

Competition Law and Data

10th May, 2016

I.	INTRODUCTION.....	3
II.	DATA.....	4
1.	WHAT IS MEANT BY “DATA”?.....	4
2.	TYPES OF DATA.....	5
a)	<i>Type of information.....</i>	5
b)	<i>Structured vs. unstructured data</i>	6
c)	<i>Ways to gather data</i>	6
3.	TECHNOLOGICAL PROGRESS	7
4.	ROLE OF DATA IN ECONOMIC ACTIVITIES	8
a)	<i>Improvements to products or services</i>	9
b)	<i>Exploitation of new business opportunities</i>	10
c)	<i>More target-oriented business models</i>	10
III.	POSSIBLE ROLE OF DATA IN THE COMPETITIVE ANALYSIS	11
1.	DATA AS A SOURCE OF MARKET POWER	11
2.	DATA, MARKET TRANSPARENCY AND COMPETITION.....	14
3.	DATA-RELATED ANTICOMPETITIVE CONDUCTS.....	15
a)	<i>Mergers and acquisitions</i>	16
b)	<i>Exclusionary conducts</i>	17
c)	<i>Data as a vehicle for price discrimination</i>	21
d)	<i>Data, market power and privacy concerns</i>	22
IV.	ASSESSING BIG DATA AND MARKET POWER	25
1.	SOURCES OF MARKET POWER IN DATA-DRIVEN ONLINE INDUSTRIES	26
a)	<i>Multisided markets and network effects</i>	27
b)	<i>Multi-homing</i>	28
c)	<i>Market dynamics.....</i>	29
2.	EVALUATION OF THE “DATA ADVANTAGE” IN PAST CASES	31
a)	<i>In “non-digital” markets.....</i>	31
b)	<i>In “digital” markets.....</i>	33
3.	ISSUES PERTAINING TO THE SCARCITY OF DATA	36
a)	<i>On the non-rivalry of data.....</i>	36
b)	<i>The availability of digital data</i>	42
4.	ISSUES PERTAINING TO THE SCALE AND SCOPE OF DATA COLLECTION	47
a)	<i>Presentation.....</i>	47
b)	<i>Discussion.....</i>	49
V.	CONCLUSION.....	52

Competition Law and Data

I. Introduction

The collection, processing and commercial use of data is often seen not as a competition law issue but rather as an issue which concerns data protection enforcement. However, several recent proceedings point to the fact that competition authorities have begun to look at possible competition issues arising from the possession and use of data, even if, in the end, none were ascertained in the specific cases.¹

Recent developments in digital markets have led to the emergence of a number of firms that achieve extremely significant turnovers based on business models which involve the collection and commercial use of (often personal) data. Some of them enjoy a very high share of users in the service sector in which they are active. The Google search engine and the Facebook social network are probably the most prominent examples. While many of the services provided by these firms are marketed as 'free', their use involves in practice making possible the collection of personal information about the users. This has spurred new discussions about the role of data in economic relationships as well as in the application of competition law to such relationships, in particular as regards the assessment of data as a factor to establish market power. It is important to note that although these questions are often examined with the examples of Google and of Facebook in mind, they are also relevant for many other industries. Indeed, the development of data collection already goes well beyond search engines, social networking or online advertising and extends today to sectors such as energy, telecommunications, insurance,² banking or transport.³ Furthermore, in the near future, the development of connected devices

¹ See e.g. European Commission, „Google/DoubleClick“, COMP/M. 4731, dated 11.03.2008, www.ec.europa.eu/competition/mergers/cases/decisions/m4731_20080311_20682_de.pdf, §§ 359-366; European Commission, „Facebook/Whatsapp“, COMP/M. 7217, dated 03.10.2014, www.ec.europa.eu/competition/mergers/cases/decisions/m7217_20141003_20310_3962132_EN.pdf, §§ 180-189.

² 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>.

³ 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.

should make data more and more relevant for product industries and not only for services.

This paper aims to feed this debate by identifying some of the key issues and parameters that may need to be considered when assessing the interplay between data, market power and competition law. For this purpose it is necessary to first clarify what can be meant by “data” or the often cited “big data”, whether there are different types of data with possibly different features, in which possible ways data can be collected and how they are used by firms (**section II**). The various theories of harm usually associated with data collection and exploitation in digital markets are presented in **section III**. Finally, in view of these two sections, **section IV** discusses some of the parameters that are to be considered in assessing the relevance and credibility of these theories of harm.

II. Data

1. What is meant by “data”?

There is no single definition for the term “data”. In a narrower sense the term is often used for the results of scientific experiments or measurements. But 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.⁴

The buzzword of choice in the current debate concerning antitrust and the digital economy, however, is often not simply “data”, but “**big data**” – another concept lacking a common definition.⁵ Aspects of “big data” that are often mentioned are large amounts of different types of data, produced at high speed from multiple sources, whose handling and analysis require new and more powerful processors and algorithms.⁶ 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” to be extracted).⁷

⁴ <http://www.merriam-webster.com/dictionary/data>;
<http://www.oxforddictionaries.com/definition/learner/data> .

⁵ 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/> .

⁶ European Data Protection Supervisor, https://secure.edps.europa.eu/EDPSWEB/edps/Consultation/big_data .

⁷ See Hu, Han et al., Toward Scalable Systems for Big Data Analytics: A Technology Tutorial, IEEE Access, Vol. 2 (2014),

Even though the term “big data” is used in different ways and is based on somewhat diffuse and cloudy features, it tries to describe the recent developments which have taken place in digital technologies and markets within the past two decades.

2. Types of data

There are different approaches to categorizing data.

a) Type of information

First, data can be categorized according to the different **types of information** they provide, which can be of different grades of usefulness and availability for a given undertaking. For instance, data can give information about individuals, economic entities or objects: they can e.g. provide information about an individual’s behavior, preferences, geographic location etc., the turnover achieved by a company with certain business transactions or refer to the current position and speed of a car.

However, so far, the current discussion concerning data and competition law focuses mostly on so-called personal data.⁸ Such personal data are subject to special data protection rules, which limit the gathering, processing and usage of such data in order to ensure consumer privacy. Personal data as defined in Article 2 (a) of Directive 95/46/EC (Data Protection Directive) means “any information relating to an identified or identifiable natural person”. Personal data may only be collected and processed in limited ways. Article 6 (1) (b) of the Data Protection Directive provides that personal data must be “collected for specific, explicit and legitimate purposes and not further processed in a way incompatible with those purposes”. Personal data processing requires a legal basis. One such basis is the data subject’s “unambiguous consent” (Art. 7 (a) Data Protection Directive), understood as the freely given specific and informed indication of his wishes by which the data subject signifies his agreement to personal data relating to him being processed (Art. 2 (h) Data Protection Directive).

<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.

⁸ In the view of European data protection agencies, the scope of personal data extends to data which can be 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 20th June 2007, see link above, pp.16-17.

b) **Structured vs. unstructured data**

Second, data may be differentiated by whether they are “**structured**” or not. This has consequences for the possibilities to extract economic value from the data and for the methods necessary to do so. Structured data follows a model that defines a number of fields, what type of data these fields contain and how they relate to each other. An example of structured data would be a consumer address data base that contains information related to each consumer’s name, surname, address, age, phone numbers, etc. Structured data can be more easily processed and used for commercial aims than unstructured data, at least if traditional methods are considered. Unstructured data does not conform to a specific model and usually needs to be processed by different and more recent algorithms in order to become of commercial value.⁹ Data can also be referred to as being semi-structured when its features borrow from the two former categories: it does not conform to a predefined model but certain elements or fields therein can be identified through a marker-type system.

c) **Ways to gather data**

Third, data can be characterized by the way it is gathered. Data can be gathered in different ways, which can in turn have an impact on its economic value, for instance in relation to of the issue of scarcity of data.

Data is often actively provided by persons on a **voluntary** basis. E. g. an online shop will typically ask the consumer to give his address, payment details and e-mail-contact in order to process the purchase and the consumer will provide these data by entering them into some type of form. Social networks as well as social communication services rely on their users inputting all kinds of (mostly personal) data. This may include personal information such as name, address, educational background as well as personal messages, photos, videos, comments on recent news, shopping preferences etc. General or specialized search engines rely on their users entering search terms and thereby revealing information about their interests; a video platform wants its users to upload new content data, a dating platform can

⁹ See Jim Harris, Bridging the Divide between Unstructured and Structured Data, <https://datascience.berkeley.edu/structured-unstructured-data/>; Stephen Pritchard, How to manage unstructured data for business benefit, ComputerWeekly, <http://www.computerweekly.com/feature/How-to-manage-unstructured-data-for-business-benefit> ; Bill Schoonmaker, Unstructured Data Can Create Chaos, Forbes, www.forbes.com/sites/emc/2013/07/11/unstructured-data-can-create-chaos/ .

match individuals only if they provide detailed information and insights into their personal lives.

Data can also be gathered by simply tapping sources (openly) available on the internet or by **observing** the user's behavior, even without his or her knowledge. Examples of the first variant are (general) search engines which rely on the systematic collection and processing of every page on the web available to them – a technique also known as “crawling”. The second variant is also widespread. A large number of firms “track” web users on their web journey from page to page – and sometimes even the user's journey within a single page¹⁰ – using different techniques. Nowadays these techniques have been further developed to work cross-device and not only on the web, but also in mobile apps, etc.¹¹ Some of the techniques used for this purpose can be avoided by users with little effort,¹² but others can be very hard to escape (if at all). Data generated by these means can also (technically) easily be combined with data given to a company on a voluntary basis in order to create highly differentiated user profiles. However, data protection laws may limit the scope for this type of behavior.

Finally, data can be generated by **inferring** new information using already existing data. E. g. an online fashion shop could analyze the individual products a visitor has been viewing to infer – albeit with some margin of error – whether the visitor is male or female. A firm which has different web services may combine the user data of these services in order to get new information about the user's behavior.

3. Technological progress

Recent developments in digital markets have made it possible to process an increasing amount of data while reducing the time needed for such processing. A number of complex, elaborate and time-consuming data processing tasks have only become possible once technical development had reached a certain point.

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

¹¹ See Dan Goodin, Beware of ads that use inaudible sound to link your phone, TV, tablet, and PC, ars technica, <http://arstechnica.com/tech-policy/2015/11/beware-of-ads-that-use-inaudible-sound-to-link-your-phone-tv-tablet-and-pc/> .

¹² Even though a high number of users seem not to know this or not to care about it.

One of the main reasons for this development has been the constant increase in computing power during the last decades. According to a rule of thumb that has become known as “Moore’s law”, the number of transistors that are packed into an integrated circuit doubled approximately every two years.¹³ The increasing number of transistors on a chip as well as other advancements paralleled by decreasing costs per transistor have led to faster processing units and bigger main memories. Developments of increased speed and capacities can also be observed with regard to those technologies used to store data permanently (such as hard disks and flash memory).¹⁴

Another factor has been the massive increase in speed and capacity the world’s network connections – necessary to transport data from one place to another – have seen. One needs to compare only what was available to the ordinary household back in the late 1980ies¹⁵ and what it can get now¹⁶ to illustrate this point. Moreover, this holds also true for networks used by corporate clients and for the whole telecommunications infrastructure.

A third factor that has to be taken into account is the development of new methods capable of extracting valuable information from extremely large accumulations of (often unstructured) data in recent years. A service like the Google search engine would have been impossible without this.

4. Role of data in economic activities

It is not a new phenomenon that businesses rely on data. Even back in the “old economy” customer data were an essential source of information for any undertaking, e. g. in order to deliver advertising to possible customers, predict their shopping

¹³ Currently there is a discussion about to what extent this development will end in the years to come, see e. g. Peter Bright, Moore’s law really is dead this time, *ars technica*, <http://arstechnica.com/information-technology/2016/02/moores-law-really-is-dead-this-time/> ; M. Mitchell Waldrop, The chips are down for Moore’s law, *Nature*, <http://www.nature.com/news/the-chips-are-down-for-moore-s-law-1.19338> ; After Moore’s law – The future of computing, *The Economist*, 12.03.2016, <http://www.economist.com/news/leaders/21694528-era-predictable-improvement-computer-hardware-ending-what-comes-next-future> .

