Rivista Italiana di Antitrust
Italian Antitrust Review
N. 1 (2017)
The New Frontiers of Innovation and Competition
Big Data
Debate

COMPETITION LAW AND BIG DATA: THE ENFORCERS' VIEW

Bruno Lasserre[1], **Andreas Mundt**[2]

Keywords: big data; digital markets; platform; Bundeskartellamt; Autorité de la concurrence

Abstract: *The increasing collection, processing and commercial use of data in digital markets has prompted a broad debate about the role of data in corporate strategies and the application of competition law to such strategies. Today several firms achieve high turnovers based on business models which involve the use of data. In a joint project the French Autorité de la concurrence and the German Bundeskartellamt have analyzed the implications and challenges for competition authorities resulting from data collection in the digital economy and other industries. They provided a comprehensive overview of the existing case law and literature helping to identify key issues and parameters that may need to be considered when assessing the relevance of data for competition law enforcement. Two aspects of particular relevance have to be carefully considered: 1) whether the data under consideration can easily be obtained by rivals and 2) whether the scale and scope of data matter. The paper illustrates the necessity of a differentiated approach and a case-by-case analysis.*

1. INTRODUCTION

Digitalization and, as one of its consequences, the role of data in business strategies, is a widely debated topic. A growing number of industries is affected by digitalization. Digitalization enables companies to systematically collect, analyse and draw conclusions from data almost in real time. Many companies achieve large profits with business models which are based on the collection and commercial use of (often personal) data. Google's search engine and Facebook's social network are two prominent examples. But data collection is not only crucial for search engines, social networks or online advertising. It became a key issue even for traditional industries such as energy, telecommunications, insurance[3], banking or transport[4].

Data-based business models certainly have a very significant economic potential. Collecting and analyzing data on a large scale can help

[1] Former Chairman of the Autorité de la concurrence.

[2] Chairman of the German Bundeskartellamt.

[3] See, e.g., UK Financial Conduct Authority, Call for Inputs: Big Data in retail general insurance (2015), https://www.fca.org.uk/news/call-for-inputs-big-data.

[4] See, e.g., Comité du débat sur l'ouverture des données liées à l'offre de transport, presided by Francis Jutand, report (2015), http://www.ladocumentationfrancaise.fr/rapports-publics/154000182/ – only available in French.

make our economies work better and be more efficient. Companies can create new products and services that were not existent before. Data and technological changes also affect traditional industries and can significantly improve companies' products and services. It can allow companies to optimize their production and sales activities. It becomes easier to identify what customers really need. Consumers can benefit from better services and recommendations on what to buy. At the same time the costs of production and distribution can be reduced.

On the other hand the collection, processing and commercial use of data can collide with public or private interest and raises questions. The issues most often discussed are probably not questions of competition law but rather issues that concerns data protection enforcement. However, in several recent proceedings competition authorities in Europe have begun to look at possible competition law issues arising from the possession and use of data.

In 2016 the French Autorité de la concurrence and the German Bundeskartellamt have jointly analysed which consequences and challenges arise out of the collection of data in the digital economy and other industries. The aim of the joint study was to feed the debate by identifying some of the key issues and parameters that can become relevant when assessing the interplay between data, market power and competition law.

This article gives an overview of the key findings of the joint study.[5] First, as a starting point, we briefly describe ways in which data is collected and analysed. Second, we analyse the various theories of harm usually associated with data collection and exploitation in digital markets. Third we discuss some parameters that are to be considered in assessing the relevance and credibility of these theories of harm.

2. DEFINITION OF "DATA" AND "BIG DATA"

To structure the debate about the interplay between data and competition law, it appears useful to first define what is meant by "data" and "big data" and to look at ways in which data is collected, processed and used by companies.

For the term "data", there is no single definition. In a wider sense the term is used to refer to (any) information, or to the representation of such information, often in combination with it being stored on a computer. In the context of many online business models the term "data" often refers to personal data[6] – providing information about

[5] The article is based on the joint study "Competition Law and Data", published by the Autorié de la Concurrence and the Bundeskartellamt on 10 May 2016, http://www.bundeskartellamt.de/SharedDocs/Publikation/DE/Berichte/Big%20Data%20Papier.pdf?__blob=publicationFile&v=2.

[6] In the view of European data protection agencies, the scope of personal data extends to data which can be

an individual's characteristic traits, preferences, geographic locations etc. Such personal data are usually subject to special data protection rules, which limit the gathering, processing and usage of data in order to ensure consumer privacy. But the data collected and used by companies is not necessarily restricted to personal data about individuals. Companies also make use of data about economic entities and objects, e.g. such as specific information about competitors, business strategies and business transactions or the status of operation of certain machines. However, so far the current discussion concerning data and competition law focusses mostly on personal data.

