controllers, they must work within the limitations imposed by the operating system.

Most developers will use their own proprietary code, but in some cases developers may use third-party 'libraries'[310] or blocks of third-party code. App developers also integrate third-party code to facilitate advertising, analytics or another service or features in an app. These libraries or 'software development kits' ('SDKs') provide the building blocks of app development. They work within an app and collect data from devices, either sending it directly to the developer or collecting and analysing the data on behalf of the developer. One interviewee told us that developers use an SDK provided by it, which then communicates with its servers to register gameplay 'events'. Data on gameplay is then presented to the game developer over an online dashboard. This data is presented at the aggregate level and cannot be used to identify any individual users.

However, complex third-party code or a lack of transparency in descriptions may hinder developers' ability to understand the data implications of a given 'library', a block of code, or, software SDK.[311]

---

[308] Data Protection Working Party for the European Commission, 16th May 2011, 'Opinion 13/2011 on Geolocation services on smart mobile devices, 881/11/EN WP 185', http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2011/wp185_en.pdf

[309] This is not necessarily true for mobile devices that are 'jail broken', where modification of the operating system by the device user could provide greater privileges to applications.

[310] Privacy Grade, 'Third Party Libraries', http://privacygrade.org/third_party_libraries

[311] Lookout mobile security, 'Mobile app advertising guidelines', https://www.lookout.com/resources/reports/mobile-ad-guidelines

Our interview with the ICO suggested that although many large developers should be sufficiently aware and conscious of data protection legislation, there is a risk that 'amateur' developers may not be informed over the exact type and volume of data collected when using third-party code and thus not able to inform users about the personal data processed by their app. Whilst there is little evidence to demonstrate the extent of this concern in practice, the ICO is working to improve awareness amongst app developers by providing guidance on data protection laws.

The number of third-party libraries and connections to URLs has also been noted as a potential concern. Research institute, Eurecom, conducted a study of 2,000 free android apps (not just games apps) and discovered that while 73.2% of the apps surveyed did not connect with a tracking website, a minority do. Some apps connected to as many as 2,000 distinct URLs.[312] Separately, games apps are reported to be the worst category of apps in terms of number of third-party libraries used.[313] For example, PrivacyGrade assigns a 'grade' to the most popular apps on Android based on the differences between a consumer's expectations of what data an app needs to collect and what data the app actually collects through third-party libraries in the app. Popular games such as, '*Despicable Me*' and '*Drag Racing*' are among the most poorly ranked apps.[314]

How data collection is controlled

Aside from the involvement of public and data protection authorities, who uphold information rights, and the OS that controls data collection through the APIs they make available, the collection and use of information by the games app will also be influenced by app store policy. App stores provide guidelines to specify what data app developers can collect, how it is collected, the extent of third-party access, how the data is stored and how consumers are informed about data collection.

[312] The Guardian, 6th May 2015, 'Free Android apps connect to thousands of tracking and ad URLs, research shows', http://www.theguardian.com/technology/2015/may/06/free-android-apps-connect-tracking-advertising-websites

[313] The Wall Street Journal, 'What They Know – Mobile', http://blogs.wsj.com/wtk-mobile and Forbes, 11th November 2014, 'Games Like Fruit Ninja – Not Facebook – Get Worst Grades On App Privacy', http://www.forbes.com/sites/parmyolson/2014/11/11/games-like-fruit-ninja-not-facebook-get-worst-grades-on-app-privacy

[314] Privacy Grade, http://privacygrade.org

Box 27: App store guidelines and policies

Apple App Store Policy

Apple requires developers to "*provide clear and complete information to users regarding collection, use and disclosure of user or device data*". Apple specifies that apps are unable to "*transmit data about a user without obtaining the user's prior permission and providing the user with access to information about how and where the data will be used"* and that *"apps that require users to share personal information, such as email address and date of birth, in order to function will be rejected*"[315]. They also specify that *"[d]evelopers that attempt to reverse lookup, trace, relate, associate, mine, harvest, or otherwise exploit Player IDs, aliases, or other information obtained through Game enter will be removed from the iOS Developer Program*".

Google Play/Android Policy

Google states that "*[i]f the users provide [the app developer] with, or [the] Product accesses or uses user names, passwords, or any other login or personal information, [the app developer] must make the users aware that the information will be available to [the] Product, and [the app developer] must provide legally adequate privacy notice and protection for those users*"[316]. Regarding advertising, Android guidelines require users to be informed and given the opportunity to opt out of sending their data to advertisers: "*It is important to respect user privacy if certain parameters, such as demographics or location, are passed to ad networks for targeting purposes. Let your users know and give them a chance to opt out of these features.*"[317]

App stores rely on developers to adhere to contractual obligations and data protection laws. It appears to be difficult to verify whether app developers comprehensively adhere to the relevant policies and whether developers are using data for the stated purpose. However, platforms may review the permissions that an app requests and operating systems can ensure that consumers are made aware of the permissions requested by an app. For example, the Google Play store identifies and may remove potentially harmful apps using their review and scanning systems, based on the permissions these apps request.[318] Manual review of apps has also been introduced for apps on Google Play to ensure that they

[315] Apple, 'App Store Review Guidelines', https://developer.apple.com/app-store/review/guidelines/

[316] Google, *Google '*Play Developer Distribution Agreement', https://play.google.com/about/developer-distribution-agreement.html

[317] Google, 'Advertising without Compromising User Experience', https://developer.android.com/training/monetization/ads-and-ux.html

[318] Google, 'Review app permissions', https://support.google.com/googleplay/answer/6014972?hl=en-GB

adhere to Google's developer policies.[319] An alternative approach used by other app stores is to have an app review team that selects and reviews apps at random to examine app compliance with guidelines. Furthermore, app stores aim to address complaints made by users about apps, as well as monitoring user feedback of apps to identify problems.

Mechanisms by which consumers are informed of data collection and usage

The Data Protection Act requires 'informed consent' from users in order for a data controller to process personal data from them. However, obtaining informed consent in apps can present difficulties due to the limited space available on mobile device screens, the number of platforms involved and the different consent methods meaning consumers may have to provide consent in a number of different ways depending on the app and the platform they are using. There are two main ways in which consumers are informed in practice. Permission requests are controlled by the operating system and app developers are encouraged to include links in-app that direct the user to an online privacy policy.