¹⁴ See e. g. Timeline: 50 Years of Hard Drives, *PCWorld*, <http://www.pcworld.com/article/127105/article.html>; and the list of resources at <http://www.storagenewsletter.com/rubriques/miscellaneous/history-of-storage-industry/> .

¹⁵ The typical service a consumer could get in the late 80ies and early 90ies of the 20th century was a modem delivering a few thousand bits per second, see Tamsin Oxford, Getting connected: a history of modems, *Techradar*, <http://www.techradar.com/news/internet/getting-connected-a-history-of-modems-657479/1> ; Benj Edwards, 60 Years of Hooking Up, *PCWorld*, http://www.pcworld.com/article/218274/modems_through_the_years.html#slide1 .

¹⁶ E. g. most households in Germany can get at least some million bits per second, in some areas also one or two hundred million bits per second.

preferences or spending capacity and analyze their employees' performance. Marketing is based on market research, which comprises systematic data collection, processing and analysis. Every undertaking is interested in learning as much as possible about its customers' interests in order to improve its products, offer personalized services as well as to target advertisements.

However, the technological changes of the digital economy have revolutionized the possibilities to collect, process and commercially use data in almost every business sector. Therefore nowadays much more attention is paid to the economic advantages of using data to foster the position of an undertaking. While the collection and use of greater volumes of data may raise competition concerns in some cases (see Sections III and IV), it is also a mechanism by which products and services are improved and economic efficiency is raised.

a) Improvements to products or services

Data can help to improve an undertaking's product or service. On the one hand this can be achieved by learning effects as in the case of web search engines. It can be safely assumed that more searches together with the possibility to observe on what results each user clicks can help improve and refine the search engine as well as the implementation of its supporting algorithm. This can improve the search result's quality, which in turn can lead to more people using the search engine. Similarly, many software products installed on personal computers or smartphones collect detailed information concerning the usage of such products. Prominent examples are web browsers¹⁷ and operating systems.¹⁸ Also, many, if not most, websites gather detailed information on the user's journey through their site and use this information in order to identify those parts which have been used intensively or to minimize technical problems. This information can be used e. g. to extend those parts of the website that were read most often or to accelerate a software product's most used functions in order to improve the product.

¹⁷ E.g., Firefox, if the user has opted in: <https://support.mozilla.org/en-US/kb/share-telemetry-data-mozilla-help-improve-firefox>.

¹⁸ E.g., Microsoft's "Windows 10" where opting out completely is not possible for private users. See Woody Leonhard, Microsoft walks a thin line between Windows 10 telemetry and snooping, InfoWorld, <http://www.infoworld.com/article/3020152/microsoft-windows/microsoft-walks-a-thin-line-between-windows-10-telemetry-and-snooping.html>; Nick Heath, Windows 10 now lets you turn off tracking - but only if you're a business, TechRepublic, <http://www.techrepublic.com/article/windows-10-now-lets-you-turn-off-tracking-but-only-if-youre-a-business/>.

Beyond the group of examples described above it is evident that more data can be especially important for those services for which data *is* more or less the product itself or is at least closely related to the product. The latter is the case for “matching platforms” such as online dating platforms: The more individuals provide the dating service with their personal profile the more attractive the service becomes for new individuals who hope to find their ideal partner. YouTube as a video platform is probably more attractive the more videos it can offer. However, in both cases the improvement to the service hinges in part on the platform’s capacity to allow for a smooth navigation through this trove of data and preserve the user experience while increasing the breadth and depth of its offering.

b) Exploitation of new business opportunities

Access to data can also enable firms to exploit new business opportunities. By reusing data gathered in the context of one service for a different purpose undertakings may provide new services based on these data. E. g. mobility data generated by mobile network operators and mobile phones are used by navigation service providers to better show traffic jams and route their users around them.¹⁹

c) More target-oriented business models

Data can also be used to better target potential customers and to provide them with individualized advertising, services or products. Companies may thus reduce their advertising costs by getting closer to addressing only their actual target audience. Companies may even set individual prices based on estimated consumer spending capacities and price sensitivities. Even though individual pricing has – up to now – not been observed to be widespread in practice,²⁰ it may become an issue in the future.

Online advertising based on so-called “behavioral targeting” can serve as an example of a business model that has been made possible in its current ubiquity only by the technical developments described above. What is meant by “behavioral targeting” is the serving of online ads to specific users based on (comprehensive) profiles of the

¹⁹ See: Gianna-Carina Grün, Der beste Staumelder ist das eigene Handy, Die Zeit, <http://www.zeit.de/digital/mobil/2012-06/staudaten-handy> – only available in German.

²⁰ See the joint French report by the CNIL and the DGCCRF, <http://www.cnil.fr/institution/actualite/article/article/ip-tracking-conclusions-de-lenquete-conjointe-menee-par-la-cnil-et-la-dgccrf/> as well as the report by the OFT, Personalised Pricing: Increasing Transparency to Improve Trust, http://webarchive.nationalarchives.gov.uk/20140402142426/http://www.of.gov.uk/shared_of/markets-work/personalised-pricing/of1489.pdf .

users generated by observing their surfing habits.²¹ This business model would not exist without the developments mentioned for three reasons. First, without the comprehensive penetration of most if not all areas of life with digital information and communication technologies it would not be possible to record the behavior of a very large number of people on a detailed basis. Only since large parts of the population have begun to read their news online, view their films and shows online, shop online, listen to streaming music services and internet radio stations, comment public policy or some celebrity's latest dress online, present themselves in social networks and always carry a device with them that can record their geographic position, has it been possible to record their actions in such a precise way that detailed and individualized conclusions on their receptiveness to specific sales messages can be drawn. Second, to implement this at very large scale, very substantial computing power and data storage capacities are needed. And third, the ubiquitous and fast digital communication channels of today are needed to actually distribute all these sales messages on an individual basis.

III. Possible role of data in the competitive analysis

As the amount of information about users and their preferences collected by companies is rapidly growing, the impact of data as a factor in the competition analysis attracts more and more attention. The following issues can be identified in the case law, existing publications and discussions: First, the collection and exploitation of data may raise barriers to entry and be a source of market power (1). It may also reinforce market transparency, which may impact the functioning of the market (2). There are finally several types of data-related conducts of an undertaking that might raise competition concerns (3).

1. Data as a source of market power

Provided that access to a large volume or variety of data is important in ensuring competitiveness on the market (which is a market-specific question), the collection of data may result in entry barriers when new entrants are unable either to collect the data or to buy access to the same kind of data, in terms of volume and/or variety, as established companies.

²¹ See smm Sales & Marketing Management: Behavioral Targeting, <https://salesandmarketing.com/article/behavioral-targeting/>; What Is Behavioral Targeting?, CBS News, <http://www.cbsnews.com/news/what-is-behavioral-targeting/>.

As mentioned in section II, there are different ways to gather data. Data may be voluntarily given by customers or prospective customers of a company. It may also be observed by this company during an interaction with a (prospective) customer or inferred by the company from such interaction, as when an online retailer observes the articles that a given customer has clicked on without purchasing them. In all those cases, the company itself has control about the collection of data because it is involved in the relationship with the (prospective) customer (“first party data”). Endowed with a smaller number of users, smaller companies or new entrants will typically collect less first-party data than larger, more established counterparts.

However, in addition to first-party data, companies may also use “third-party data”, i.e. data collected by another entity, if such data is available to it. There may be limits to the availability of third party data, for example where such other entity is not willing to share with or sell these data to its competitors. As regards cost, the fixed costs associated with third-party data can be lower than for first-party data but the associated variable costs can be higher. Above all, resorting to third-party data also allows, at least in principle, the creation of larger datasets which are not limited to the company’s own (prospective) customers.

But even though every company could in theory buy “third-party data” in order to match the incumbent’s data trove, this might not be possible in practice due to the quantity and quality of the established company’s data set. In some sectors, the leading companies may have such a large base of customers and of information that the question arises of whether any third party is able to match the same volume and variety of data. This may particularly (but not exclusively) be the case of online services such as search engines or social networks where “free” attractive services are offered to a wide base of users, which, in turn, generate a large volume of data which may not be accessible to competitors.

As an example, the 2014 DoJ’s action against the merger of *Bazaarvoice* and its leading rival *Power-Reviews* established that data can serve as an entry barrier in the market for “rating and review platforms”. The horizontal merger between these two undertakings would have lessened competition for “rating and review platforms” in the U.S by creating a near monopoly on that market, compounded by the significant entry barriers in this market, including network effects from syndication, switching

costs, moderation, analytics and reputation.²² Such a possible increase in entry barriers could isolate established competitors from smaller rivals and potential entrants, thereby allowing them to increase their prices. In itself, the prospect of higher entry barriers, which still has to be evaluated on a case-by-case basis as shown in section IV, has important consequences for competition only when the level of market concentration is relatively high or if the market characteristics are favorable to tacit collusion. In that case, while both consumers and economic efficiency would benefit from the productivity gains associated with the development of data collection and usage possibilities, they could be also harmed by the increase in entry barriers and the reduced competition associated with these developments.

As it turns out, the economic sectors where the collection and use of data is often seen as particularly important, such as search engines or social networking for instance, are often particularly concentrated, with a few operators already holding very high user shares. Furthermore, the alleged existence of strong scale and network effects in these cases is also described as limiting the intensity of competition (see section IV). 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, the 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 converging towards a monopolization of data-related markets.²³

²² DOJ, Antitrust Division, Competitive Impact Statement, 13-cv-00133 WHO, dated 08.05.2014, <http://www.justice.gov/atr/case-document/file/488826/download>, p. 5.

²³ In some cases, the development of data can also reduce entry barriers, for instance when those data gained on a given market can be used to identify and satisfy the needs of consumers on another market.

2. Data, market transparency and competition

The increasing collection and use of digital data is often associated with greater online market transparency. From an economic point of view, such transparency has ambiguous effects on the functioning of markets.

On the one hand, consumers may benefit from greater market transparency if it allows them to compare more easily prices or characteristics of competing goods or services. For instance, price comparators or platforms like Tripadvisor allow consumers to make more informed choices resulting in a higher intensity of competition both in terms of price and quality. Market places are another illustration of the benefits that market transparency may bring. Amazon Marketplace or E-Bay host many online shops including smaller ones that might have been prevented from entering the market without such a platform. Moreover, they allow the comparison of prices and conditions offered by their hosted merchants, thereby contributing to market transparency. In some cases, greater transparency can also facilitate entry by new competitors who have more information about consumer needs and market conditions.

On the other hand, the greater information resulting from expanded data collection, especially about competitors' pricing, may also be used by undertakings in ways that could limit competition. Market transparency can indeed enhance the stability of a (tacit or explicit) collusion. By making the detection of a deviation from an agreement easier, market transparency lowers the expected profit and therefore the incentive of a party to deviate from a tacit or explicit collusion. To a certain extent, the increasing availability of data on prices on the Internet, and the fact that these data are displayed in real time, could give online markets an unprecedented level of transparency.²⁴

Data collection may also facilitate collusion when these data are used to fix prices through the use of algorithms. Even though market transparency as a facilitating factor for collusion has been debated for several decades now, it gains new relevance due to technical developments such as sophisticated computer algorithms. For example, by processing all available information and thus monitoring and analyzing or anticipating their competitors' responses to current and future prices, competitors may easier be able to find a sustainable supra-competitive price

²⁴ Although e. g. the use of personalized pricing systems could counteract this effect to some extent.

equilibrium which they can agree on.²⁵ Furthermore, data-crunching algorithms can also be used to implement an agreement, detect deviations and more generally let the collusive prices react in a more precise manner to changes in exogenous market conditions.