The buzzword of choice in the current debate concerning competition law and digitalization, however, is not simply "data" but the term "big data" – a concept lacking a common definition.[7] It is often used to describe the recent developments of collection and utilization of data in the digital economy. Aspects of "big data" that are often mentioned in the debate are large amounts of different types of data, produced at high speed from multiple sources, whose handling and analysis require new algorithms, new and more powerful processors, storage and data transport technology .[8] In a shorter form "big data" is often characterized by the three "V"s – Velocity, Variety and Volume – or the four of them adding "Value".[9]

As regards data collection by companies, there are different ways to gather data. Personal data are often provided voluntarily by consumers, e.g. in social networks or online shops. Customers provide information in exchange for – often zero-priced and advertising financed – products and services. Thereby the companies get information not only about address, email-contact, date of birth or payment details, but also about shopping preferences or in some cases even photos or videos. Another possibility to learn about the consumers' interests and preferences is to observe their behavior on the internet and

assigned to IP addresses and cookies, even if the name as such of the user is not identified, see Article 29 Data Protection Working Party, Opinion 1/2008 on data protection issues related to search engines, dated 4 April 2008, http://ec.europa.eu/justice/policies/privacy/workinggroup/wpdocs/index_en.htm, pp.6-8, and Opinion 4/2007 on the concept of personal data, dated 20 June 2007, see link above, pp.16-17.

[7] See Hu, Han et al., Toward Scalable Systems for Big Data Analytics: A Technology Tutorial, IEEE Access, Vol. 2 (2014), http://ieeexplore.ieee.org/xpl/articleDetails.jsp?reload=true&arnumber=6842585, p. 652; Gil Press, 12 Big Data Definitions: What's Yours?, Forbes, http://www.forbes.com/sites/gilpress/2014/09/03/12-big-data-definitions-whats-yours; The Big Data Conundrum: How to Define It?, MIT Technology Review, http://www.technologyreview.com/view/519851/the-big-data-conundrum-how-to-define-it/ .

[8] European Data Protection Supervisor, https://secure.edps.europa.eu/EDPSWEB/edps/Consultation/big_data .

[9] See Hu, Han et al., Toward Scalable Systems for Big Data Analytics: A Technology Tutorial, IEEE Access, Vol. 2 (2014), http://ieeexplore.ieee.org/xpl/articleDetails.jsp?reload=true&arnumber=6842585, p. 652 (654); German Monopolies Commission (Monopolkommission), Special Report No. 68: Competition policy: The challenge of digital markets (2015), http://www.monopolkommission.de/images/PDF/SG/s68_fulltext_eng.pdf, § 67.

assess which web pages they visit.[10] Companies can also infer new information about consumers using already existing data, e.g. drawing conclusions about gender or age by analyzing the consumers' shopping activities. All these possibilities to collect data are usually referred to as "first party data", because the company itself deals with the data collection and is related to their own customers. Besides there is the possibility to use "third party data". This means that another entity collected the data and shares or sells these data.

For companies, which collect data in different business units or subsidiaries it is possible to combine these data from different sources in a big pool. In this way the companies are able to cross-connect information and create sophisticated customer profiles.

Finally it is worth noting that the importance of data for companies is not a new phenomenon and not restricted to digital markets. Data has always been important input factor for companies in many traditional markets – e.g. to improve their products and services or to better target their products customers or to sell advertising space. However, the technological changes induced by digitalization have significantly improved the possibilities to collect, process and commercially use data in almost every business sector. In many cases that increases economic efficiency. However, it also has the potential to raise competition concerns.

3. POSSIBLE ROLE OF DATA IN THE COMPETITIVE ANALYSIS

As described above, data and technological changes of the digital economy can have very significant economic advantages. Thereby the collection and use of data can create economic efficiencies and can have pro-competitive effects.

However, under certain circumstances, the collection and analysis of data can be a factor contributing to competition concerns. The debate usually circles around three aspects: 1) Data can be a factor contributing to market power. 2) Data can increase market transparency among suppliers and thereby facilitate collusion. 3) Data can be an instrument for certain anticompetitive conducts.

3.1 Data as a source of market power

Data can be a factor contributing to market power. If – and this is a crucial precondition – the access to data is important to compete on the markets, the possession of or the access to data can constitute a barrier for market entry if new entrants are unable to either collect similar data or to buy access to sources of such data as the incumbents. Of course this has to be assessed on a case-by-case basis.

[10] That it is technically possible to monitor which part of a web page a user actually sees has prompted demands by advertisers of a minimum "viewability" of display ads to measure audience (web pages are generally larger than the screens they are viewed on). See Invisible ads, phantom readers, The Economist, 26 March 2016, http://www.economist.com/news/business/21695388-worries-about-fraud-and-fragmentation-may-prompt-shake-out-crowded-online-ad.