The operating system handles permissions requests

Permission requests by an app to access data from a device are handled by the device's operating system. This automates the system of presenting requests to users for access to consumer data at the appropriate time. These permission requests may be presented to users in the app store at the point of installing the app, or requests may be made as an app is about to access data from the consumer ('just-in-time notifications'). Users are also able to read a summary of app permissions that will be requested in the app store when reading the app summary.

There are some slight differences in the processes for gaining consent used by the two most popular operating systems, Apple's iOS and Google's Android, mainly related to the point at which consumers are notified.

Google Android presents a set of permissions at the point of download

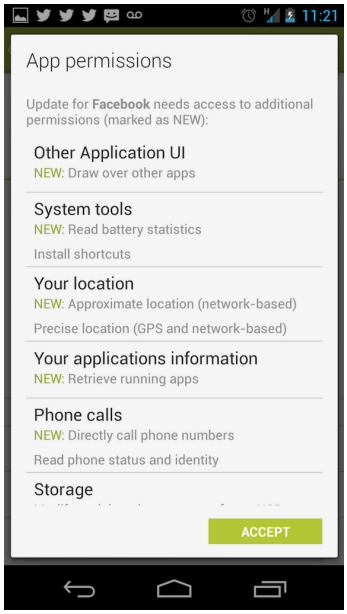
A device using the Android operating system will present the permissions required by an app to a would-be user at the point of download.[320] The permissions are listed in order of items of data that Google believes are most important to the user. This is an all-inclusive set of permissions. The process of consent does not

[319] Google, 17th March 2015, 'Creating Better User Experiences on Google Play', http://android-developers.blogspot.com.es/2015/03/creating-better-user-experiences-on.html

[320] This also applies to Windows Phones.

permit the users to select or de-select categories of data within the list that it is presented with, it is required to accept the sharing of all items on the list at the point of download.[321] [322] However, for some features, such as collection of location data, there are OS-level features that enable the user to override this permission. An example of the user notification process on Android is presented in Figure 15 below. Note in the illustration that permissions relating to a user's location are set out separately and explicitly.

Figure 15: Consenting to data usage within Android



Apple iOS presents 'just-in-time' prompts to gain permission

A device using Apple's operating system, iOS, seeks to gain user permission to allow apps to access consumer data by sending request permissions to the user at the point at which an app wishes to access their personal data. Such a notification is given when an app wants to access both information stored on the device (for example, contacts, calendars, photos, reminders) and access to certain services and sensors (such as location services, Bluetooth sharing, microphone, and camera). When a user grants an app permission to access a certain type of data, the app has access to

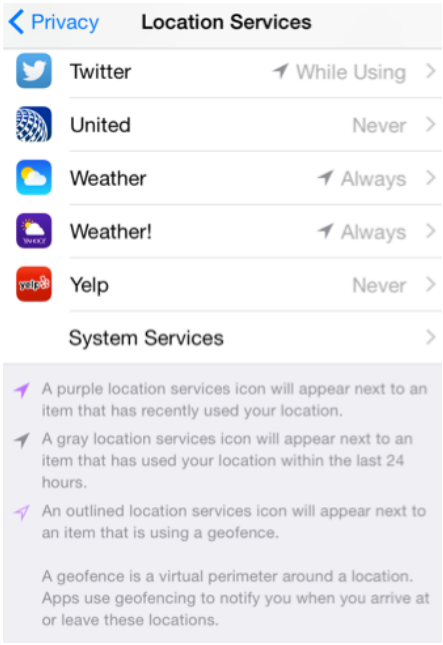
[321] We note that Android did temporarily offer users an 'app ops' feature that allowed the management of individual permissions, but the feature has been removed amid concerns that it could interfere with the functioning of legacy apps. CNET, 19th December 2013, 'Why Android won't be getting App Ops anytime soon', http://www.cnet.com/uk/news/why-android-wont-be-getting-app-ops-anytime-soon/

[322] We note that in May 2015 Google announced that they will change their stance on this and move to something more in line with Apple's "just-in-time" notification method. See http://www.theguardian.com/technology/2015/may/28/google-android-m-software-privacy-battery-life

that data until a user changes this permission, which can be done by accessing the settings menu and reviewing all permissions that have previously been granted, either by looking at all apps accessing each type of data or looking at all data accessed by a specific app.

As with Android, location information is given special distinction within iOS. Users can access a 'location services' section within the Settings menu, which lists all apps accessing location information and the times at which they last accessed this information. The user can change permissions in relation to location data here (choosing between always allowing access, never allowing access or allowing access only when the user is using the app). An illustration of the 'location services' options is shown below:

Figure 16: Location services options in iOS8



Developers often provide a link to their privacy policy in-app

In terms of notifying customers about how data will be shared with third parties, some app developers have a link to their privacy policy in the app, which may include which data is sent to third parties or which third parties the developer uses. For example, one games developer we spoke to told us how they take privacy seriously and have a 'privacy' link in all their apps. When the customer clicks on the link they are redirected to their mobile web-browser and the developer's privacy policy is displayed in full. However, privacy policies are not always available within the app and there is no compulsion for users to read them before using the app. Without an in-app link, users would have to know who the developer/publisher is and go directly to their online site to read the privacy policy.

As well as app developers, third-party data handlers, such as ad networks or analytics firms, have privacy policies publicly available online or within an app, setting out what information is collected

and how it is used. For example, Flurry, a major analytics firm, includes a statement on its website to inform consumers that it will use only anonymous and aggregated data for the purposes of their analysis and that it does not engage in collection, storage, buying or selling of consumer personal information. Flurry sets out clearly that it does not profile consumers but instead profiles apps, which does not require personal information.[323]

However, users may still not be entirely aware of the collection and use of consumer data

There appears also to be some variation in the degree of consumer concern about the use of their personal data, with a significant proportion of users either not aware that this might be an issue, not concerning themselves with this issue actively, or relying on third parties to ensure appropriate conduct in relation to consumer data. Supporting the view that consumers may not be fully informed, we note that:

- MEF research in 2013 revealed that, of the top 100 free mobile apps available from the Apple App Store and Google Play, around half did not make their privacy policy available before download, around a third had a privacy policy accessible within the app, and the remainder did not provide a privacy policy at all. The average privacy policy was over 3,000 words long and 18% of privacy policies were not written in plain English.[324]
- Similarly, a 2014 survey by the Global Privacy Enforcement Network found that "*85% of the 1,211 apps surveyed failed to clearly explain how they were collecting, using and disclosing personal information and more than half (59%) of the apps left users struggling to find basic privacy information*." They also found that "*almost one in three apps appeared to request an excessive number of permissions to access additional personal information*."[325]