Finally, some analysts²⁶ also add to these risks two more sophisticated concerns. First, even in the absence of explicit horizontal coordination, the use of similar pricing algorithms, for instance if these algorithms are provided by the same company, could attenuate competition by reducing uncertainty and behavioural biases favourable to price competition. Second, data-based algorithms could also limit competition by integrating in their price-fixing mechanisms the competitors' reactions, drawn from data collected in past experiences of price variations. For instance, even algorithms designed by different companies could be unilaterally targeted to follow competitors' price increases, punish deviations, etc. Alternatively, tacit collusion could also be the result of sophisticated machine-learning. All in all, prosecuting such conducts could prove difficult: first, market transparency is generally said to benefit consumers when they have – at least in theory – the same information as the companies and second, no coordination may be necessary to achieve such supra competitive results.

3. Data-related anticompetitive conducts

When data confer significant competitive advantages to their owners, firms will need to acquire more data and/or analyse and exploit it better in order to remain competitive and/or to get a competitive edge over market rivals. The conducts

²⁵ Real examples of those scenarios include *United States v. Airline Tariff Publ'g Co.*, 836 F. Supp. 9, 12 (D.D.C. 1993) as well as the *Topkins* case. The ruling in that latter case (http://www.justice.gov/sites/default/files/opa/press-releases/attachments/2015/04/06/topkins_information.pdf and <http://www.justice.gov/opa/pr/former-e-commerce-executive-charged-price-fixing-antitrust-divisions-first-online-marketplace>) stated that “in order to implement this agreement, TOPKINS and his co-conspirators agreed to adopt specific pricing algorithms for the agreed-upon posters with the goal of coordinating changes to their respective prices. In furtherance of the conspiracy, TOPKINS wrote computer code that instructed Company A’s algorithm-based software to set prices of the agreed-upon posters in conformity with this agreement. For the purpose of reaching agreements on prices, enforcing adherence to the agreements reached, and monitoring the effectiveness of the pricing algorithms, TOPKINS and his co-conspirators, exchanged, monitored, and discussed information on the prices and sales of the agreed-upon posters. In accordance with the agreements reached, TOPKINS and his co-conspirators sold, distributed, and accepted payment for the agreed upon posters at collusive, non competitive prices on Amazon marketplace.”²⁶ See Maurice E. Stucke and Ariel Ezrachi, *Artificial intelligence and collusion: when computers inhibit competition*, University of Tennessee Legal Studies Research Paper No. 267 (2015), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2591874.

²⁶ See Maurice E. Stucke and Ariel Ezrachi, *Artificial intelligence and collusion: when computers inhibit competition*, University of Tennessee Legal Studies Research Paper No. 267 (2015), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2591874.

implemented to acquire these data can be examined from a competition law perspective. In parallel, undertakings can use their data-based market power to gain a competitive edge on adjacent markets. Concerns are also often voiced as regards data-related price discrimination. Finally, the possible interplay between competition law and privacy rules is touched upon.

a) Mergers and acquisitions

To get a better access to data, a first corporate strategy is to acquire other companies owning large datasets or to merge with them. The OECD (2015)²⁷ reports that in sectors related to data, *“the number of mergers and acquisitions (M&A) has increased rapidly from 55 deals in 2008 to almost 164 deals in 2012”* (p. 94).

In many markets, a merger between an established undertaking and an innovative newcomer has only a low impact on the existing market structure because of the newcomer’s low market shares or even the absence of horizontal overlap. However, in data-related markets, such a merger could result in differentiated data access and increase the concentration of data related to this market if the newcomer has access to a large database (gained on another market for instance).²⁸

Also, in assessing possible restraints of competition resulting from a merger, competition authorities may have to cast a close look at the advantages the new entity will have by combining different sets of data. In particular, a combination of different data troves could raise competition concerns if the combination of data makes it impossible for competitors to replicate the information extracted from it.

Furthermore, a merger of two companies which already hold strong market positions in separate upstream or downstream markets can foreclose these markets for new competitors. For instance, online service providers consuming big personal data volumes may want to acquire producers of computers, smartphones or softwares in order to make sure to continue to access important amounts of data through users of these services.

²⁷ OECD, Data-Driven Innovation: Big Data for Growth and Well-Being, OECD Publishing, Paris (2015), <http://dx.doi.org/10.1787/9789264229358-en>.

²⁸ 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, 478; 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. 123..

Such considerations are already taken into account by competition authorities and reflected in their decision making practice. For instance, in the context of the Facebook/WhatsApp merger, the European Commission assessed whether a potential integration between Facebook's social networking platform and the consumer communications application WhatsApp would allow Facebook to have access to additional data from WhatsApp users and whether this would alter competition. Likewise, in its *Telefónica UK/Vodafone UK/Everything Everywhere* merger decision, the Commission assessed "*whether the JV Co would foreclose competing providers of data analytics or advertising services by combining personal information, location data, response data, social behavior data and browsing data and by so creating a unique database that would become an essential input for targeted mobile advertising that no competing provider of mobile data analytics services or advertising customer would be able to replicate*" (§539).

Finally, mergers or acquisitions relating to data might also yield efficiency gains which would – as far as the applicable merger control regime accepts an efficiency defense – have to be compared to the risks they entail for competition. In some merger cases for instance, the parties have used the scale of data resulting from the operation as an efficiency defense. In *Microsoft/Yahoo!*²⁹, *United States v. Bazaarvoice* and *Tomtom/Tele Atlas*³⁰, efficiency claims were made by the merging parties stating that the merger would allow a company to produce better products faster because of data.³¹

b) Exclusionary conducts

Conducts depriving some competitors from access to data could also weaken competition and even lead to exclusion of competitors in different situations.

aa) Refusal to access

Refusal to access to data can be anticompetitive if the data are an "essential facility" to the activity of the undertaking asking for access. However, the ECJ has

²⁹ European Commission, „Microsoft/Yahoo!“, Comp/M. 5727, dated 18.02.2010, http://ec.europa.eu/competition/mergers/cases/decisions/M5727_20100218_20310_261202_EN.pdf, § 184.

³⁰ European Commission, „Tomtom/Teleatlas“, Comp/M. 4854, dated 14.05.2008, http://ec.europa.eu/competition/mergers/cases/decisions/M5727_20100218_20310_261202_EN.pdf, §§ 238-250.

³¹ Allen P. Grunes and Maurice E. Stucke, No Mistake about it: The Important Role of Antitrust in the Era of Big Data, University of Tennessee Legal Studies Research Paper No. 269 (2015), <http://ssrn.com/abstract=2600051>, p. 11.

circumscribed compulsory access to essential facilities to only a limited number of cases as even a dominant company cannot, in principle, be obliged to promote its competitor's business. More precisely, according to the ECJ's rulings in "Bronner", "IMS Health"³² and "Microsoft",³³ an undertaking can request access to a facility or network if the incumbent's refusal to grant access concerns a product which is indispensable for carrying on the business in question, if the refusal prevents the emergence of a new product for which there is a potential consumer demand (this condition being applicable when the exercise of an intellectual property right is at stake), if it is not justified by objective considerations and if it is likely to exclude all competition in the secondary market.³⁴ Furthermore, in *Bronner*, the ECJ ruled that a product or service is indispensable only if there are no alternative products or services and there are technical, legal or economic obstacles that make it impossible or unreasonably difficult for any undertaking seeking to operate on the downstream market to develop, possibly in cooperation with other companies, products or services.³⁵

These ECJ requirements would only be met, if it is demonstrated that the data owned by the incumbent is truly unique and that there is no possibility for the competitor to obtain the data that it needs to perform its services.³⁶ Improved data access may also lessen incentives for rivals to develop their own sources of data. Finally, access to a company's data may raise privacy concerns as forced sharing of user data could violate privacy laws if companies exchange data without asking for consumer's consent before sharing their personal information with third companies with whom the consumer has no relationship.³⁷

bb) Discriminatory access to data

Refusal to access data could also be deemed anticompetitive if it is discriminatory. As an illustration, consider the case of Cegedim.³⁸ Cegedim, the leading provider of

³² ECJ, „IMS Health“, C-418/01, judgment of 29.04.2004, §§ 34-52.

³³ GC, „Microsoft“, T-201/04, judgment of 17.09.2007, §§ 320-336.

³⁴ ECJ, „IMS Health“, C-418/01, judgment of 29.04.2004, § 37.

³⁵ ECJ, „Bronner“, C-7/97, judgment of 26.11.1998, §§ 44-45.

³⁶ 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.

³⁷ Darren S. Tucker and Hill B. Wellford, Big Mistakes Regarding Big Data, Antitrust Source, American Bar Association, December 2014, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2549044, p. 11.

³⁸ 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

medical information databases in France, refused to sell its main database (called *OneKey*) to customers using the software of Euris, a competitor of Cegedim on the adjacent market for customer relationship management (CRM) software in the health sector, but would sell it to other customers. The French Competition Authority considered such behaviour as discriminatory and concluded that, given that *OneKey* was the leading dataset on the market for medical information databases and that Cegedim was a dominant player on the market for medical information databases, such a discriminatory practice had the effect of limiting Euris's development between 2008 and 2012.

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.

cc) Exclusive contracts

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.³⁹ Exclusive agreements can exclude rivals, especially when they are concluded by dominant firms. A network of exclusive agreements might be even more problematic, not only under Art. 102 TFEU but also under Article 101 TFEU.⁴⁰ For instance, in its Art. 102 TFEU proceeding against Google, the European Commission looked into a series of exclusive contracts

médicales. This decision has been confirmed on appeal but is still pending in front of the Cour de Cassation (the French Supreme Court).

³⁹ Allen P. Grunes and Maurice E. Stucke, No Mistake about it: The Important Role of Antitrust in the Era of Big Data, University of Tennessee Legal Studies Research Paper No. 269 (2015), <http://ssrn.com/abstract=2600051>, p. 3.

⁴⁰ ECJ, "Delimitis", C-234/89, judgment of 28.02.1991 (cumulative effects of a network of similar agreements).

concluded by Google in the search advertising market that might foreclose competitors from being able to challenge the company.⁴¹

dd) Tied sales and cross-usage of datasets

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. For instance, in its report,⁴² the UK *Competition and Markets Authority* mentioned the possibility of tied sales whereby a company owning a valuable dataset ties access to it to the use of its own data analytics services. As it noted, such tied sales may increase efficiency in some circumstances but they could also reduce competition by giving a favorable position to that company which owned the dataset over its competitors on the market for data analytics.

More generally, in an Opinion of 2010, the French Competition Authority emphasized that cross usage of data, i.e., the use of data collected on a given market onto another market, can, in certain circumstances, have foreclosing effects.⁴³ In particular, former monopolies having a privileged access to data in the context of their public service activity can use these data to propose tailored offers to consumers on adjacent markets, therefore possibly gaining a strong competitive advantage that could not be matched by competing suppliers. This reasoning recently led the French Competition Authority to impose interim measures to GDF-Suez, ordering that gas supplier to grant its competitors an access to some of the data it collected as a provider of regulated offers, in particular consumption data.⁴⁴ The aim of this interim measure was to allow all suppliers to have the same level of relevant information to make offers to consumers (no public information or private database exists on households subscribing to gas contracts).

⁴¹ European Commission, "Google", case 38740, press release dated 30.11.2010, http://europa.eu/rapid/press-release_IP-10-1624_en.htm?locale=en.

⁴² 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.

⁴³ French Competition Authority, Opinion 10-A-13 on the cross-usage of customer databases.

⁴⁴ French Competition Authority, Decision 14-MC-02 of 09.09.2014. Due to privacy laws, the transmission of GDF data to competitors was conditional to an approval by consumers. A significant share of the consumers did refuse that their data be transferred from GDF to competing operators.

c) Data as a vehicle for price discrimination

Data is also said to facilitate price discrimination.⁴⁵ Indeed, by collecting data about their clients, a company receives better information about their purchasing habits and is in a better position to assess their willingness to pay for a given good or service. Provided that it has market power,⁴⁶ the company would then be able to use that information to set different prices for the different customer groups it has identified thanks to the data collected.

However, different effects of price discrimination can be considered. On the negative side, price discrimination is often viewed as an unfair breach of consumer equality. Also, while consumers can choose to shop elsewhere to escape unfavourable price discrimination, greater price discrimination may increase search costs, thus reducing the ease of substitution. Finally, due to price discrimination, some consumers end up paying higher prices for their purchases than before the discrimination was implemented.