For example big, established online services such as search engines or online shops are able to collect very huge amounts of data generated by transactions and interactivity with their users or (prospective) customers due to their large user or customer base. To be clear: holding a large amount of data is not a problem in itself. But smaller companies or new entrants might in the specific case not be able to collect data to the same extent as the incumbents as they usually have less or even no users or transactions. While such companies principally could also use third-party data, i.e. data collected by another entity, in practice there may be limits to the availability of third party data. Data usually is highly differentiated and data collected by third parties can have less value to a company than data generated by its own interaction with users and customers. Also, third parties might not always be willing to share data or sell data to its competitors.

In addition, markets where the collection and use of data are often seen as important, such as search engines or social networks, are often highly concentrated. In that context, the existence of strong scale and network effects in these cases is also described as limiting the intensity of competition. The development of data collection and usage on those markets may thus reinforce the market power of leading companies on these markets. Also, the marginalization of smaller competitors due to differentiated data access might be self-reinforcing: access to a larger amount of data may support better services, which in turn attract more customers – and more data ("snowball effects"). By contrast, smaller companies might attract fewer consumers and as a result have less data. As the gap in market share increases, so might the gap in data collection, which could further increase the gap in the quality of services proposed to customers. Finally, higher revenues earned by larger undertakings could fuel higher investments (such as new algorithms, new functionalities, entry on adjacent markets, etc.), thereby attracting even more customers and more data. Such a trend could harm competition by ultimately converging towards a monopolization of data-related markets.

3.2 Data, market transparency and competition

Data can increase market transparency. It can increase transparency between suppliers and customers which usually has positive effects on competition. Consumers can benefit from higher market transparency. It can reduce information asymmetries between consumers and suppliers. Consumers can compare prices and view ratings in a simple way, e.g. at platforms like TripAdvisor or market places like Ebay or Amazon. As a result competition can get more intensive with regard to both price and quality. In addition transparency can ease the access by new providers of information about customer requirements and competing offers, thereby lowering the cost of market entry.

On the other hand, data collection increases transparency between suppliers. In certain cases that can limit competition. Particularly concentrated markets with a high transparency among suppliers can be vulnerable to (tacit or explicit) collusion. In transparent markets it is easy to reveal deviations from an (tacit or

explicit) agreement. Furthermore it may become easier to agree on collusive prices, adapt collusive prices to changes or to detect deviations by using analytical algorithms on that data.

3.3 Data-related anticompetitive conducts

The collection and access to data can be an instrument for certain anticompetitive conducts.

First, the collection and access to data can raise concerns in the assessment of merger cases. For companies a merger can be a possible strategy to obtain access to new data by acquiring or merging with a – often small and innovative – company that possesses large amounts of relevant data. In data-related markets such a merger could increase the concentration of relevant data and restrict entry and expansion for new companies. For example, in its Facebook/WhatsApp decision, the European Commission examined the impact of the merger on the possibility of data access for the purpose of subsequent utilization for the advertising market. Here, the Commission not only explored the status quo, which is that WhatsApp has not collected any user data, but over and above this examined the hypothetical case in which WhatsApp were to commence collecting user data. The Commission found here that, even in the latter case, there would be no concentration of advertising-relevant data on Facebook which would be problematic in terms of competition law, since many other companies were also collecting extensive data. The merger was ultimately approved without competitive concerns.[11]

Furthermore, a merger in data-related markets can also give rise to vertical or conglomerate effects if the merger increases the ability and incentive of a large company to restrict up- or downstream competitors' access to data. In any of these scenarios, competition concerns are more likely the more difficult it is for competitors to replicate the information extracted from the relevant data.

Second, competition can be restricted, if the access to data is restricted by a dominant company in an anticompetitive manner. Such a restriction can evolve in different situations. A refusal to access to data to a competitor can be anticompetitive if the data are considered as an "essential facility" to the activity of the company requesting access. However, the requirements of the European Court of Justice to classify data as an essential facility are relatively strict.[12] Among other preconditions, it would have to be demonstrated that data owned by the dominant company is truly unique and there is no possibility for the

[11] European Commission, Decision of 3 October 2014, M.7217 –Facebook/WhatsApp, paras. 164 ff ; German Monopolies Commission (Monopolkommission), Special Report No. 68: Competition policy: The challenge of digital markets (2015), http://www.monopolkommission.de/images/PDF/SG/s68_fulltext_eng.pdf, § 109

[12] ECJ, „IMS Health", C-418/01, judgment of 29.04.2004, §§ 34-52, GC, „Microsoft", T-201/04, judgment of 17.09.2007, §§ 320-336, ECJ, "Bronner", C-7/97, judgment of 26.11.1998, §§ 44-45.

competitor to obtain the data needed to perform its service.[13]

A refusal to allow access to data could also be anticompetitive if it is discriminatory, i.e. if a dominant company grants access to certain customers while denying access to customers of a downstream competitor.[14] More generally, vertical integration can entail discriminatory access to strategic information with the effect of distorting competition. For instance, some market place operators also operating as online retailers may get access to information about their competitors selling on that market place and about the behaviour of consumers. By identifying the range of products that are globally more in demand, an integrated platform could then be able to more efficiently adjust the range of products it sells as well as the pricing of its products. A similar effect could be achieved by such a platform, if it restricted the information that their competitors operating on the marketplace get about the transactions they are involved in. Such information transfers and limitations could make the integrated platform operator more competitive than its competitors operating on its market place.