[323] Flurry, 7th December, 2010, 'Audience Analysis While Respecting Consumer Privacy', http://www.flurry.com/bid/52136/Audience-Analysis-While-Respecting-Consumer-Privacy#.VNy4i4tLH8s

[324] MEF research, reported in Mobiforge, 20th August, 2013, '28 percent of the top 100 apps still don't have a privacy policy', http://mobiforge.com/news-comment/28-percent-top-100-apps-still-dont-have-a-privacy-policy

[325] Information Commissioners Office, 10th September, 2014, 'Global survey finds 85% of mobile apps fail to provide basic privacy information', https://ico.org.uk/about-the-ico/news-and-events/news-and-blogs/2014/09/global-survey-finds-85-of-mobile-apps-fail-to-provide-basic-privacy-information/

- A Communications Consumer Panel research report found, in May 2011, that 55% of UK internet users were not aware that mobile apps are capable of accessing personal data.[326]
- Separately, a Systamex Norton survey found that 54% of those surveyed were not aware that apps could track location, 73% did not know that apps can have access to personal information and photos, 77% were not aware apps can access a phone's camera and microphone and 85% did not know that apps could modify web browser bookmarks.[327]
- A survey by Intel Security found that 65% of 18-14 year olds chose not to read terms and conditions, 20% of respondents said that they did not care about them and 20% said that they trusted the app stores. [328]
- On the other hand, an ICO YouGov poll found that 62% of app users are concerned about the way apps can use their personal information and 49% of have decided not to download an app due to privacy concerns.[329]

4.3.2 Social network apps

Types of data collected

As with mobile games, app developers of games on social network sites also collect pseudonymous data related to gameplay. However, when users access games through Facebook, developers can access a significantly broader range of consumer data compared with mobile apps that do not use a social login. The difference is due to the substantial amounts of consumer data that may be held on users' Facebook profiles.

[326] Communications Consumer Panel, May 2011, 'Online personal data: the consumer perspective', http://www.communicationsconsumerpanel.org.uk/Online%20personal%20data%20final%20240511.pdf

[327] Norton, December 2014, 'Norton Mobile Apps Survey Report', http://www.slideshare.net/symantec/norton-mobile-apps-survey-report

[328] Help Net Security, 10th February, 2015, 'Unsurprisingly, adults don't read terms and conditions of mobile apps', http://www.net-security.org/secworld.php?id=17931

[329] Information Commissioners Office, 19th December, 2013, 'Online YouGov survey on apps and privacy for ICO', http://ico.org.uk/news/latest_news/2013/~/media/documents/library/Data_Protection/R esearch_and_reports/yougov-ico-survey-apps-and-privacy-20131219-csv.csv

For example, games apps on Facebook can collect 'public profile' information which includes name, public profile pictures, username, user ID (account number), networks (such as university network) and any other information a user makes publicly available on the profile. An app may also extend permissions to email and a list of the user's friends who have also downloaded the app. App developers are able to request information on Bio, birthday, education history, events, relationships, Likes, News Feeds, interactions, age range, location and other preferences and behaviours on the social network.

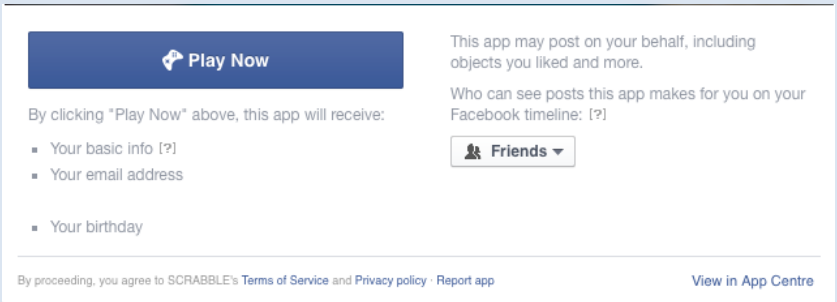
For example, some apps such as 'Scrabble' (described in Box 28 below) collect a large amount of data about the consumer which is not obviously necessary for the core function of the app. However, we note that app developers will no longer be able to access data about a user's activities and interests (based on their likes of activity and interest related pages) from the 23 June 2015.[330]

Box 28: Example of data collected by the Scrabble games app on Facebook

Scrabble app

When a user attempts to use the Scrabble games app through their Facebook account, Scrabble requests permissions to collect:

- *"Your profile info: description, activities, birthday, education history, groups, hometown, interests, likes, location, relationship status, relationship details, religious and political views, website and work history*
- *Your stories: events, notes, photos, status updates and videos*
- *Friends' profile info: descriptions, activities, birthdays, education histories, groups, hometowns, interests, likes, location, relationship statuses, relationship details, religious and political views, websites and work histories*
- *Stories shared with you: events, notes, photos, status updates and videos"*[331]



[330] Facebook, 'Developer Guidelines: Permissions', https://developers.facebook.com/docs/facebook-login/permissions/v2.3#adding

[331] Data permissions requested through a pop up window when playing on Facebook, See https://www.facebook.com/ScrabbleEA

Whilst apps are able to collect a large amount of information, there may be some disincentives to collect this information, especially where it would not appear to add significant value to the app developer. For example, one app developer that we spoke to told us that although they could collect a wide range of data from its users on Facebook, adding that extra permission request lowers the conversion rate and keeping conversion as high as possible is more valuable to it than getting more data on individual players. The data they do collect is typically limited to public profile data and friend lists (profile data and friends lists are required for the leader boards showing the score ranking among friends). Note also that if an app collects more data than simply the public profile, email address and the user's contacts the specific purpose for collecting this additional information is reviewed by Facebook[332].