Yet, the economic analysis also shows that the effects of price discrimination, assuming that it is made easier thanks to data, are more ambiguous. Indeed, as a consequence of data-based price discrimination, some consumers would end up paying higher prices for a given good or service but some others would receive better price offers than in the absence of discrimination. These better-off customers include some consumers that could not afford buying the product if only a single, identical-for-all, price was allowed. They are also likely to be more price-sensitive and to have lower revenues than others.⁴⁷ In that respect, price discrimination may improve the social welfare, i.e., the sum of the undertakings' profits and the consumers' surplus, by increasing the number of transactions in comparison with the situation without price discrimination. However, the maximization of the social welfare does not necessarily translate into a maximization of the consumers' surplus. Indeed, price discrimination can help suppliers to set prices according to the consumers'

⁴⁵ 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>.

⁴⁶ Without any market power, a supplier is not able to set its prices in reference to the willingness to pay of consumers. Indeed, a competitor that would set a single price in reference to its costs would be more attractive than that supplier vis-à-vis most consumers.

⁴⁷ Executive Office of The President of The United States, Big Data And Differential Pricing (2015), https://www.whitehouse.gov/sites/default/files/docs/Big_Data_Report_Nonembargo_v2.pdf ("if historically disadvantaged groups are more price-sensitive than the average consumer, profit-maximizing differential pricing should work to their benefit").

willingness to pay, so that suppliers could appropriate most of the consumers' surplus.

Besides this, data-based price discrimination can also reinforce competition. For instance, thanks to price discrimination, an undertaking can propose lower prices to those customers with a strong preference for another product, resulting in greater price competition.⁴⁸ Therefore, even assuming that data allows for greater price discrimination, its effects are difficult to assess.

Finally, one may query whether price discrimination in itself is within the scope of European competition law. To be considered as either an abuse of 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.⁴⁹

d) Data, market power and privacy concerns

Another concern that has been raised in the literature exploring the competition law issues related to the collection and use of (personal) data relates to reduction in privacy protection.

Privacy concerns are not, in and of themselves, within the scope of intervention of competition authorities. The European Court of Justice has stated in *Asnef-Equifax* (2006) that any issues relating to the sensitivity of personal data are not, as such, a matter for competition law, but may be resolved on the basis of the relevant provisions governing data protection.⁵⁰ The European Commission confirmed this view in "Facebook/Whatsapp" by stating: "*Any privacy related concerns flowing from the increased concentration of data within the control of Facebook as a result of the Transaction do not fall within the scope of the EU competition law rules but within the*

⁴⁸ See Mark Armstrong, Price discrimination, University College, London (2006), <http://else.econ.ucl.ac.uk/papers/uploaded/222.pdf>.

⁴⁹ 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>.

⁵⁰ ECJ, „Asnef-Equifax“, C-238/05, judgment of 23.11.2006, <http://curia.europa.eu/juris/showPdf.jsf?jsessionid=9ea7d0f130d5bea0e088b08f44b3b853d5eb7ffd88fa.e34KaxilLc3eQc40LaxqMbN4Och0Qe0?text=&docid=65421&pageIndex=0&doclang=EN&mode=lst&dir=&occ=first&part=1&cid=236269>, § 63.

scope of the EU data protection rules“.⁵¹ Its analysis in this case was focused on the market for advertisements on which advertisers could be harmed by an increase in Facebook’s market power through increased data collection capabilities, but not on whether personal data from end users may be collected to a greater extent due to the operation. As it explained, *“for the purposes of this decision, the Commission has analysed potential data concentration only to the extent that it is likely to strengthen Facebook’s position in the online advertising market or in any sub-segments thereof”* (§164).

However, the fact that some specific legal instruments serve to resolve sensitive issues on personal data does not entail that competition law is irrelevant to personal data. Generally speaking, statutory requirements stemming from other bodies of law may be taken into account, if only as an element of context, when conducting a legal assessment under competition law. In *Allianz Hungária* (2013) for instance, the ECJ held that the impairment of objectives pursued by another set of national rules⁵² could be taken into account to assess whether there was a restriction of competition (in this instance, by object).⁵³ Referring to 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.⁵⁴

Indeed, even if data protection and competition laws serve different goals, privacy issues cannot be excluded from consideration under competition law simply by virtue of their nature. Decisions taken by an undertaking regarding the collection and use of personal data can have, in parallel, implications on economic and competition dimensions. Therefore, privacy policies could be considered from a competition standpoint whenever these policies are liable to affect competition, notably when they are implemented by a dominant undertaking for which data serves as a main input of

⁵¹ European Commission, „Facebook/Whatsapp“, COMP/M.7217, dated 03.10.2014, http://ec.europa.eu/competition/mergers/cases/decisions/m7217_20141003_20310_3962132_EN.pdf, § 164.

⁵² In the cited case, Hungarian domestic law required that insurance dealers acting as intermediaries or insurance brokers must be independent from the insurance companies.

⁵³ ECJ, „Allianz Hungária“, C-32/11, judgment of 14.03.2013, <http://curia.europa.eu/juris/document/document.jsf?text=&docid=135021&pageIndex=0&doclang=EN&mode=lst&dir=&occ=first&part=1>, §§ 46, 47.

⁵⁴ German Federal Court of Justice (Bundesgerichtshof), „VBL-Gegenwert“, KZR 61/11, judgment of 16.11.2013, available via <http://juris.bundesgerichtshof.de/cgi-bin/rechtsprechung/list.py?Gericht=bgh&Art=en&Datum=Aktuell&Sort=12288>, § 68.

its products or services. In those cases, there may be a close link between the dominance of the company, its data collection processes and competition on the relevant markets, which could justify the consideration of privacy policies and regulations in competition proceedings.

A look at cultural diversity issues might be interesting, too. Cultural diversity has explicitly been taken into account by the European Commission, notably in the field of merger control (see Universal/EMI decision). Article 167(4) of the Treaty on the functioning of the European Union provides that “*The Union shall take cultural aspects into account in its action under other provisions of the Treaties, in particular in order to respect and to promote the diversity of its cultures*”. Article 16 TFEU, while not explicitly mandating that data protection aspects be considered in all actions conducted by the Union pursuant to the Treaties, does affirm that “*Everyone has the right to the protection of [their] personal data*”. Even as it remains open to question whether these provisions carry specific obligations for competition authorities, the *European Data Protection Supervisor* has advocated, in 2014, a shift in policy and a “*more holistic approach to enforcement*”, in which a more systematic dialogue is maintained between competition, consumer and data protection authorities, “*wherever a specific case arises in which consumer welfare and data protection concerns appear to be at stake*”.⁵⁵

In merger control cases, the question of data privacy might particularly become relevant from a competition standpoint if a given undertaking benefits from a strong market power towards its end-users. Indeed, firms that gain a powerful position through a merger may be able to gain further market power through the collection of more consumer data and privacy degradation. If two horizontal competitors compete on privacy as an aspect of product quality, their merger could be expected to reduce quality.⁵⁶ For some analysts,⁵⁷ a reduction of privacy is indeed tantamount to a reduction of product quality. To date, this has remained, however, a theoretical

⁵⁵ European Data Protection Supervisor, Privacy and competitiveness in the age of big data, preliminary opinion (2014), https://secure.edps.europa.eu/EDPSWEB/webdav/site/mySite/shared/Documents/Consultation/Opinions/2014/14-03-26_competition_law_big_data_EN.pdf.

⁵⁶ CMA, 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. 95.

⁵⁷ Allen P. Grunes and Maurice E. Stucke, No Mistake about it: The Important Role of Antitrust in the Era of Big Data, University of Tennessee Legal Studies Research Paper No. 269 (2015), <http://ssrn.com/abstract=2600051>, p. 4.

scenario since privacy has so far not emerged as a significant parameter of competition in competition authorities' practice.

Further, reductions in privacy could also be a matter of abuse control, if an incumbent collects data by clearly breaching data protection law and if there is a strong interplay between the data collection and the undertaking's market position. So far, competition authorities understood exploitative conduct mostly as an instrument against excessive pricing.⁵⁸ Such an intervention against excessive pricing faces many practical difficulties since it involves finding a comparable market or complex cost-based price comparisons and a determination of useful "benchmark" prices. Some argue that these practical difficulties and the risk of competition authorities arriving at the wrong result are so great that enforcement actions against exploitative conduct should only be taken as a last resort.⁵⁹ However, looking at excessive trading conditions, especially terms and conditions which are imposed on consumers in order to use a service or product, data privacy regulations might be a useful benchmark to assess an exploitative conduct, especially in a context where most consumers do not read the conditions and terms of services and privacy policies of the various providers of the services that they use.⁶⁰

IV. Assessing Big Data and market power

Among the aforementioned potential impacts of the collection and use of data, the issue of data possibly contributing to market power is most likely to arise and is, in many respects, the most interesting one from a competition standpoint. Several of the theories of harm presented in section III are conditional on the extent of the competitive advantage generated through data collection, and more precisely, on the advantage specifically conferred to those undertakings that are able to collect data to a greater extent. This question is complex. Especially in online industries the issue of data does not exhaust the somewhat "non-traditional" aspects to be considered when assessing market power. This must be kept in mind (1). Then, to assess whether

⁵⁸ Emil Paulis, „Art. 82 EC and Exploitative Conduct“, in Ehlermann/Marquis, European Competition Law Annual 2007: A Reformed Approach to Article 82 EC, p. 2.

⁵⁹ Emil Paulis, „Art. 82 EC and Exploitative Conduct“, in Ehlermann/Marquis, European Competition Law Annual 2007: A Reformed Approach to Article 82 EC, p. 3.

⁶⁰ Data protection Eurobarometer (June 2015) shows that only a fifth of respondents fully read privacy statements (18%) http://ec.europa.eu/justice/data-protection/files/factsheets/factsheet_data_protection_eurobarometer_240615_en.pdf; The Bundeskartellamt has recently initiated a proceeding against Facebook to investigate suspicions that with its specific terms of service on the use of user data, Facebook has abused its possibly dominant position in the market for social networks.

data indeed contribute to creating or preserving market power, competition authorities will have to evaluate the extent of the economic advantage that data provides. This is an exercise that has already been implemented in several cases although, as far as personal data on digital markets are concerned, only merger cases have considered this issue (2). These merger decisions as well as several contributions from academics or economic consultants underline the attention that should be paid to the availability of data on digital markets (3) as well as to the specific advantages that derive from a greater volume of data (4).

1. Sources of market power in data-driven online industries

Many industries, including those involving the sale of physical goods through offline channels, could be affected by present and future possibilities of collecting and using data. However, most of the debate on whether and how these possibilities are likely to affect competition has focused so far on online services, specifically on search engines, social networking and online retailing. Indeed, thanks to their size and digital channels and to devices being the base of their market interactions, online businesses often collect more data than brick and mortar counterparts currently do. Online businesses may also be more able to use these data at short notice, thanks to automated processes, for example for advertising or recommendation purposes. Furthermore, these data can economically support many online business models, especially those financed by advertising. Finally, some online markets are perceived by various observers as excessively concentrated and they fear that the development in data collection and exploitation may increase this concentration process.

Online businesses and their economic characteristics have been a field for controversial academic discussions within the last decade.⁶¹ Economists have identified some aspects which characterize several online industries and which have to be taken into account while assessing market power of undertakings. These include in particular network effects, multi-homing and market dynamics.