Anticompetitive data-driven strategies may also include preventing rivals from accessing data through exclusivity provisions with third-party providers or foreclosing opportunities for rivals to procure similar data by making it harder for consumers to adopt their technologies or platforms. And finally, data collected on a given market could be used by a company to develop or to increase its market power on another market in an anti-competitive way, e.g. by way of tied sales whereby a company owning a valuable dataset ties access to it to the use of its own data analytics services. [15]

Third, data can also be a vehicle to facilitate price discrimination.[16] The more a company knows about the preferences and reservation prices of its customers, the better it can adapt the prices to individual customer groups.

Price differentiation by itself does not necessarily raise competition concerns. Setting differentiated prices is a key element of competition. Customers with low reservation prices can benefit from price differentiation as they can buy a product at a lower price. Some customers might not even have afforded the

[13] Damien Geradin and Monika Kuschewsky, Competition law and personal data: preliminary thoughts on a complex issue (2013), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2216088 , p. 15.

[14] French Competition Authority, Decision n° 14-D-06, dated 08.07.2014, relative à des pratiques mises en œuvre par la société Cegedim dans le secteur des bases de données d'informations médicales. This decision has been confirmed on appeal but is still pending in front of the Cour de Cassation (the French Supreme Court).

[15] Competition and Markets Authority, The Commercial Use of Consumer data (2015), https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/435817/The_commercial_use_of_consumer_data.pdf, p.90.

[16] See Nathan Newman, The Costs of Lost Privacy: Consumer Harm and Rising Economic Inequality in the Age of Google, 40 WM. MITCHELL L. REV. pp. 850 (865-873), available at http://open.wmitchell.edu/cgi/viewcontent.cgi?article=1568&context=wmlr.

product at a higher price. Thus, the social welfare may increase compared to the situation without price discrimination. However, price discrimination can also have negative effects. Price discrimination can increase the information asymmetry between consumers and suppliers, resulting in higher search costs for consumers. For some consumers price discrimination may lead to higher prices. Economic evidence suggests that price discrimination can lead to higher prices for "naïve" (non-rational) consumers, while rational consumers are not or less harmed by price discrimination.[17]

Thus, one may query whether price discrimination in itself is within the scope of European competition law. To be considered as either an abuse of a dominant position or a vertical restraint, its effects on competition, and not only on consumer welfare, would have to be demonstrated; the absence of any proportionate justification would also have to be ascertained. The situation under national competition law, however, might be different. Provisions governing unilateral conduct may indeed be stricter.[18]

Finally, in data-related market there can be a close link between the effects of data on market power and privacy concerns. The supervision of data protection is basically not the responsibility of competition authorities. Nevertheless, data protection can possibly be an issue of competition law. In German Competition law, the German Federal Court of Justice has stated that contract terms which are incompatible with the laws regulating general conditions and terms of trade might be an abuse of a dominant position if the use of the terms is based on the company's market dominance.[19] In that respect, data protection regulations can be subject of competition proceedings. This can particularly be the case, if the suspected violation of data privacy is committed by a dominant company whose business model is essentially based on the collection and analysis of data.

4. ASSESSING BIG DATA AND MARKET POWER

As indicated before, several competition issues associated with data concern the question of market power being supported by the collection and use of data. If this question is

[17] Heidhus and Köszegi (2016), Naivité-Based Discrimination, The Quarterly Journal of Economics; Heidhus and Köszegi (2014), Using Information to Price Discriminate, Working Paper.

[18] For instance, in Germany, the Federal Supreme Court has stated that the provision against the abuse of a dominant position can include a consumer protection dimension as regards price discrimination, see German Federal Supreme Court (BGH), "Entega II", KZR 5/10, judgment of 07.12.2010, available via http://juris.bundesgerichtshof.de/cgi-bin/rechtsprechung/list.py?Gericht=bgh&Art=en&Datum=Aktuell&Sort=12288.

[19] German Federal Court of Justice (Bundesgerichtshof), "VBL-Gegenwert", KZR 61/11, judgment of 16.11.2013, available via http://juris.bundesgerichtshof.de/cgibin/rechtsprechung/list.py?Gericht=bgh&Art=en&Datum=Aktuell&Sort=12288, § 68.

not new for competition enforcers - i.e. data is important for undertakings whether they are operating on an online or offline market -, it has however taken a whole new dimension in the digital markets with the recent advent of Big Data. The massive collection and use of data which are allowed by technological progress requires a renewal in our analysis of the possible contribution of data to market power, taking into account typical key specificities of online services.