Mechanisms for data collection

Social network platforms have a similar but somewhat more involved role than operating systems and app stores for mobile apps in data sharing. Similar to mobile app stores, social network platforms publish terms and conditions for apps that can be published on their platform. These regulate the amount of data an app developer will be able to source from a user's profile. For example, Facebook's policy for app developers in described in Box 29 below:

[332] Facebook, 'Developer Guidelines: Permissions', https://developers.facebook.com/docs/facebook-login/permissions/v2.3#adding

Box 29: Facebook's platform policy for app developers

Facebook's terms and conditions for developers submitting apps to Facebook require app developers to "*provide a publicly available and easily accessible privacy policy that explains what data you are collecting and how you will use the data*". App developers "*may use Account Information in accordance with your privacy policy and other Facebook policies. All other data may only be used outside your app after you have gained explicit user consent*".

Terms and conditions also specify that app developers must "*[o]btain consent from people before using their data in any ad*", and "*[o]nly use friend data including friends lists in the person's experience in your app*." Facebook also regulates the transfer of data from the app developer: "*[d]on't transfer any data that you receive from us (including anonymous, aggregate, or derived data) to any ad network, data broker or other advertising or monetization-related service*." [333]

As in the case of mobile games apps, which use APIs interfacing with a device's operating system to access consumer data, social network apps use APIs interfacing with the social network site to source a user's data. However, the majority of API calls on Facebook need to include an 'access token' which is obtained when someone connects with the app and provides temporary, secure access to Facebook APIs.[334] The most commonly used access token is the 'user access token' which is needed any time an API is used to read a specific user's Facebook data. Users must permit the app developer to obtain a user access tokens, often through a login process.[335].

Social networks are also able to provide data to app developers about the demographic characteristics of their users, leveraging the data social networks held relating to their consumers. For example, Facebook Analytics for Apps provides app developers with market segmentation data (such as age, gender, other demographic information or device used for playing), detailed game activity and high-level game insights based on cohort behaviour (that is, people

[333] Facebook, 'Platform Policy', https://developers.facebook.com/policy

[334] There are four types of access tokens: App Access Token, User Access Token (and Extended User Access Token), Page Access Token (and Extended Page Access Token) and Client Token. See https://developers.facebook.com/docs/facebook-login/access-tokens

[335] In terms of how this is implemented in practice, when setting up a Facebook app, the developer will have the option to 'Get Access Token', with options that can be selected for each data permission, once all the necessary permissions are selected the developer will be returned a section of code that will request these permissions when the app is published. The app token is then pasted into the Custom Facebook Feed plugin. See Smash Balloon, 'How to get an extended Facebook 'User' Access Token', https://smashballoon.com/custom-facebook-feed/docs/get-extended-facebook-user-access-token/

who took the same action in a game at the same time). This data is shared at an aggregated and anonymous level.

Mechanisms by which consumers are informed of data collection and usage

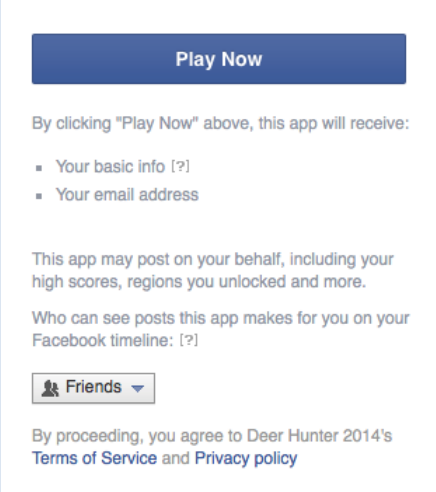
Apps on social networks request permissions when a user 'installs' an app. Users are presented with the relevant categories of data requested in a pop-up window. As mentioned above in Section 0 all apps can collect public profile information and this is presented to users as 'Your basic info' that the app in question is accessing.

Facebook makes potential games users aware of this data collection, along with the discretion that apps have to collect further personal information held by Facebook, at the point of play (see Box 30 below).

Box 30: Example of 'point of play' notification of data collected when playing the app

Deer Hunter by Glu

For example, Deer Hunter, an action game published by Glu Games on Facebook (as well as on iOS and Android), requests permissions as below, including a link to Glu's privacy policy.[336]



On clicking the privacy policy link, users are able to read the privacy policy of the app developer on its website detailing the information that it collects from users of the game. All games published in the App Center are required to follow a similar process.

[336] Deer Hunter 2014 on Facebook's App Center, https://www.facebook.com/games/deer_hunt_two/?fbs=110

App developers tend to provide a single privacy policy across all of their games. Privacy policies for social network games include details of information on cookies from a consumer accessing a website, information that will be collected for competitions and information collected from permissions. For example, Glu's privacy policy details the information that the user provides to Glu, how information is automatically provided to Glu when a user interacts with one of its games, use of cookies and push notifications, information collected from other sources, use of information including sharing of information for social sharing, third-party services and security.[337]

However, Facebook has recently offered customers the option to play games without having to provide any data to the app. In response to consumers who wish to play social network games but who are concerned about the about of data that may be collected about them, Facebook introduced 'Anonymous Login'. However, we understand that uptake thus far is very low.[338]

4.4 Use of consumer data

The data collected through apps may be necessary to enable the functionality of apps (e.g. in so-called 'augmented reality' games that rely on observing a user's real-world location) or for functionality in any app where 'phone state' information is used in order to pause the game when there is an incoming call.

Beyond these uses related to the game's functionality, data collected by developers is generally analysed to understand app usage patterns and draw insights that can be valuable as they seek to:

- drive user acquisition;
- maximise user engagement and retention; and
- monetise users.

User acquisition

In seeking to attract users, developers may choose to advertise their games using mobile advertising, advertising on Facebook and other social networks, or advertising their games within other apps. Targeted advertising can be used to focus adverts on certain market segments based on aggregate and anonymous data about their existing customer base or target market. Developers can also

[337] Glu Games, 'Glu Mobile Privacy Policy', http://www.glu.com/privacy

[338] Recode, 6th March 2015, 'Whatever Happened to Facebook's Anonymous Login', http://recode.net/2015/03/06/whatever-happened-to-facebooks-anonymous-login/

benefit from players 'sharing' their gaming experience on a social network or sending game invites to their friends as a means of reaching out to potential new users.

Engagement and retention

Consumer data can be analysed to derive insights about user engagement, which can then help developers identify possible improvements to the app. Typically this involves using pseudonymous gameplay data to analyse gameplay patterns, which is then used to inform future development and improvements to the game's design. For example, Halfbrick Studios' Data Policy specifies that the purpose of data collection is to "*continually improve [its] games, social features and [its] Services to [the user]*"[339], as well as to deliver services that the user has requested.