The academic discussion on these factors is still ongoing. Competition practice, however, has shown that phenomena like these may have ambiguous effects on

⁶¹ David S. Evans and Richard Schmalensee, *The industrial Organization of Markets with Two-sided Platforms*, Competition Policy International, 2007, Vol. 3, p. 164; Bernard Caillaud and Bruno Jullien, *Chicken & egg: competition among intermediation service providers*, RAND Journal of Economics 2003, 34(2), pp. 309-328; Mark Armstrong, *Competition in two-sided markets*, RAND Journal of Economics 2006, 37(3), pp. 668-691; Lapo Filistrucchi/Damien Geradin/Eric van Damme, [Identifying two-sided markets](#), World Competition: Law and Economics Review 2013, 36 (1), pp. 33-60.

competition and that there are no universally valid conclusions that can be drawn without looking at the specific case involved. Therefore, competition assessment needs to be supported by extremely refined and case-related considerations.

a) Multisided markets and network effects

Many online markets are so-called “multi-sided markets”, where the undertakings are active towards more than one group of users/customers. Such a characteristic can pose questions when defining relevant markets. Furthermore, some services are provided without monetary consideration on a given side of the market, which raises the question, whether this should prevent a relevant market from being defined (as has been traditional practice e. g. in Germany) despite companies on the “free side” of the market competing on other aspects than price, such as the quality of services or, probably to a lesser extent so far, the amount of data collected on each individual user.⁶²

Furthermore, so-called “network effects” are often to be found here. The term “network effects” refers to how the use of a good or service by a user impacts the value of that product to other users. Such effects may be “direct”, when the benefit that users of one group get from a specific service depends on the number of other users from this group using the service. Telecommunication networks are the classic example. The more people use them and can be reached, the more useful they are. Network effects can also be “indirect”, when the benefit that users of one group get from the service depends on the number of users from a different group using the service. A dating platform bringing together men and women can serve as an example here. Direct and indirect network effects may also coexist in some cases. For instance, the value of a social network for a given user is likely to increase with the total number of users of that network (direct network effects). Meanwhile, a higher number of users of a social network also increases the value for advertisers (indirect network effects). It should be pointed out that indirect network effects are not necessarily symmetric. For instance, the value of a social network for advertisers increases with the number of users but it is unclear and even doubtful that users directly value a higher number of advertisers or advertisements, although they may value the various investments made by the social network to improve its services thanks to these advertising revenues. Hence, the extent of network effects has to be

⁶² See GC, “Microsoft”, T-201/04, judgment of 17.09.2007, §§ 966-970; GC, “Cisco Systems”, T-79/12, judgment of 11.12.2013, §§ 65-74.

evaluated on a case-by-case basis: both their absolute importance and how they evolve as new customers are gained can vary depending on the service under consideration.

Network effects can have both negative and positive impacts on competition. Often, network effects are mentioned with a self-strengthening scenario in mind that can favour market concentration. They are also seen as a potential barrier to entry or an element of such barrier and thus as a factor which limits competition. In such context data collection and exploitation could possibly also reinforce network effects, when an increase in a firm's user share enables it to collect more data than its competitors, leading to higher quality products or services and to further increases in market shares (see section III above).

However, network effects may also be beneficial to new market participants if they are able to attract a high number of users for other reasons (e. g. because of an innovative feature), thereby increasing their attractiveness to future users thanks to network effects. Therefore, network effects can also stimulate competition by giving an entrant the potential for a rapid growth of its consumer base. Depending on various parameters such as the level of fixed costs or the differences in the undertakings' market shares, network effects could thus reinforce or attenuate competition.

b) Multi-homing

Consumers are said to multi-home when they use several providers to get the same kind of service.

Many academics see multi-homing as a factor likely to reduce market power.⁶³ But a situation of perfect multi-homing, where every user of a service practices multi-homing, will be rather rare, due to various switching costs (network effects, learning costs, etc.). In particular, data collection may increase switching costs as the provider most used by an individual has more information on him or her and is able to tailor his service offerings to that particular individual. Overall, something in between single homing and multi-homing is far more likely in day-to-day life. Further, the interplay of multi-homing (to some extent) on one side of a two-sided market and single-homing

⁶³ 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.

(to some extent) on the other side has to be considered when analyzing the effects in a specific case.

It is also frequently argued that markets where services are free for end users are particularly prone to multi-homing. There are some limits to this assertion however. Switching costs can prevent consumers from using various providers in equal proportions. Consumers may pay more attention to quality when services are free because in this context, quality is the only dimension of competition between platforms; yet, in a market characterized by network economies and experience effects (possibly related to the volumes of data collected)⁶⁴, new entrants may not be able to propose services of a quality as high as those of established undertakings and could not compensate this lower quality by proposing lower prices.

Finally, while the potential for multi-homing and, more generally, low switching costs, may decrease the market power of established undertakings (by making substitution much easier if the quality of services provided by established undertakings diminishes), this potential multi-homing is not necessarily relevant when considering data-based market power. Indeed, user-based data may only make a difference if end-users multi-home and use rival providers sufficiently frequently – a rather tall order when network and experience effects are at stake.

c) Market dynamics

An often heard argument is that established market power on digital markets can be vulnerable to displacement by innovative products. A *“tremendous amount of entry and rapid gains often enjoyed by innovative new challengers”*⁶⁵ on online markets could be an indication that entry and development barriers, be they related to data requirements or to other kinds of assets, are low. Proponents of this “dynamic competition” argument frequently refer to the displacement of Yahoo, Lycos or AltaVista by Google and of MySpace by Facebook within a few years. When they entered the market, Google and Facebook did not have access to the same network and experience effects and scale economies as their larger rivals. Likewise, they did not benefit from the same volume of data as the market leaders; still, they overcame that disadvantage. In its Facebook/Whatsapp decision, when analyzing the market

⁶⁴ Experience economies designate the cost reduction (or, possibly, quality increases) of products as their volumes of production/consumption increase.

⁶⁵ Darren S. Tucker and Hill B. Wellford, Big Mistakes Regarding Big Data, Antitrust Source, American Bar Association, December 2014, <http://ssrn.com/abstract=2549044>.

for consumer communication applications, the Commission considered that *“consumer communications apps are a fast-moving sector, where customers' switching costs and barriers to entry/expansion are low. In this market any leading market position even if assisted by network effects is unlikely to be incontestable. The market of consumer communications apps has a long track record of entry by new players. Also, competing consumer communications apps are able to grow despite network effects, both over time and following disruptions in the market. Such threat from new players constitutes and is likely to keep constituting a significant disciplining factor for the merged entity, regardless of the size of its network”* (§ 132).

While dynamic competition could be strong enough to mitigate competition law concerns related to static market power on at least some markets, this balance should be evaluated on a case-by-case basis and the difficulties for new undertakings to enter and grow on online or digital markets should not be underestimated and therefore be carefully assessed, along with the possible role of data collection in this regard. Depending on the case under consideration, the expenses necessary to enter the market can include research and development expenses, tangible assets to operate it, marketing expenses to make the service known by its potential users, etc. Also the extent of entry on online markets might deserve closer inspection. Many new entrants in the online world focus on niches, be they particular functionalities, customer segments, or user interests. Also, some promising new entrants, even well-known names of the digital world, have failed to live up to their promises, the most prominent example maybe being the “Google+” social network. Finally, some past examples of successful entrants, such as Google or Facebook on their respective markets, are not necessarily illustrative of present entry conditions because market conditions can have changed since their time of entry. In particular, it remains to be assessed to which extent the importance of data in developing new services is higher today than a few years ago. Finally, dynamic competition may be stifled if established undertakings have a proven ability to buy out competitively significant new entrants with a turnover, asset-base or market share too low to trigger merger control.

2. Evaluation of the “data advantage” in past cases

a) In “non-digital” markets

Evaluating the advantages associated with specific datasets is an exercise that competition authorities have already pursued in several cases.

In its opinion of 14th June 2010 on the cross-usage of customer datasets⁶⁶, i.e. the use by a company of datasets relating to its customers on one market as an input to develop its activity on another market, the French Competition Authority explained that the criteria used to determine whether the use of such datasets could result in a restriction of competition include the conditions under which the dataset was constituted, whether the dataset could be replicated under reasonable conditions by competitors and whether the use of the dataset was likely to result in a significant competitive advantage.

It has applied this line of reasoning in several instances, related notably to the use of customer datasets by former public monopolies. In the already mentioned decision of 2014 concerning GDF, the French Competition Authority noted that the information contained in the dataset collected by GDF referenced contact details, technical information and consumption data. It was used by the company to propose tailored offers to former and current customers and therefore helped GDF to sell its services on the gas market. The French Competition Authority considered that the dataset was not the result of any particular innovation by GDF since it was inherited from its past monopoly position. Neither was this dataset replicable under reasonable financial conditions and in a sufficiently short time. In this context, although this dataset was not considered to be an essential asset, the French Competition Authority concluded that it gave GDF an unjustified significant advantage relatively to its competitors and that GDF was likely to have abused its dominant position on the gas market by using this dataset.⁶⁷

This kind of reasoning and of evaluation of the “data advantage” can be found in several other previous cases. For instance, the French Competition Authority⁶⁸

⁶⁶ French Competition Authority, Opinion n°10-A-13 of 1406.2010, <http://www.autoritedelaconcurrence.fr/pdf/avis/10a13.pdf>. That opinion of the Authority aimed at giving general guidance on that subject. It did not focus on any particular market or industry although it described a possible application of its analysis to the telecom industry.

⁶⁷ This interim injunction was confirmed in appeal.

⁶⁸ French Competition Authority, Decision n°13-D-20 of 17.12.2013, confirmed on that points by the court of appeal on 21.05.2015.

sanctioned EDF for having abused its dominant position on the electricity supply market by using data (in particular contact details) collected from its subscribers on the electricity market to promote effectively and at a low cost the offers of one of its subsidiaries on another market (photovoltaic electricity generation). That dataset, partly inherited from its legal monopoly on the electricity supply under regulated prices, was not accessible to competitors and was considered as not replicable under reasonable financial conditions and in a short period in view in particular of its size (20 million customers).⁶⁹

A similar reasoning has also been used in some merger cases. For instance, in its EDF-Dalkia merger decision,⁷⁰ the European Commission considered how information about electricity consumption could affect competition on the market for the provision of services allowing the optimization of electricity consumption. It noted that access to such information was unlikely to give the new entity a significant competitive advantage given that the customer (businesses and co-ops which procure these services *via* a bidding process) could be given such information by the network manager (ERDF) and then provide it to any supplier competing with EDF-Dalkia. Also, in its Enerest/Electricité de Strasbourg merger decision,⁷¹ the French Competition Authority expressed the concern that the new entity which would result from the merging of the two historical suppliers of gas and electricity which were, respectively, entrusted with a legal local monopoly on the retail supply of gas and electricity at regulated tariffs in the area of Strasbourg, would be the only market player to have exclusive access to extensive electricity and gas consumption data derived from their respective legal monopolies. These data would allow it to propose offers combining gas and electricity specifically adjusted to each potential customer, thus giving the new entity a competitive advantage. In order to have the same level of information, competitors would have to send requests to their prospects, which would represent a commercial cost that the merged entity would not have to bear.⁷² The Authority cleared the operation following the commitment by Enerest and Electricité de Strasbourg to send every competitor that would request it the information necessary to design tailored combined commercial offers (§§ 79-80, §§ 87-90).

⁶⁹ The sanction of this conduct was confirmed on appeal.

⁷⁰ European Commission, “EDF/Dalkia en France”, COMP/M.7137, dated 25.06.2014, http://ec.europa.eu/competition/mergers/cases/decisions/m7137_20140625_20310_3804903_FR.pdf

⁷¹ French Competition Authority, Decision n°12-DCC-20 of 07.02.2012.

⁷² § 71.

The context of the aforementioned Cegedim decision is different in the sense that Cegedim, although it enjoyed a dominant position on the market for the provision of medical information datasets to pharmaceutical companies and medical representatives, was not a former public monopoly. Furthermore, the theory of harm was not related to the exclusive use of its data by Cegedim. Yet, as in the other aforementioned decisions, the French Competition Authority also stressed the high quality of the medical information dataset OneKey designed and sold by Cegedim, in terms of coverage and regular updates as well as in consideration of the high costs of replicating this dataset. Although the dataset OneKey was not considered to be an essential facility for competitors in the downstream market for customer relationship management (CRM) software (which used datasets of the same type as OneKey), the French Competition Authority concluded that the access conditions to that dataset which discriminated against Euris clients significantly impeded competition by degrading the quality of the services and the commercial image of Euris relative to its competitors.

b) In “digital” markets

Although the role of data in non-digital markets can be very significant in some instances and is certainly likely to increase in the future as connected devices, for instance, become ubiquitous, it is really in the case of digital markets that the collection and use of large volumes of data has attracted much attention. The next sections describe some of the criteria used to assess how data can shape competition and competition assessment in these markets. Indeed, there have already been several instances – exclusively in the context of merger control – where competition authorities have examined the advantages conferred to undertakings on digital markets by data collection. In all of these cases, the competition authorities involved, i.e., the US authorities and the European Commission, considered that the data advantage potentially enjoyed by the new entity did not lead to any risk for competition on the considered markets.