4.1 Sources and limits of market power in data-driven online industries

The first main characteristic of digital markets is that they often involve so-called multi-sided markets, allowing undertakings to be active toward more than one group of users/customers and to subsidize services to one group of users by extracting revenue from another group of users.

This widespread characteristic may have implications for market definition, raising in particular the question whether a relevant market can be defined when services are provided for free, notwithstanding the fact that competition is in any event at work on non-price factors. Moreover, these multi-sided markets are characterized by direct and/or indirect network effects, i.e. the use of a good or service by a user directly or indirectly impacts the value of that product or service to other users. These network effects may favor market concentration, by their self-strengthening nature and the corresponding raise in entry barriers. At the same time, they may stimulate competition by making possible the swift and effective entry of a new market participant on the back of a product or service innovation. Whether network effects are ultimately beneficial or detrimental to competition will depend on various parameters, including the level of fixed costs needed to attract a sufficient large number of users to implement positive network effects, or the difference in undertakings' market shares. In this context, data can be seen as reinforcing traditional network effects, for better or worse.

Another characteristic of digital markets is the basic possibility for users to multi-home, i.e. the ability to resort to different service providers. Multi-homing can facilitate the collection of similar data by a plurality of undertakings and is consequently often seen as a factor likely to reduce market power[20]. However, there are some limits to this assertion. A situation of perfect multi-homing is rather rare and switching costs often prevent consumers from using various providers in equal proportion. Especially in a market where services are provided for free, consumers may pay more attention to quality and if these markets are characterized by network and experience effects, new entrants may not be able to provide the same level of quality as incumbents. Furthermore, even if switching costs are low, what matters when considering data-based markets is the actual level of multi-homing, as the critical mass of relevant data

[20] David S. Evans and Richard Schmalensee, The Industrial Organisation of Markets with Two-Sided Platforms, Competition Policy International, 2007, Vol. 3, Nr. 1, pp. 151-179; Mark Armstrong, Competition in two-sided markets, RAND Journal of Economics, Vol. 37, Nr. 3, pp. 668-691.

may oftentimes only be obtained through the regular use – rather than a sporadic use – of a given service. And finally, the interplay of e. g. multi-homing (to some extent) on one side of a two-sided market and single-homing (to some extent) on the other side has to be considered when analyzing the effects in a specific case.

Finally, when considering digital markets, it is often argued that these are characterized by rapid changes, strong innovation and easy entry of new challengers. Experience (e.g. displacement of Yahoo! By Google) has indeed shown that established market power is a relative concept when assessing these markets. However, it would be a mistake to consider that lessons learnt from past experiences can be relied on to solve present or future concerns. Different parameters, such as the level of barriers to entry (i.e. research and development expenses, tangible assets, marketing expenses, etc.) but also the role of data collection should be taken into consideration on a case-by-case basis to buttress or reject the claim that (dynamic) competition can, in and of itself, act as a credible disciplining factor. Indeed, the emergence of all-encompassing ecosystems, which provide multiple services and buy out new entrants as they begin to thrive, should lead competition enforcers to envisage with renewed caution arguments turning around the ease of displacement in a dynamic competitive environment.

4.2 Evaluation of the "data-advantage" in past cases

Recent cases have given many occasions, both in digital and non-digital sectors, to assess the competitive advantage that data collection and use may confer.

It is historically in non-digital sectors that competition enforcers have tackled this issue. In its opinion of 14th June 2010 on the cross-usage of customer datasets, the Autorité detailed its criteria to determine when the use by a company of a dataset on its users to develop its activity on another market is anticompetitive. This assessment revealed in particular the importance of the conditions under which the dataset was constituted, whether it could be replicated under reasonable conditions by competitors and was likely to result in a significant competitive advantage. The Autorité applied this analytical grid to various cases concerning former public monopolies – in particular in the energy sector (gas and electricity) – or involving practices by dominant private undertakings which were not former public monopolies[21].

Even if, in the future, as a result of the growth of connected devices, separation between online and offline markets concerning the use and importance of data may be mitigated, it remains for the time being in the digital sector that the question of the advantage arising from the use of data is growing in relevance. Very often, the purpose of data in digital markets is to improve the quality of service to customers – more accurate and relevant search results, information, recommendations or advertising.

[21] See case 14-D-06 http://www.autoritedelaconcurrence.fr/user/standard.php?id_rub=592&id_article=2403.

Several merger decisions[22] have exposed the role of data in providing better services.

Although the European Commission noted in its merger decisions the role that the combination of several kinds of data from different sources could play in driving the efficiency of a service or of the ads targeting, it concluded in each of these cases that the data advantage potentially enjoyed by the new entity did not lead to any competition issue. It underlined in particular the practical difficulties that such combination could face, especially potential reluctance from client advertisers or even from users[23], and that, even if it could be implemented, it was unlikely that such an advantage could not to be matched by competitors. Furthermore, the Commission considered that possible post-merger combinations of data did not raise any competition concerns to the extent that there would still be, after the operation, a large amount of internet user data to draw from.