Monetisation

Consumer data is also important for optimising developers' approaches to monetising users. Analysis of gameplay data can help to boost take-up of in-app purchases by providing insight into in-game behaviour and the circumstances that are most likely to induce purchasing or making social connections to grow the user base. Where third-party in app advertising is used to generate revenues, consumer data may be used to help provide targeted adverts.

Analysis used to drive the above outcomes may be carried out in-house by a developer and/or using third-party tools. A survey of developers has indicated that there are a large variety of third-party tools available, including specialist solutions aimed at the games apps sector and other more wide ranging tools such as Google Analytics and Facebook Analytics for Apps.[340]

We discuss each of these main uses of consumer data in turn.

4.4.1 Driving user acquisition

Driving take-up of an app is the first step towards monetisation for any developer. Where app downloads are paid-for, and where the freemium model is used, attracting users is a prerequisite for any monetisation strategy.

[339] Halfbrick, 'Privacy Policy', http://docs.halfbrick.com/PrivacyPolicy.htm

[340] VB Insight, 18 April 2015, 'Mobile App Analytics: What winning mobile developers use', http://insight.venturebeat.com/report/mobile-app-analytics-what-winning-mobile-developers-use

Developers advertising their apps

As in most consumer-facing sectors, developers may seek to attract customers by advertising through various media, including online and mobile, but also for example through TV advertising and celebrity endorsements in the case of large developers.[341] They may also seek to improve user acquisition by optimising their 'ranking' on the app store or social network, analogously to how firms in any sector might seek to optimise their online search ranking.

Targeted mobile advertising appears to be an important source of app installs

Given that the number of apps available is enormous, it is difficult for an individual app to attract a significant number of users without reaching out to those who are most likely to be interested. The shelf life of many games apps is relatively short "*targeting users' likes and dislikes, their buying habits, their age and their gender*" may be crucial in driving profit.[342] As such, targeted advertising of an app is preferable to generic advertising.

Facebook plays a role a major advertising platform for games apps, serving adverts for both mobile games and social network games. With its precise targeting options that leverage consumer data, which may include details of the apps that a Facebook user currently has installed,[343] combined with the fact that Facebook is increasingly used on mobile devices, Facebook mobile advertising may be particularly effective for developers seeking to acquire new mobile app users.[344] For example, a developer can target Facebook users on the basis of a set of Facebook 'Likes' that indicate a specific interest that is expected to be strongly related to the app being advertised.[345] Facebook is effective at targeting particular gamers and developers have been reported to spend substantial amounts of money advertising on the social network.[346]

---

[341] See e.g. Forbes, 16 November 2014, 'A $40M Ad Budget Buys 'Game of War: Fire Age' Kate Upton', http://www.forbes.com/sites/insertcoin/2014/11/16/a-40m-ad-budget-buys-game-of-war-fire-age-kate-upton/

[342] International Business Times, 4 October 2012, 'Social-Gaming: How Facebook Game Developers are Using Your Data to Build Games for You', http://www.ibtimes.co.uk/zynga-social-gaming-facebook-tableau-wild-tangent-391028

[343] Business Insider, 6 July 2012, 'Here's How Facebook Is Going To Rake It In With Mobile Ads For Games And Apps', http://www.businessinsider.com/facebook-mobile-ads-games-apps-2012-7

[344] The Wall Street Journal, 24 April 2014, 'Why Mobile App Install Ads Are Suddenly Such a Huge Deal', http://blogs.wsj.com/cmo/2014/04/24/why-mobile-app-install-ads-are-suddenly-such-a-huge-deal/

[345] Gamasutra, 7 June 2012, 'Using Facebook Ads to Find Your Game's Audience', http://www.gamasutra.com/view/feature/171895/using_facebook_ads_to_find_your_.php

[346] Ibid.

However, employing such targeted advertising strategies does not typically involve the use or sharing of any data that the app developer collects about its users. Whilst the app developer may have a preference for the broad demographic it wants to advertise its game to, the serving of adverts is based on the data Facebook holds about its users and this is not shared with the advertiser as part of the process.

Developers may also 'cross-promote other apps in-game'

Where a developer already has an established user base on some apps, it may opt to leverage this to try and induce users to try its other apps. One developer told us that it only used ads for its own apps, rather than any third-party ads, and that it would use the information it held on its users in order to tailor the ads they see. One app developer we spoke to told us that this typically relies on simple logic; for example not advertising a game a user has already downloaded. Applying this simple logic to decide which adverts to serve to a particular user is based on pseudonymous data held about the user, including any information linked to the user's ID including a log of all the developer's games the user already uses and any past in-app purchases they have made.

Social networks offer opportunities for user acquisition

Developers of social games apps have a particularly rich set of tools available to try to acquire new users, exploiting the connections that exist on social networks between an app's existing users and non-users by encouraging users to 'share'. For example, games can request permission to send invites to the game to the user's friends by "*strategically positioning invite screens when the user runs out of lives*". This strategy is reported to have been successful for popular games such as King's Candy Crush.[347] In this instance, the developer effectively uses consumer data – the user's friends list – to send invite notifications to all friends on the user's behalf, though there are controls that enable social network users to turn off such notifications.[348] However, social networks or app stores may be moving towards restricting developers' opportunities to incentivise social sharing.[349] One of our interviews with a developer suggested that its users seems to be more reluctant to use Facebook logins when playing games because of the number of requests sent

[347] Gamesauce, 9 September 2014, 'Why mobile game user acquisition cant be an afterthought', http://gamesauce.org/news/2014/09/09/why-mobile-game-user-acquisition-cant-be-an-afterthought/

[348] International Business Times, 12 November 2013, ''Candy Crush Saga': How To Turn Off Facebook Notifications For The Popular Social Media Game, Forever', http://www.ibtimes.com/candy-crush-saga-how-turn-facebook-notifications-popular-social-media-game-forever-1466754

[349] See e.g. Techcrunch, 9 June 2014, 'Apple Begins Rejecting Apps That Offer Rewards For Video Views, Social Sharing', http://techcrunch.com/2014/06/09/apple-begins-rejecting-apps-that-offer-rewards-for-video-views-social-sharing/

through notifications. Therefore, we expect to see a decline in these practices in future.