More precisely, as described in section II, the purpose of data in these digital markets often is to improve the quality of service offered to customers. Most well-known search engines use collected data about the searches made by their users to improve the quality of future search results. Online shops use their data on made and unmade purchases to recommend products to their customers. Social network providers select the most relevant pieces of information for a given user thanks to the

activity of that user on the social network. Finally, access to user data allows the platform to target the ads that it publishes to the characteristics of each of the users of the platform or to a group of them.

How these processes of data collection help their owners in providing better services to their customers is well acknowledged in these merger decisions. For instance, in its Google/DoubleClick decision, the Commission considered that *“as opposed to offline advertising, online advertising is considered to be capable of reaching a more targeted audience in a more effective way. Advertisers can precisely target their audience by combining information regarding geographical location, time of day, areas of interest, previous purchasing record of the user and search preferences”* (§ 45, emphasis added). Thus, as early as 2008, the Commission effectively acknowledged the role that data can play in helping publishers as well as advertising intermediation agencies to target their ads to user characteristics. The Commission analyzed the possibility that the new entity would try to combine DoubleClick’s and Google’s data collections, which could have contributed a better quality of service: *“Such a combination, using information about users’ IP addresses, cookie IDs and connection times to correctly match records from both databases, could result in individual users’ search histories being linked to the same users’ past surfing behaviour on the internet. For instance, after such a match, the merged entity may know that the same user has searched for terms A, B and C and visited web pages X, Y and Z in the past week. Such information could potentially be used to better target ads to users.”*⁷³ However the Commission considered that such a combination could face practical difficulties, especially in view of the potential reluctance of advertisers (§§ 361-363), and that even if it could be implemented, it would be unlikely to give the merged entity a competitive advantage that could not be matched by competitors. Indeed, several competitors both run a search engine and offer ad serving, like the merged entity. In addition the Commission held that competitors may purchase data or targeting services from third parties, which would compensate for a lower access to data compared to the merged entity (§§ 364-366). The possibility for competitors to access similar data was therefore one of the main criteria taken into account by the Commission (see section below).

⁷³ European Commission, “Google/DoubleClick”, COMP/M. 4731, dated 11.03.2008, http://ec.europa.eu/competition/mergers/cases/decisions/m4731_20080311_20682_en.pdf, § 360.

Likewise, in its Facebook/Whatsapp decision, the Commission mentioned the possibility that “the merged entity could start collecting data from Whatsapp users with a view of improving the accuracy of the targeted ads served on Facebook’s social networking platform”.⁷⁴ However it noted that the incentives of the merged entity to do so were mixed since, in particular, such a data collection could prompt some users to switch to other consumer communications apps that they perceive as less intrusive (§ 186), and that, in any event, the amount of Internet user data valuable for advertising purposes and that are not within Facebook’s exclusive control remained very large (§ 189). In this context, the Commission considered that the merger did not give rise to competition concerns on the market for online advertising services.

In its Microsoft/Yahoo! Search Business merger decision, the Commission went a bit further by considering that access to data (e.g. the searches made by its users or the links on which the users clicked) was important in determining the performance of a search engine but also that the increase in its scale of data collection would help the new entity to improve its ability to provide better services to users. Hence, the Commission indicated that “*it is plausible that the merged entity through innovation and through its access to a larger index will be able to provide personalized search results better aligned to users' preferences*” (§§ 225-226).⁷⁵ The Commission then considered that the operation did not raise any risk for competition – but in fact was likely to have a positive impact for competition – because, to put it shortly, the search engines of Yahoo and Microsoft were both lagging behind that of Google, so that a greater scale, in particular a greater scale of data collection, would be helpful in reinforcing the competitive pressure that the merged entity could exert on Google.

These merger decisions related to digital markets help to delineate two factors likely to be of much relevance when considering whether data can contribute to market power, i.e., the scarcity of data (or ease of replicability) and whether the scale/scope of data collection matters to competitive performance.

⁷⁴ European Commission, “Facebook/WhatsApp”, COMP/M.7217, dated 03.10.2014, http://ec.europa.eu/competition/mergers/cases/decisions/m7217_20141003_20310_3962132_EN.pdf, § 180.

⁷⁵ The Commission also indicated that the returns to a larger are likely to be decreasing. See below.

3. Issues pertaining to the scarcity of data

Data may impact competition if they provide a competitive advantage to their owners. If the data are widely available, no undertaking, even a dominant one, would benefit from a competitive advantage thanks to these data because competitors are able to get the same information and efficiency from data as the leading operators of the market.

In particular, there are three features that could contribute to a high availability of data and therefore reduce the risk that differentiated access to these data between competitors could harm competition. First, by definition, data are “non rivalrous” and an access to data by a given operator does not necessarily, in itself, preclude others from accessing the same data. Second, data brokers have developed, particularly in the U.S., and could increase the availability of data. Third, digital markets are particularly prone to data collection, which may further increase the availability of data. These three aspects are considered in turn.

a) On the non-rivalry of data

aa) Presentation

Data are non-rival goods in the sense that someone having and using a dataset does not prevent others, be they competitors or not, from having and using the same data as well (provided they can access them).⁷⁶ Hence, if an undertaking, be it an ad network, a sales company or any other undertaking, collects the home address, phone number, gender, date of birth, income, tastes and current interests, etc. of an individual, the use of such information does not prevent its competitors from having the same information on that individual.

Although it did not explicitly refer to non-rivalry when examining issues related to data in its Telefonica UK/Vodafone UK/Everything Everywhere joint venture decision, the Commission mentioned that “*customers generally tend to give their personal data to many market players, which gather and market it. Therefore, this type of data is generally understood to be a commodity*” (§ 543).

Non-rivalry of data may be a particularly prominent feature in the context of “multi-homing”, i.e., when individual users make use of several providers to perform the same kind of services and thus give data about themselves to several providers,

⁷⁶ 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.

none of which has any exclusivity on those data. It may furthermore remain relevant when consumers tend to single-home because a competitor may still have access to the same kind of information as another operator by proposing different services (FTC, 2014;⁷⁷ Tucker and Welford, 2014⁷⁸). For example, a new entrant onto a market for social networks may not have much data related to the interests of potential users through its social network activity. However, it may have been able to access to that information by interacting with social network users differently if he already had other commercial activities, for instance through a commercial website asking its customers for their interests and/or through their past purchases.⁷⁹

Finally, the non-rivalry of data may contribute to limit the price of data, at least from a theoretical standpoint. Indeed, potential customers of data brokers can threaten them with collecting these data themselves. Furthermore, if data are non-rival, several data brokers could simultaneously collect and sell them, thereby further reducing the price of these data and facilitating their access.

However, the fact that data are non-rival does not imply that they are accessible to all competitors. In fact, as already mentioned, there are several examples of antitrust cases where competition authorities have considered that although data were non-rival as such, access to these data turned out to be prohibitively costly, so that having an exclusive access to them was deemed a significant competitive advantage. In the following section, the practical difficulties that could prevent a competitor from access data are discussed.

⁷⁷ “Several of the data brokers share the same sources. And each data broker utilizes multiple sources for similar data. For example, one of the data brokers in this study obtains consumers’ contact information from twenty different sources.” (p. 14).

⁷⁸ “Big data is non-rivalrous. In other words, collecting a particular piece of data does not prevent other companies from collecting identical data by similar or other means. Using multiple providers for the same service (user multi-homing) and the common practice of website operators using multiple ad networks and analytic firms make it easier for multiple providers to collect relevant user data. For example, if one ad network determined that the user of a particular mobile device lived in Connecticut, liked to travel, and owned a dog, there is nothing to prevent another ad network from learning the same information—indeed, for a frequent Internet user, it is likely that dozens of firms will create a similar profile. Redundant data are so common as to cause problems for data brokers.” (pp. 3-4).

⁷⁹ Some observers have also argued that the information on past searches could be retrieved not only by search engines but also by browsers and operating systems (see Andres V. Lerner, *The Role of ‘Big Data’ in Online Platform Competition* (2014), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2482780 , p.24, citing Matt Rosoff, *Yes, Bing Has Been Copying Google Search Results FOR YEARS*, Business Insider, February 1, 2011, http://articles.businessinsider.com/2011-02-01/tech/29975847_1_bing-director-stefan-weitz-satyanadella-msn-toolbar).

bb) Factors that may limit the possibility to access data by oneself

The possibility to collect data directly may be subjected to different types of costs.

First, in order to collect data, firms may have to engage in significant investments. The emergence and development of vast data centres illustrates the potentially high level of fixed costs that have to be invested in order to collect and exploit massive amounts of data. The level of those costs associated with the accumulation of large datasets may prevent small companies and new entrants to make use of the same volume and/or variety of data as large incumbents.

Second, data are often collected from users as they make use of a product or service. This is particularly true of multi-sided platforms (such as search engines or social networks) who provide (often free) services to a first category of users and extract value from the data collected from those users by providing other products or services to a second type of customers (ad space sold to advertisers for instance). Therefore, to directly access these kinds of data, an entrant needs to build a platform that is able to provide the same kind of services (or other kinds of services that allow it to collect similar data) to a sufficiently large number of users, which may require significant investments, notably in research and development. Due to network and experience effects and scale economies, building a sufficiently large customer base may not be straightforward as the quality of the service that is proposed depends on the size of the customer base (see point 1 above). In some settings, convincing users of giving access to their personal data could also be difficult: indeed, some people can be reluctant to transfer their personal data unless they are given sufficient guarantees (e.g. secured storage, indications on the use of data, reputation of the service provider) or are provided high added-value services. Furthermore, the ability of a new entrant to build a large customer base may be limited by switching costs which can be high in some sectors, either due to exogenous reasons (if the service in question needs learning for instance) or to the incumbent's behavior, which may then call for antitrust scrutiny.

cc) Factors that may limit the access to data through third parties

Data are not solely collected by companies as an input to gain a competitive edge over their rivals. They can also be collected to be resold on a 'market for data'. Indeed, the last ten to twenty years have seen a surge in the emergence of data

intermediaries,⁸⁰ especially in the U. S., which can collect, store and analyze data for third parties. Examples of such data intermediaries include Acxiom, Datalogix, Experian, Epsilon or Teradata. These data brokers collect data from a variety of sources, including their own data collection technology,⁸¹ arrangements with website owners allowing the intermediary to implement user tracking technologies (such as cookies and pixels), public information (phone numbers, information available on social networks), data from public authorities and third-party companies (websites, banks, online stores, other data brokers). Data brokers' customers are businesses active in various sectors and can include other operators collecting large datasets. For instance, some data brokers have entered into partnerships with companies such as Facebook or Google, in order to help them improve the targeting of their advertising offers.⁸² Conversely, an operator active on a data-related market may find it profitable to give access to the data it collected to other operators in order to extract additional value from those data.

The use of third parties' data may be an alternative to the direct collection of data. Indeed, such an intermediated access to data can be less costly: the fixed costs of data collection are mutualized over a greater number of using undertakings, a company may buy from the data broker only the data that it needs in terms of volumes and variety without incurring a large fixed cost. Furthermore, the services proposed by data intermediaries are numerous and can include data analytics, thereby further reducing the fixed costs associated with data exploitation. Resorting to a data intermediary can also help a company expand the volumes or/and scope of its own datasets or the quality of its data exploitation services.

On the other hand, third party data present several drawbacks. First, the scope of data accessible through data brokers may be limited compared to the volume and variety of data directly collected by the largest operators on a data-related market.

⁸⁰ Data brokers appeared some time ago. For instance, Acxiom, a company selling both physical and digital data, was created in 1969.