These precedents help to delineate two relevant factors when considering whether data can contribute to market power: the scarcity of data (or ease to replicability) and the impact of the scale and scope of data collection on competitive performance.

4.3 Issues pertaining to the scarcity of data

The main issue when considering the competitive impact of the ownership of a dataset by one undertaking is whether a similar dataset, in terms of scale and scope, is potentially accessible to an as efficient undertaking. Three features may contribute to a high availability of data and therefore reduce the risk that differentiated access to these data between competitors could harm competition.

Firstly, data are non-rival goods. Hence, the collection and use of one particular dataset does not prevent other undertakings from constituting and using the same data (provided they can access them)[24]. Indeed, customers can provide similar data to different services providers whether they are competitors - in particular in the case of multi-homing – or not. Moreover, potential customers of data brokers can threaten them with collecting these data themselves thereby contributing to limit the price of data. However, non-rivalry of data does not necessarily mean that it is accessible to all undertakings. Indeed, access to data may, at the same time, be subject to different types of costs such as significant investments to collect data (construction and development of data centers), to build a sufficiently large customer base by proposing complete or innovative services (R&D expenses, acquisition costs, etc.) and finally to analyze the data

[22] See the European Commission's decisions Google/DoubleClick, Facebook/WhatsApp or more recently Microsoft/LinkedIn.

[23] See e.g. the European Commission decision Microsoft/LinkedIn, §248.

[24] Nils-Peter Schepp and Achim Wambach, On Big Data and its Relevance for Market Power Assessment, Journal of European Competition Law & Practice, 2016, Vol. 7, No. 2, p. 121.

(development of specific algorithms). In addition, some legal constraints can also limit the accessibility of data for some specific uses, especially personal data (i.e. preventing the communication of personal data to third parties).

Data are not solely collected by companies as an input to gain a competitive edge over rivals. They can also be collected to be resold by data intermediaries. Data brokers rely on various sources of data, including their own data collection technology, using tracking technologies (cookies or pixels), public information, and data from public authorities and third-party companies. They can constitute a credible alternative for undertakings which may in some cases be less costly than in-house data collection, especially for new entrants who may need to quickly obtain a large volume of data. On the other hand, data provided by third parties may present important drawbacks. They can be imperfect substitutes to directly obtained data which may have a larger scope, in particular in the case of complex data obtained by major companies which may cross-reference data from different services. Past decisions have also underscored that mergers aiming at integrating users and providers of data, for instance navigational data, can yield efficiencies which cannot be obtained from third-party data procurement.[25] Special technical costs associated with third-party sourcing may also arise, especially if the data needs to be updated frequently or even in real-time to remain relevant. Finally the incentive to share data with competitors from operators active in the market may be insufficient, as the OECD underlined in its study of "data-driven innovation".[26]

With the ever-growing digitization of the economy and connectivity of goods and devices, the volume of data that can be collected has never been as high as today. Data is arguably everywhere and the real value of data will in fact significantly depend on the knowledge that can be extracted from it. Indeed, if the same kind of data can be extracted from different datasets, obtained through different mechanisms, then the competitive advantage conferred by this dataset would probably be low. Undertakings, especially if there are not competitors, can benefit from diversified sources of data and may agree to share and cross-exchange the data they collect in order to increase their level of information. However, one has to keep in mind that this *prima facie* massive availability needs to be assessed concurrently with its actual accessibility – as pointed out before. Finally, if competition authorities have considered in the recent merger decisions that the amount of accessible data will remain sufficient, it does not imply that all types of data are substitutes for one another.

[25] See the European Commission decision in Tom-Tom/Tele-Atlas.

[26] 2015 OECD study, "Data-driven innovation" – Big Data for growth and well-being, p. 192.

4.4 Issues pertaining to the scale and scope of data collection

The competitive advantage associated with data may depend on whether the said data need to be collected on a large scale and/or scope, which possibly could only be attained by a large or/and diversified incumbent. The importance of scale and scope of data has been the object of intense debates.

First of all, as underlined by Lerner (2014)[27], the marginal value of data used for inference purposes can decrease rapidly once a certain amount of data has been collected. Such a circumstance, particularly for some specific uses of data, like the feeding of a search engine's algorithm or, more generally, when data is used as an input to make prediction, may significantly nuance the competitive advantages resulting from large amount of data. The statistical sampling error associated with any extrapolation from a dataset always decreases if the size of the dataset increases, but at a rate that keeps decreasing with sample size. E.g. in the case of a search engine, the more observations are collected, the more the sample of observations will be reliable. At the same time, as the number of searches increases, the marginal information value of each search query will decrease. This may mean that the number of necessary searches for an internet search engine to be competitive may be lower than the actual level of searches accumulated by a large platform such as Google.