Social networks also present opportunities to reach new users. For example, Facebook offers developers the possibility to offer 'in-feed gaming', where developers publish a story on behalf of a user that includes a small, embeddable version of the game; this will then appear on friends' news feeds, where new users can easily experiment with the game within the news feed and can be encouraged to go on and play the full version.[350]

Consumer data analysis will help inform the user acquisition process

Developers will typically analyse consumer data for the purposes of monitoring the process of user acquisition and identifying any possible improvements or opportunities. For example, they will monitor users' origins – where a user has come from in order to access the game - which could be through a direct search on an app store or social network, through browsing or following recommendations on an app store or social network, or by following an ad for the game. This information can then inform the developer's future marketing practices.

Third parties may assist in user acquisition

Firms such as App Annie and GameAnalytics also offer freely available analytics services that allow developers to monitor key metrics over time, including where users are coming from and what the average acquisition costs are.[351] The information provided by these analytics companies will typically be at the aggregate and anonymous level.

Other specialist third-party marketing firms exist that will assist developers in devising targeted mobile marketing campaigns to optimise their user acquisition process. For example, one such service provider - Fiksu - tracks its client's applications as they are downloaded and determines when users respond to an advert. This is done by storing a mobile ID when a user clicks on a client's advert or application delivered by Fiksu. This is then used to deliver relevant adverts to the user based on app download history and previous ads clicked on.[352]

Note that in some cases, consumers can install ad-blockers such as Adblock Plus[353], which prevents ads from being shown. However,

[350] Facebook, 'Feed Gaming', https://developers.facebook.com/docs/games/feed-gaming

[351] GameAnalytics, 'Features', http://www.gameanalytics.com/features.html and AppAnnie, 'Advertising Analytics', https://www.appannie.com/tours/advertising-analytics/

[352] Fiksu Privacy Policy, https://www.fiksu.com/privacy-policy

[353] Adblock Plus, https://adblockplus.org/en/android-install

ad blocking is not possible for apps on iOS devices[354] and ad blocking apps were removed from the Google Play store in 2013. Some are still available on alternative app stores and on developers' websites.[355] It is also possible to install apps, like Disconnect[356], that focus solely on protecting the personal data of a user. This software, which is installed as an app on the device, protects the identity of the device user from trackers. However, there is not currently a 'Do Not Track' standard and appropriate responses have not been defined, meaning that some third parties will simply not respond to these do not track signals. The World Wide Web Consortium (W3C) is currently working to develop a Do Not Track Standard.[357]

4.4.2 Maximising user engagement and retention

Engaging users is key for monetisation

A developer has strong incentives to make a game as enjoyable as possible and to make its appeal as enduring as possible, since this will enable it to achieve a large group of regular users, which is essential for the predominant freemium model of monetisation. Developers we interviewed indicated that the longer a user plays, the more likely the developer is to monetise that user. A study also indicates that ARPU (average revenue per user) can increase dramatically as users move from 1-10 plays to 11-50 and more.[358]

Data can key game features

Consumer data can play an important role in keeping users engaged by allowing certain aspects of functionality to be delivered in-game, or allowing the customisation of the game's features. For example, data on a user's contacts or social network friends can be used to connect users in-app. Users may play each other through an app, for example, in EA's *Scrabble*, or in games with levels, users may be able to see the in-game progress of their contacts, as in King's *Candy Crush*, for example. Further information that friends can observe or share can include their username, avatar image (which

[354] Adweek, 22nd April 2015, 'Adblock Plus Develops First Mobile Ad Block for iPhones', http://www.adweek.com/news/technology/adblock-plus-develops-first-mobile-ad-blocker-iphones-164212

[355] Adblock Plus, 14th March 2013, 'Adblock Plus for Android removed from Google Play store', https://adblockplus.org/blog/adblock-plus-for-android-removed-from-google-play-store

[356] Disconnect, https://disconnect.me

[357] World Wide Web Consortium, 24 April 2014, 'Tracking Preference Expression (DNT) W3C Last Call Working Draft', http://www.w3.org/TR/tracking-dnt/

[358] Kongregate, 2013, 'Maximising Player Retention and Monetization in Free-to-Play Games', http://www.slideshare.net/DavidPChiu/kongregate-digital-taipei-2013

may be their profile picture from their Facebook profile) and any other aspects of their gameplay. They may also be able to interact in-game in various other ways, for example by sending 'lives' or virtual gifts and currency, as in King's *Farm Heroes*.

Data allows personalisation

App developers often collect data to personalise the games experience for the user. At the simplest level, this may include using location data to provide the game in the right language. App developers may also use personal data to offer more personalised gameplay. For example, a profile picture or user name can be chosen to create an avatar for a game. Birthday dates may be collected to send out special offers. Fruit Ninja, published by Halfbrick Studios, collects geolocation to offer "*relevant games/features/promotions/advertisements*" in-game.[359]

Gameplay data can help enhance game design to boost engagement

One of the main uses of data collected by games apps is to learn more about users' behaviour and game play activity to provide the feedback needed to continuously develop and adjust the gaming experience.[360] The developers we spoke to suggested that gameplay data is the most important data they collect. The pseudonymous data collected about gameplay helps a developer infer key information such as which levels, game modes or characters are associated with user engagement, when and why users drop-out, propensity to connect with others or propensity to make in-app purchases. These insights can be used to further improve the app itself to maximise user engagement and retention.

In this regard, developers benefit from the ability to release frequent updated versions of apps, which become available to new or existing users alike. Updates may be used to make changes to the game's design and functionality or to introduce new content (e.g. new levels). More broadly, our interviews indicated that insights from the analysis of gameplay data would also be useful in the development of new apps. Learning key lessons from previous apps, their structure and how customers interact with certain elements or features of the game allow the developer to incorporate 'successful' features into new games.

Specifically, it can help set an adequate difficulty level

For example, our interviews with app developers indicate that setting an appropriate difficulty level is a highly important aspect of game design. Games that are deemed too easy will quickly lead to boredom; games that are too difficult will lead to frustration. In either case, users are more likely to stop playing the game and are also more likely to leave negative reviews about the game,

[359] Halfbrick, 'Privacy Policy', http://docs.halfbrick.com/PrivacyPolicy.htm

[360] idean, November 2013, 'Mobile Content Market in Finland 2012-2016', http://www.teleforum-ry.fi/wp-content/uploads/2013/11/Mobile-content-market-in-Finland-2012-2016-desk-top.pdf

worsening the prospects for capturing more users in the future, given the importance of user reviews in influencing a consumer's download decision.[361] Data collected on number of levels played in a session, number of attempts to complete a level or data on which level consumers stop playing can all help optimise the level of difficulty to encourage users to play the game for as long as possible.