⁸¹ For instance, Twitter created a platform giving access to the flow of tweets and to associated raw data. In this context, Twitter first entered into partnerships with data resellers (GNIP, Datasift and NTT), which acted as intermediaries with companies using tweets. After the acquisition of GNIP in 2014, Twitter decided to put an end to those partnerships to set up a direct commercial relationship with companies purchasing these data. Also, Facebook entered into a partnership in March 2015 with Datasift in relation to the processing of topic data from its users, which will be made available to marketers via a Datasift platform.

⁸² For instance, Facebook uses data from databrokers (including visited webpages, subscribed newsletters, money spent online and offline, etc.) to enrich its own datasets about its users' friends and likes (Evelyn M. Rusli, Buy Signal: Facebook Widens Data Targeting, <http://www.wsj.com/articles/SB10001424127887324504704578412960951909032>).

For instance, the behavioral data collected through tracking cookies or pixels may have a limited scope and consequently be imperfect substitutes to personal data entered directly by the user of a service, because tracking cookies only give information on the websites visited by an individual but may not allow to collect detailed socio-demographic information which may be particularly valuable. Facebook builds up detailed user profiles based on highly sensitive personal information of all areas of user's lives (family, education, job, political interests, hobbies etc.) which are given while using the social media platform. Moreover, data exploitation can exhibit increasing returns to scope, where having more diverse kinds of data brings new insights, for instance in terms of customer profiling. Hence, Google is said to benefit not only from a vast amount of data collected through queries on its search engines but also from data collected on its other services such as mail, video services, phones, and so on. Overall, therefore, if third parties are limited to specific items, they could be less relevant to improve an undertaking's performance than an undertaking's own data.

Second, there can be some specific costs supported by data intermediaries which are passed through to their customers.⁸³ Those technical constraints may be even more challenging in the case of data that have a rapidly decreasing value over time, such as may be true for localization data for instance, because the technical solution will have to allow for a frequent update of the dataset provided to the buyer of the data. Those different factors explain why, as set forth by the OECD, *“creators and controllers of data do not necessarily have the incentives to share their data. One reason is that the costs of data sharing are perceived as higher than the expected private benefits of sharing”*.⁸⁴ Furthermore, data users integrated into data collection may benefit from various efficiencies compared to those resorting to independent data brokers. For instance, in its TomTom/Tele-Atlas merger decision, the Commission pointed out that the vertical integration between TomTom (a maker of portable navigational device) and Tele-Atlas (a seller of map database) could allow Tele-Atlas to benefit from the feedback data (such as error corrections) gathered by

⁸³ For example, the OECD points out that *“the provision of high-quality data can require significant time and up-front investments before the data can be shared”*, including “i) datafication, ii) data collection, iii) data cleaning and iv) data curation (...), data models and algorithms for data storage and processing, and even secured IT infrastructures for (shared) data storage, processing, and access”.

⁸⁴ OECD, Data-Driven Innovation: Big Data for Growth and Well-Being, OECD Publishing, Paris (2015), <http://dx.doi.org/10.1787/9789264229358-en>, p.192.

TomTom through its larger customer base (§ 246), although the magnitude of such efficiencies remained uncertain.

Third, those entities that collect valuable data may have only limited possibilities to share them due to legal or contractual provisions that may prohibit or constrain the possibility to share data with third parties. In particular, when collecting personal data, the operator generally guarantees its users that their personal data will not be communicated to a third party without their consent. Without such a guarantee users may be reluctant to communicate their personal data. Privacy rules in most countries strictly limit the provision of data to third parties for commercial purposes. Actually, the more sensitive the data, the less likely it is that it can be collected or transferred to a data intermediary, who, for their most part, remain unknown to the consumers and have less incentives to protect their reputation for privacy protection and therefore are less trusted by consumers.⁸⁵ Overall, the role played by data brokers may vary depending on the country and on the type of data under consideration as well as on the extent of privacy protection enjoyed by individual users pursuant to national regulations.

Finally, the incentives of operators active on the market to share their data with competitors could be insufficient. If the data are valuable, sharing them with competitors could significantly reduce the competitive advantage of the collecting undertaking more than it increases its revenues. As a result, most companies active on the markets on which they collect data are frequently reluctant to share them with competitors. In some cases, this reluctance may go as far as to diminish the ability of third parties to access the same data. For instance, Graef et al. (2015) report that Facebook prevents other undertakings from collecting data on the users of its social network without its consent.⁸⁶ On a more general level, there are several markets on

⁸⁵ See also OECD, *Data-Driven Innovation: Big Data for Growth and Well-Being*, OECD Publishing, Paris (2015), <http://dx.doi.org/10.1787/9789264229358-en>, when it argues that “*better data governance regimes are needed to overcome barriers to data access, sharing and interoperability. These regimes can have an impact on the incentives to share and the possibility of data being used in interoperable ways. The elements to consider for effective data governance include data access and reuse; portability and interoperability; linkage and integration; quality and curation; “ownership” and control; and value and pricing.*”

⁸⁶ Inge Graef/Sih Yuliana Wahyuningtyas/Peggy Valcke, *Assessing data access issues in online platforms* (2015), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2647309, refer to the Facebook’s Statement of Rights and Responsibilities on Safety, which prohibits automatic collection of user content: “*You will not collect users’ content or information, or otherwise access Facebook, using automated means (such as harvesting bots, robots, spiders, or scrapers) without our prior permission*”, available at <https://www.facebook.com/legal/terms>.

which competitors complain about a lack of access to relevant data, regardless of whether this conduct can be deemed as anticompetitive.⁸⁷

To conclude, the fact that data are non-rival does not solve all competition issues associated with data. Indeed, non-rivalry does not necessarily imply that data are accessible to all competitors or, more precisely, that all competitors are equally able to collect these data. These are issues that deserve a case-by-case analysis.

b) The availability of digital data

aa) Presentation

The volumes of data that can be collected have never been as high as today. As the OECD explains, the digitalisation of many activities and the deployment of connected services capturing offline activities lead to a massive increase in the amount of data generated by users⁸⁸ and available for collection by undertakings. In this context, data is said to be everywhere, thereby reducing the risk that they could generate market power.

The “data is everywhere” argument is all the more relevant since the value of data may frequently depend on the knowledge that can be extracted from it, not from the data themselves. If the same kind of knowledge can be extracted from different datasets which may also be obtained through different mechanisms, the risk that an undertaking may not be able to have access to the knowledge enjoyed by his competitors could be low.

Graef (2015) for instance gives the example of a search engine provider getting to know the music preferences of a particular user thanks to his search queries and of a social network provider gaining the same information by looking at the profile information that the user has shared on its platform.⁸⁹ A report to the President of the United States⁹⁰ also states that social network analysis can allow for an ad targeting which may be as good as the targeting derived from search data. Lerner (2014) also

⁸⁷ These include the energy markets already mentioned. Some other cases hinge on the transfer of information product markets onto some aftermarket operators (markets for maintenance of products, for training of employees and so on).

⁸⁸ OECD, *Data-Driven Innovation: Big Data for Growth and Well-Being*, OECD Publishing, Paris (2015), <http://dx.doi.org/10.1787/9789264229358-en>.

⁸⁹ Inge Graef, *Market definition and market power in data: the case of online platforms*, *World Competition: Law and Economics Review*, Vol. 38, No. 4 (2015).

⁹⁰ Executive Office of the President, *President's Council of Advisors on Science and Technology*, , Report to the President - Big Data and Privacy: a Technological Perspective (2014), https://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_big_data_and_privacy_-_may_2014.pdf.

argues that the data collected by Amazon regarding consumers' actual purchases could be as effective as those owned by Google to enhance ad targeting. Also, mobile Internet service providers could access considerable data about consumers' real-time location, close to what Google and Facebook may have.⁹¹ Furthermore, some companies, especially when they do not compete with each other, may agree to cross-exchange the data they collect in order to increase their level of information.

Both the American and the European competition authorities have considered in some merger cases that increased access to certain types of data through acquisitions did not raise any competition concerns because large amounts of data remained available to competitors. Hence, according to the 2008 Google/DoubleClick merger decision by the European Commission, the *"combination of data about searches with data on users' web surfing behavior [generated following the merger] is already available to a number of Google's competitors today"*. Furthermore, the decision stated that *"[c]ompetitors may also purchase data or targeting services from third parties"* and that *"[d]ata is also available from internet service providers"* (§ 365). A similar reasoning was presented by the FTC: *"a number of Google's competitors have at their disposal valuable stores of data not available to Google. For instance, Google's most significant competitors in the ad intermediation market, Microsoft, Yahoo!, and Time Warner have access to their own unique data stores. These firms own popular search engines, and will have access to consumer information from their internal ad servers, ad intermediation services, other web properties, and software. The entry and expansion of these well-financed competitors has transformed the ad intermediation marketplace over the last six months. All of these firms are vertically integrated, and all appear to be well-positioned to compete vigorously against Google in this new marketplace."* (pp. 12-13).

While it may be considered that those decisions are too old to be still relevant in view of the fast moving nature of that industry, more recent decisions have held the same reasoning. In its 2012 Telefonica UK/Vodafone UK/ Everything Everywhere JV joint venture decision, the Commission considered that the joint venture would not have access to a database for mobile advertising that no competitor could match as *"information available to the [new entity] is also available to a large extent to both*

⁹¹ Geoffrey A. Manne and Ben Sperry, The problems and perils of bootstrapping privacy and data into an antitrust framework, CPI Antitrust Chronicle, May 2015, p.11, also via http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2617685 .

existing and new market players such as Google, Apple, Facebook, card issuers, reference agencies or retailers. (...) Customers generally tend to give their personal data to many market players, which gather and market it. Therefore, this type of data is generally understood to be a commodity. (...) Other ways to reach large numbers of consumers exist, such as utilities providers for instance. Additionally, Internet service providers also appear to be in a position to reach a very large number of consumers” (§§ 543-544). In 2014, both the US and the European competition authorities held the same kind of reasoning to unconditionally clear the acquisition of Facebook by WhatsApp. In particular, the European Commission explained that “there are currently a significant number of market participants that collect user data alongside Facebook. These include, first of all Google, which accounts for a significant portion of the Internet user data and, in addition, companies such as Apple, Amazon, eBay, Microsoft, AOL, Yahoo, Twitter, IAC, LinkedIn, Adobe and Yelp among others.” (§ 188). In addition, “there will continue to be a large amount of Internet user data that are valuable for advertising purposes and that are not within Facebook’s exclusive control.” (§ 189).

bb) Discussion: from availability of data to substitutability between data

The volumes and variety of data have been constantly increasing for years. This holds both for the online environment and for the physical world thanks to connected devices. That being said, the relevance of the “*data is everywhere*” argument depends crucially on the accessibility of data (see above) and on the substitutability between data of different types. Appreciating the extent of this substitution can be hard.

For instance, as mentioned above, the information gained by networks on their users may be as rich or even richer than that earned by search engines through user queries. Yet, unlike social network data, the data collected by search engines have the advantage of allowing to identify consumers who are in an active search for a given service or good. A similar ambiguity could prevail when comparing the data collected by offline retailers (obtained through loyalty cards for instance or through the data-mining of check-out receipts) and those collected by online retailers, with the latter including those products a user has considered buying before he made his actual purchase. Likewise, data collected and used through PCs and laptops could not be substituted to data collected and used on mobile phones because only those

data that contain a real time user's location could be fit for use through mobile phones.⁹²

The extent of substitutability between different types of data has been considered in several past cases, although not in direct relation to so-called digital markets. Hence, in its TomTom/Tele Atlas merger decision, where TeleAtlas is a seller of map databases and TomTom a seller of portable navigational device using these databases (decision M.4854, § 22), the Commission concluded that digital map databases for non-navigational purposes were not substitutes to those with a navigational purpose because the latter must include more details (such as road type or traffic information, turn restrictions).

When assessing substitutability between data, competition authorities also consider supply-side substitutability and entry barriers. A case where this criteria has particularly been considered is that of Thomson Corporation/Reuters Group merger decision (decision M.4726). This operation led to significant overlaps for some specific database/content sets on several markets for financial services. Competition concerns arose due to the new entity's high market shares as well as to the costs and lengthy time of building datasets similar as the ones proposed by the entity. Indeed, these data compile contributions from many different brokers and financial analysts, which may be willing to share their research with a distributor only if it has a significant customer base or who may have exclusive deals with either Thomson or Reuters, are retrieved through specific software infrastructure and may need to be standardized to be attractive for customers.