Moreover, the ability to extract information from data does not rely exclusively on the volume of data collected, but also on the technical ability to analyze that data and to separate the wheat from the chaff, as collected data are not all of the same quality. The role of data analytics will thus be central in order to determine whether a small-sized or a poor dataset can be compensated (or exacerbated).

Finally, another limiting factor of the advantage enjoyed by larger incumbents relative to new entrants is that the value of data may decrease quickly in time. Indeed, historical data may have a relative low value for some operators such as the search engines.

The apparent willingness of some companies to collect, combine, store and use ever greater volume of data or even their external growth strategies toward user data are sometimes considered as proof that scale and scope of data are a key parameter of competition. However, such strategies could merely indicate that the marginal costs of collecting data are very low. Thus, unless large fixed costs or legal risks are incurred when processing these supplementary data, the ever larger volumes of collected data are not necessarily the proof that a large scale and scope of data collection confers market power.

One should keep in mind that the benefits derived from a large scale and scope of data will greatly depend on market circumstances, the type of data concerned and the applications foreseen. For example data obsolescence will vary greatly: sociodemographic data for

[27] Andres V. Lerner, The Role of 'Big Data' in Online Platform Competition (2014), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2482780.

instance may prove valuable even as months and years elapse. Similarly, the notion that data exhibits a decreasing marginal value is true if used for inference purposes, but data is used for a wide array of purposes, above and beyond inference. Moreover, decreasing returns to scale are one thing, but should the threshold beyond which this decrease is observed prove high, then ample room remains for market power to derive from the sheer volume of data detained. Finally, with regard to services for which demand shifts rapidly and which require constant updating or information on niche queries – such is the case presumably of search engine services –, a steady influx of new and varied data will be required, which in turn requires a large and loyal customer base. All these characteristics need to be considered, and pronouncements on whether the volume of data matters can only result from a case-by-case approach.

5. THE FACEBOOK INVESTIGATION OF THE BUNDESKARTELLAMT

In March 2016 the Bundeskartellamt has initiated a proceeding against Facebook. The Bundeskartellamt is investigating if Facebook has abused its possibly dominant position in the market of social networks. There is an initial suspicion that Facebook's conditions of use violate data protection provisions. First it has to be investigated whether Facebook has market power or a dominant position in a possible market of social networks. In this context the findings described above on data as a factor of market power are very valuable. Second, it has to be assessed whether Facebook's specific terms of service on the use of user data constitute an infringement of German competition law. Here the analysis of possible theories of harm mentioned above come into play.

The Bundeskartellamt is conducting the proceeding in close contact with the competent data protection officers, consumer protection associations as well as the European Commission and the competition authorities of the other EU Member States.

6. SECTOR INQUIRY: DATA-RELATED MARKETS AND STRATEGIES

On 23 may 2016, the Autorité de la concurrence launched a sector inquiry into online advertising, in order notably to assess how significant data processing has become in this industry. Six years after a previous inquiry that established Google's dominance in the market for online "search" advertisement, the focus is now on online display advertisement, in relation to the growth of real-time bidding and the role of data in managing advertising campaigns. The inquiry seeks to disentangle the webs of a complex ecosystem, characterized by growing integration and risks of conflict of interest. The Autorité is assessing whether there is room for independent intermediaries in this industry as major players tend to integrate throughout the entire advertisement supply chain, with the risk that this integration can create input and/or customer foreclosure.

The Autorité is looking at three main sets of issues: data and the definition of relevant markets with particular reference to the substitutability between the different forms of targeted advertising; data and market power, and especially the weight and strategies of operators such as Google and Facebook and the place of data in the exercise of this market power; and data and business practices, and whether merit-based competition may be hindered.

The opinion of the Autorité, to be issued in 2017, will be preceded by a far-reaching public consultation aimed at gathering the observations of all stakeholders (advertisers, publishers, data brokers, etc.).

7. Conclusion

Big data has a very significant potential to improve services and products. Effective enforcement of competition rules can ensure that stakeholders are getting the best out of this potential.

The collection and use of data by companies in order to optimize their business, e.g. the use of customer databases, has already been common practice for a long time. Really new are the developments of the last few years, made possible by the fast technological progress, increased digitalization and connectivity, which have radically extended the type, volume and sources of data. The economic relevance of data has significantly increased. Competition authorities have to take these developments into account when assessing market positions and market behavior. Due to the different and sometimes ambiguous effects of data utilization described above, the market situation should always be individually assessed on a case-by-case basis. In this process the characteristics of digital markets like network effects, multi-homing and market dynamics have to be considered.

The examples of the proceedings of the Bundeskartellamt against Facebook and the sector inquiry of the Autorité de la concurrence into online advertising markets illustrate how competition authorities respond to these new developments. Obviously, competition authorities are still in the early stages of coping with these new phenomena and their approach has to be constantly refined. In this process a close cooperation of competition authorities throughout Europe, together with growing interactions with other concerned regulators such as data protection agencies, stand to benefit all. In doing so the authorities are well-positioned to deal with the challenges of the digital economy and big data.