We understand that apps often offer the same difficulty level to all users at present, and any changes to the overall level of difficulty, as might be introduced by an updated version of the game, would apply to all users. However, one developer we interviewed is investigating options to adjust difficulty dynamically, varying the game design or difficulty among different groups of players as a way to significantly increase engagement. Updates or changes could be served to individuals based on the pseudonymous data held about the user i.e. that linked to their unique user ID.

App developers may also take advantage of any features allowed by the social network or operating system, in order to try to optimise user retention. For example, data collected about a user's usage of the app (pseudonymous data) allows developers to serve timely push notifications to remind users about the game when they have not played it for a period of time. A reminder may be accompanied by additional incentives, such as discounts on in-app purchases. There is evidence that this can boost retention, though equally developers risk causing annoyance.[362]

Analysis may be in-house or outsourced

Analysis of gameplay data can be done in-house or through a third party, in which case there is typically a transfer of pseudonymous data to an analytics firm. For example, Tableau is a third-party software tool being used by some developers to analyse large volumes of gameplay data – in the context of a shooting game, this was used to produce 'heat maps' for hundreds of thousands of users showing the in-game locations where players were 'killed', allowing the developer to judge whether the level design was appropriate or could be improved to improve customer engagement and encourage continued gameplay.[363]

[361] See slide 23, Newzoo, 'Spotting the mobile spenders', https://images.eurogamer.net/2014/gamesindustry/pdf/newzoo/Newzoo_Free_Report_Spotting_The_Mobile_Spenders_V1.pdf

[362] Kiip, 26 July 2014, '3 Tricks to Increase User Retention', http://blog.kiip.me/developers/increasing-user-retention/

[363] International Business Times, 4 October 2012, 'Social-Gaming: How Facebook Game Developers are Using Your Data to Build Games for You', http://www.ibtimes.co.uk/zynga-social-gaming-facebook-tableau-wild-tangent-391028

Third parties may be engaged commercially, often charging a subscription fee, but there are also currently some free third-party solutions that may assist developers in conducting analysis of gameplay data. As mentioned above, GameAnalytics provides a free service that allows developers to monitor various metrics on an interactive dashboard and this includes gameplay metrics such as number of sessions per user and average sessions per user. GameAnalytics claims that this can allow developers to estimate the quality of players, improve retention rates, eliminate design decisions that drive players away and compare the effectiveness of different game design options.[364]

In some cases, app developers will pay a subscription fee to obtain market data from analytics firms. For example, AppAnnie offers some such data for free (e.g. related to app rankings), but charges for access to more detailed information that allows developers to benchmark their performance against competitors, on metrics such as download volumes and revenues provided at an aggregated and anonymous level.[365]

Where third parties are involved, they would typically have access to data collected through the app in order to provide analytics services and display key metrics. One provider of gameplay analytics services told us that as part of its terms and conditions requires developers to notify their users of the fact that their game usage is being tracked by the analytics providers, and to incorporate at least part of its terms and conditions as part of the games' own T&Cs. In other cases, where data collected by developers is shared with third parties, this is often stated in privacy policies, although it may not always be entirely clear who the parties involved actually are. For example, Glu Games privacy policy explains that data may be shared "*[w]ith vendors, consultants and other service providers who need access to such information to carry [Glu's] work on [Glu's] behalf*"[366]. However, this lack of clarity about the involvement of third parties does not seem limited to the games apps sector.

4.4.3 Monetisation

Having acquired a substantial base of users who are engaged with the game, a developer will then face the challenge of monetising at least some proportion of those users, ideally whilst minimising

[364] GameAnalytics, 'Benefits', http://www.gameanalytics.com/benefits.html

[365] AppAnnie, 'Store Stats Pro', https://www.appannie.com/tours/store-stats-pro/

[366] Glu, 'Glu mobile privacy policy', http://www.glu.com/privacy/

negative impacts on the user experience. As discussed in Section 4.2.3, the primary means of monetisation is via in-app purchases, though advertising is still used to an extent.

In-app purchasing as a source of revenue

Major app developers commit significant resources to analysing monetisation strategies and deciding how best to encourage in-app purchasing in their games, since incremental improvements to the ways in which users are offered or induced to make in-app purchases could have substantial revenue implications. In practice, developers can adjust many aspects of their strategies, including the types and range of in-app virtual items that are offered, their price points, the ways in which they are presented, the times at which they are offered to the user (including any in-game 'triggers' that would cause this) and so on. Designing these strategies may effectively amount to price discrimination, for example where patient users are able to play for free or for little cost, whereas impatient users are induced to make in-app purchases.

Data can be a valuable input in optimising such strategies. At the simplest level, developers will monitor the relative success of different virtual items – e.g. different levels or packs – and use this insight to try to optimise the mix of options and price points offered to the users. This decision may be closely analogous to pricing decisions that firms in many sectors face when selling 'real-world' goods. For example, developers may opt to design the game in such a way that purchasing an additional virtual item creates substantial additional value to the user at little expense, which could lead to many users making a purchase and doing so relatively early on. On the other hand, they may offer high price points[367] on the expectation that a sufficient number of highly engaged users will still make the purchase.

In our interviews, developers told us that, as well as hoping to increase the user base in general by improving the game and therefore increase the volume of in-app purchases. It is particularly desirable to acquire 'whales' – users with a high propensity to make in-app purchases. Where developers opt for the prevalent freemium model, revenue is typically generated from a small minority of users who make in-app purchases. Illustrating this, King

[367] The two approaches are not necessarily mutually exclusive. For examples, see GameAnalytics, 11 December 2014, 'Soft touches on Clash of Clans monetization', http://blog.gameanalytics.com/blog/soft-touches-on-clash-of-clans-monetiation.html

reported 350 million monthly unique users ('MUUs')[368] on average over 2014, but only 9.8 million monthly unique payers ('MUPs')[369] on average. While the vast majority of users generated no revenue, among those users who made at least one payment during a month the average spend (or 'bookings') over the month was $20.21.[370] Similarly, Zynga reported 71 million average MUUs in Q4 2014 and 1.1 million average MUPs.[371] Monthly bookings per MUP were $37.[372] Therefore, learning more about these users' behaviour is valuable to the developer.