8. BIBLIOGRAPHY

Autorité de la Concurrence and Bundeskartellamt, joint study «*Competition Law and Data*», http://www.bundeskartellamt.de/SharedDocs/Publikation/DE/Berichte/Big%20Data%20Papier.pdf?__blob=publicationFile&v=2, 10th May 2016.

Mark Armstrong, Competition in two-sided markets, RAND Journal of Economics, pp. 668-691, Vol. 37, Nr. 3.

Comité du débat sur l'ouverture des données liées à l'offre de transport, http://www.ladocumentationfrancaise.fr/rapports-publics/154000182/, 2015.

Competition and Markets Authority, *"The Commercial Use of Consumer data"*, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/435817/The_commercial_use_of_consumer_data.pdf , p.90, 2015.

Court of Justice, C-7/97 Bronner, 26th November 1998.

Court of Justice, C-418/01 IMS Health, 29th April 2004.

Data Protection Working Party, Opinion 1/2008 on data protection issues related to search engines, pp.6-8, 4th April 2008, and Opinion 4/2007 on the concept of personal data, pp.16-17, 20th June 2007, http://ec.europa.eu/justice/policies/privacy/workinggroup/wpdocs/index_en.htm.

David S. Evans and Richard Schmalensee, The Industrial Organisation of Markets with Two-Sided Platforms, Competition Policy International, pp. 151-179, Vol. 3, Nr. 1, 2007.

European Commission, M.4731 Google/DoubleClick, 11th March 2008.

European Commission, M.4854 Tom-Tom/Tele-Atlas, 14th May 2008.

European Commission, M.7217 Facebook/WhatsApp, 3rd October 2014.

European Commission, M.8124 Microsoft/LinkedIn, 6th December 2016.

European Data Protection Supervisor, https://secure.edps.europa.eu/EDPSWEB/edps/Consultation/big_data.

French Competition Authority, Decision n° 14-D-06, 8th July 2014.

General Court, T-201/04 Microsoft, 17th September 2007.

Damien Geradin and Monika Kuschewsky, *Competition law and personal data: preliminary thoughts on a complex issue*, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2216088 , p. 15, 2013.

German Federal Court of Justice (Bundesgerichtshof), "*VBL-Gegenwert*", KZR 61/11, http://juris.bundesgerichtshof.de/cgibin/rechtsprechung/list.py?Gericht=bgh&Art=en&Datum=Aktuell&Sort=12288, 16th November 2013.

German Federal Supreme Court (BGH), "*Entega II*", KZR 5/10, http://juris.bundesgerichtshof.de/cgi-bin/rechtsprechung/list.py?Gericht=bgh&Art=en&Datum=Aktuell&Sort=12288, 7th December 2010.

German Monopolies Commission (Monopolkommission), Special Report No. 68: Competition policy: The challenge of digital markets, http://www.monopolkommission.de/images/PDF/SG/s68_fulltext_eng.pdf, § 109, 2015.

Gil Press, 12 Big Data Definitions: What's Yours?, Forbes, http://www.forbes.com/sites/gilpress/2014/09/03/12-big-data-definitions-whats-yours.

Heidhus and Köszegi, *Naivité-Based Discrimination*, The Quarterly Journal of Economics, 2016.

Heidhus and Köszegi, *Using Information to Price Discriminate*, Working Paper, 2014.

Hu, Han et al., *Toward Scalable Systems for Big Data Analytics: A Technology Tutorial*, IEEE Access, http://ieeexplore.ieee.org/xpl/articleDetails.jsp?reload=true&arnumber=6842585, p. 652, Vol. 2 (2014).

Andres V. Lerner, *The Role of 'Big Data' in Online Platform Competition*, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2482780, 2014.

MIT Technology Review, The Big Data Conundrum: How to Define It?, http://www.technologyreview.com/view/519851/the-big-data-conundrum-how-to-define-it/.

Nathan Newman, *The Costs of Lost Privacy: Consumer Harm and Rising Economic Inequality in the Age of Google*, 40 WM. MITCHELL L. REV. pp. 850 (865-873), http://open.wmitchell.edu/cgi/viewcontent.cgi?article=1568&context=wmlr.

Nils-Peter Schepp and Achim Wambach, O*n Big Data and its Relevance for Market Power Assessment*, Journal of European Competition Law & Practice, No. 2, p. 121, Vol. 7, 2016.

OECD study, Data-driven innovation – Big Data for growth and well-being, p. 192, 2015.

The Economist, http://www.economist.com/news/business/21695388-worries-about-fraud-and-fragmentation-may-prompt-shake-out-crowded-online-ad, 26th March 2016,

United Kingdom's Financial Conduct Authority, Call for Inputs: Big Data in retail general insurance, https://www.fca.org.uk/news/call-for-inputs-big-data, 2015.