Developers experiment with different strategies – either in the same game at different times (e.g. in different versions), or across different games – and monitor monetisation performance in order to assess the relative merits of each strategy. This may be done in-house and/or with the use of third-party analytics.

One developer told us that the key to monetisation from in-app purchases was to keep users playing as long as possible and that insights from analysis would help to achieve this, but the functionality or design of the game is not altered for different users at present. However, they were looking for ways in which it could vary the functioning of the game according to individual user characteristics or behaviour patterns. This could extend to using different monetisation strategies for different users. However, there is little evidence that this occurs in practice at present.

Where in app purchases are made, data on payment and billing, which is personal data, will be used. However, in this case, financial

---

[368] "*MUUs are the number of unique individuals who played any of our games on a particular platform in the 30-day period ending with the measurement date. We calculate average MUUs by adding the total number of unique users as of the end of each month in a given period and dividing by the number of months in the period.*" See King Digital Entertainment PLC, 'Annual Report 2014 Form 20-F', http://investor.king.com/investors/financial-information/quarterly-reports/default.aspx

[369] "*MUPs are the number of unique individuals who made a purchase of virtual currency at least once on a particular platform in the 30-day period ending with the measurement date. We calculate average MUPs by adding the total number of unique payers as of the end of each month in a period and dividing by the number of months in the period.*" Ibid.

[370] This figure refers to Monthly Gross Average Bookings per Paying User (MGABPPU), "*calculated by dividing (1) our total gross bookings in a given period by (2) the number of months in that period, divided by (3) the average number of MUPs during the period.*" Ibid.

[371] Zynga Inc, 'Annual Report From 10-K', http://files.shareholder.com/downloads/AMDA-KX1KB/112322736x0xS1193125-15-60187/1439404/filing.pdf

[372] "*Monthly unique payer bookings per MUP is calculated by dividing average monthly unique payer bookings by average MUPs.*" Ibid.

transactions are usually handled by the app store, which holds user details necessary to process payments, rather than the application itself. For example, on Google Play, a purchase is initiated when the app sends a billing request to the app store relating to a specific in-game product; Google Play processes the transaction and, once complete, sends to the app details such as order number, date and time and price paid.[373]

Advertising as a source of revenue

Some developers choose to display third-party adverts in their apps to generate additional revenue. We understand that this is not always targeted (so does not always require the exchange of any consumer data). Further, and as explained earlier, the relative importance of third-party in-app advertising appears to be declining. Whilst the usage of third-party advertising varies from game to game, one app developer we interviewed estimated that 10% of revenue originates from in-app advertising.

Where targeted advertising is used, this may simply be based on inferred information about the user, purely based on the fact that they are using a particular app, so will be linked only to the pseudonymous data linked to the user's unique ID. For example, players of a sports game might be relatively likely to be interested in sports generally, which could then make them a desirable target audience for a sports retailer's advertising.

Targeting advertising in-app

Beyond this, targeted in-app advertising through ad networks, on the basis of more specific information – whether observed or inferred – may be possible. For example, Facebook's Audience Network offers over 50 targeting options for in-app advertising, leveraging the information it holds on Facebook users (and not using any information collected by the developer). This does not require an app to offer a Facebook login option; rather, Facebook will use an advertising identifier on the mobile device and match it to Facebook users, where possible.[374] Flurry also offers targeting based on characteristics such as age, gender and 'personas', which seem to mainly be inferred from a device's app usage (e.g. extent of usage of different types of apps) and attached to a device ID.[375] Developers relying on advertising revenue are likely to monitor the performance of different ads or different advertising approaches, which again may involve the use of third-party analytics tools.

[373] Google Play, 'Billing Overview', http://developer.android.com/google/play/billing/billing_overview.html#checkout

[374] Facebook, 'Audience Network', https://developers.facebook.com/docs/audience-network/faq#a1

[375] Flurry, 'Brands: What we do', http://www.flurry.com/solutions/advertisers/brands

However, even where no personal data is used, the common use of permanent device IDs for advertising purposes has raised privacy concerns in the past.[376] Recently, Apple released a new identifier (known as an Identifier for Advertising 'IDFA') to be used for anonymous targeted advertising to replace the previous identifier (UDID)[377] and Google has followed a similar approach with a new Android Advertising ID.[378] With the new identifiers, users may reset these advertising identifiers to prevent ad networks from using these as permanent identifiers, while in some cases users also have an option to opt out of targeted advertising that uses the identifiers. Other identifiers that could be used for targeting functionality that do not have the same hardware or subscriber specific nature include Pasteboard Conversion Tracking, a hashed MAC address, browser redirects (similar to cookies). These may also have options for user opt-out, for example, Pasteboard Conversion Tracking in iOS. Note however, that app store guidelines aim to discourage the use of any permanent identifiers that users have no control over; for example, Apple requires developers to comply with terms that state that IDFA is the only way to offer targeted ads.[379]

Whilst third-party advertising in-app may generate some additional revenue for the app, we understand that for games apps, advertising tends to generate only a relatively small income stream. For example, one app developer we spoke to reported that while third-party advertising was used, this only represented about 10% of their total revenue stream. Throughout our research, and interviews with developers, we understand that there is more value associated with reducing the number of third-party adverts, and either providing no adverts at all (to help further enhance the customer experience) or using the space to cross-promote their own games. A strategy of cross-promotion is often popular as it drives the number of people accessing a developer's apps, creating a larger market for monetisation through in-app purchases. Therefore, the common perception is that driving in-app purchases remains the most important monetisation tool for app developers.

[376] See for example, AVG Now, 18 June 2013, 'Apple iOS 7 puts an end to unique device IDs', https://now.avg.com/apple-ios-7-puts-unique-device-ids/

[377] Though this process started in 2011, see e.g. Techcrunch, 19 August 2011, 'Apple Sneaks A Big Change Into iOS 5: Phasing Out Developer Access To The UDID', http://techcrunch.com/2011/08/19/apple-ios-5-phasing-out-udid/

[378] Android, 'Advertising ID', https://developer.android.com/google/play-services/id.html

[379] Apple, 'Submitting the App to App Review', https://developer.apple.com/library/ios/documentation/LanguagesUtilities/Conceptual/iTunesConnect_Guide/Chapters/SubmittingTheApp.html