In most merger cases examined by competition authorities in the area of data-related markets, the parties involved used their data solely as an input for their own production, e.g. did not sell their data to third parties, and thus were not active in any possible market for data. As a result, a market for data did not have to be defined.

Nonetheless, as explained above, in its Google/DoubleClick and Facebook/Whatsapp decisions as well as in its *Telefónica/Vodafone/Everything Everywhere* joint venture decision and *Publicis/Omnicom* merger decision, the Commission has tended to consider that in spite of the expanded means of data

⁹² In this context, some analysts consider that mobile ISPs have access to considerable more data about consumers' real-time location, close to what Google and Facebook may have. See Geoffrey A. Manne and Ben Sperry, The problems and perils of bootstrapping privacy and data into an antitrust framework, CPI Antitrust Chronicle, May 2015, p.11, also via http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2617685.

collection induced by those mergers/joint ventures, the amount of accessible data that could be used for advertising purposes or for data analytics would remain sufficient for competitors to match the advantages gained by the merging parties through the operation. On the contrary, in *TomTom/Tele Atlas* decision, the Commission considered that producing a map database for navigational purposes would be very costly and resource intensive because part of the features of such a database would have to be compiled manually by a fleet of vehicles and regularly updated (§ 24). The same conclusion was attained in the *Thomson/Reuters* merger decision. All in all, such considerations could tentatively lead to conclude that data may indeed be easier to collect on digital markets, thereby attenuating any competition concern. On a more general level, one should however keep in mind that Big Data is still in its infancy and that developments in the exploitation of data are probably still to come. Data which could appear as substitutes today may no longer be so in the future and vice versa. Possible distinctions between offline and online data (see above), between data generated through mobile applications and those collected through fixed devices, between data retrieved from social networks and those retrieved from search queries could become more prominent as the means of exploiting these different sorts of data are further developed. By contrast, some data, such as navigational information, which are costly to collect today because they cannot be uploaded automatically through online or connected processes may, may partially or may not (e. g. for data quality reasons) be substituted by data actively and voluntarily collected by a high number of individuals – or in the future by their cars. With regard to mapping data, such a resource exists in the form of the data collected for the Open Street Map project.⁹³

Some of the above mentioned decisions already leave room for more sophisticated analysis of data substitutability. Hence, the Commission already made a clear distinction between offline and online advertising services in its *Google/DoubleClick* and *Microsoft/Yahoo* merger decisions, in particular because of the differing targeting possibilities offered by online advertising (see notably *Google/DoubleClick*, §§ 50-52; *Microsoft/Yahoo!*, §§ 64-65). But more recently, in its *Telefónica/Vodafone/Everything Everywhere* decision, the Commission noted that a vast majority of respondents to the market investigation considered that “*the provision of data analytics services for static online advertising cannot be substituted by the provision of data analytics*

⁹³ www.openstreetmap.org .

services for mobile advertising. (...) Indeed, the two services collect a different type of information and amount of consumer details. For example, the information collected via mobile data analytics is usually more personal, geo-located, and can be cross referenced with call behavior, which cannot be offered by online data analytics to a comparable extent. The online and mobile data analytics services are generally considered as complementary and will continue to be used in parallel” (§ 200). The Commission also considered however that data relative to the geolocation of customers are sufficiently widespread for the new entity not to have any competitive advantage over its competitors: *“some of the most common example of players who gather this type of information are Apple, Facebook or Google (with Google Maps)” (§ 545)* as well as Microsoft.

Likewise, in its Facebook/Whatsapp decision, the Commission noted that “a number of respondents considered that other forms of non-search advertising are not as effective as advertising on social networking websites and notably on Facebook, due to Facebook's large and highly engaged audience and its ad targeting opportunities” (§ 77, underline added), possibly indicating that the data collected by Facebook allow for a better targeting of advertisements than those collected by other operators on the market for online advertising services.

To conclude, the fact that the costs of collecting data have dramatically decreased due to the development of online, mobile and connected offline activities does not imply that all types of data are substitutes for one another.

4. Issues pertaining to the scale and scope of data collection

a) Presentation

The significance of the competitive advantages associated with collecting and exploiting data may hinge crucially on whether those data need to be collected on a large scale and/or a scope, i.e. one that could only be attained by a large and/or diversified incumbent. If that is not the case, it could be asserted that competitors can easily attain the volume of data necessary to benefit from a similar advantage, especially given the availability of data on digital or connected markets and their non-rival character.

In this regard, Lerner (2014) argues that the marginal value of data used for inference purposes can decrease rapidly once a certain amount of data has been collected. Data could therefore display decreasing marginal returns to scale, which

would invite to tame down the competitive advantages resulting from large amounts of data.⁹⁴

This assertion could particularly be relevant 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 predictions. Indeed, predictions are based on probabilities. The higher the number of observations, the more reliable the measure of a probability is and the more reliable a prediction will be. However, the increase in the precision of the prediction decreases as the size of a dataset increases: 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.⁹⁵

For instance, most search engines take into account the observed choices of its users among the results they get for a given search query in order to improve and make more relevant the ranking of those results for future similar queries. If the number of observations is relatively high, the search engine will have a more representative sample of observations on which to rely on in order to improve the ranking of the results with a limited risk of being biased by unusual observations. However, as the number of search queries increases, the marginal informational value of each search query decreases.

Hence, in the context of search engines, the number of necessary searches for an internet search engine to be competitive may be lower than Google's accumulated number of searches. Some considerations in the Microsoft/Yahoo decision tend to support this view, the Commission indicating that "*Microsoft has submitted a benchmarking study that compares its algorithmic search result relevance against that of Google and Yahoo [...]. The results of this study illustrate that Microsoft and Yahoo [...]. However, for the most frequent queries, the overall relevance gap between engines is very small [...]*" (§ 166).

In addition, it should be kept in mind that the ability to extract information from data does not rely exclusively on the amounts of data available but also on the algorithms that analyze data, which are not all of the same quality.⁹⁶ The different levels of

⁹⁴ 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.

⁹⁵ For instance, see Andres V. Lerner, The Role of 'Big Data' in Online Platform Competition (2014), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2482780, p. 35.

⁹⁶ 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. 122.

competitiveness or of quality on a data-related market are therefore not fully attributable to a larger or poorer set of data collected by competitors. To illustrate this, in the aforementioned Microsoft/Yahoo decision, the Commission concluded that *“while (...) Google appears to perform better in terms of relevance especially for [...] queries, this does not provide evidence that scale leads to higher relevance for users, since the above studies do not take into account the technology of the different search engine which are not related to scale”* (§ 168). The role of data analytics in compensating (or exacerbating) a small-sized dataset should therefore not be underestimated.

Another limiting factor of the data advantage enjoyed by larger incumbents relative to new entrants is that the value of data may decrease quite quickly in time. For instance, historical data, while useful for analyzing trends in advertising markets, may have comparatively little value for instant decision making such as the choice of which ad to display in real-time bidding. Moreover historical data may be of relatively low value for some actors like search engines in view of the high rate of new search queries: as reported by Google, 15 % of every day people’s searches are new (see Lerner (2014), § 64), implying that algorithms continuously need new data to be effective in providing the most relevant ranking of results to those new queries.⁹⁷

b) Discussion

As a preliminary point, observers⁹⁸ consider that the strategies followed by some major operators regarding their investment choices and their conducts towards the data from their users show that they are willing to access even more data than they already do, thus suggesting that scale and scope of data collection are key parameters of competition. Yet, there can be several motives behind the acquisition of a given company (such as acquiring a large user base on which to monetize advertising campaigns or new services). Also, the apparent willingness of some companies to collect, combine, store and use ever larger volumes of data could merely indicate that the marginal costs of collecting data are very low. Thus, unless some large fixed costs or legal risks are incurred when processing these supplementary data, the ever larger volumes of collected data are not necessarily

⁹⁷ Note however that Google may remain advantaged relative to new entrants because it has a larger pool of users making these new searches and will therefore upgrade its search engine more rapidly than its competitors.

⁹⁸ See Nathan Newman, Search, Antitrust and the Economics of the Control of User Data, Yale Journal on Regulation, Vol. 30, No. 3, 2014, also available at SSRN: <http://ssrn.com/abstract=2309547> or <http://dx.doi.org/10.2139/ssrn.2309547>

proof, in themselves, that a large scale of data collection is indispensable to compete on the related markets.

Still, although several of the arguments developed above indicate limits to the advantages of the scale and scope of data under some circumstances, their universal validity is disputable.

First, as regards the impact of data obsolescence, it should be kept in mind that although the value of some data may diminish relatively quickly over time, not all data are transient in value. Depending on the market, some data, such as gender, names, address, date of birth, job, etc., may not lose their value over time. Thus, a company having such data at its disposal may have a lasting advantage over its competitors. Furthermore, if collected data have to be updated very frequently, undertakings may then have to collect their volumes of data in a very short time, thereby increasing the data constraint.

Second, the decreasing marginal value of data results pointed out by Lerner and others in the case of search engines and recommendation systems derives from the fact that these data are used for inference purposes. However, not all data are used only for inference purposes. For instance, in more 'classical' markets, datasets often compile contact information and usage information. In those cases, each observation in the dataset may have the same importance as each of them represents a new prospect.

Third, to make useful inferences, companies are interested both in the quality of inferences that can be made thanks to a given dataset as well in the quantity of observations contained in the dataset. For instance, a small dataset may have a sufficient size to help an advertiser or an ad agency make inferences about which consumers are likely to buy a given good. Yet, for this inference to be valuable, the undertaking needs to have matching information about a large pool of individuals. For instance, thanks to a relatively small dataset, an undertaking may have inferred that brown-haired women buy a certain kind of make-up. For that inference to be of any use, the undertaking would still need a large dataset describing the hair colour of its potential customers.⁹⁹

⁹⁹ See D. Evans, The online advertising industry: economics, evolution and privacy, *Journal of Economic Perspectives*, 2009, 23 (3), pp. 37-60.

Fourth, decreasing returns to the scale of data collection are not informative of the amount of data under which an undertaking may be disadvantaged compared to incumbents. That level may be quite high for at least some activities. For instance, considering search advertising as opposed to the search engine itself, the European Commission noted in the context of the Microsoft/Yahoo merger investigation (2010) that almost all advertisers considered that user scale was an important factor to effectively compete in search advertising since it ensures a larger audience.¹⁰⁰

Fifth, the volume of data necessary to make a given inference may be relatively low. However, on some markets, the number of inferences that have to be made daily is huge, thereby dramatically increasing the volume of data that is necessary. For instance, search queries are quite diverse and the number of different search queries is very high. Therefore, a significant proportion of search queries can be quite rare at the level of a search engine. In this context, the low number of users of small search engines making rare queries does not allow them to improve the relevance of their ranking, leading consumers to opt for another, larger, engine for their whole set of queries. The fact that users' queries and behaviours evolve at a high pace makes the number of queries received by a search engine even more valuable in order to identify these changes, adapt the search results and thus improve their relevance in a short period of time.¹⁰¹

The same kind of reasoning could apply for other types of websites (think for instance of recommendations for online purchases). Therefore, a relatively large scale of data collection can contribute positively to the competitiveness of an online platform.

Finally, depending on the usage of data, the *scope* of a dataset may be as important as its *scale*: the value of a dataset does not rely only on the number of individuals it describes but also on the amount of information collected for each individual. In this regard, data collection through the offering of different services allows the collecting entity to gather knowledge on the multiple aspects of users' behaviour and tastes.

¹⁰⁰ "The first phase market investigation tends to support the view of the notifying party that scale is an important element to be an effective competitor. Almost all advertisers responding to the market investigation consider that Microsoft does not have enough traffic volume to be an attractive alternative to Google" (§153). "All the advertisers responding to the market investigation consider that Google is a 'must have' for search advertising campaigns. Also, all of the advertisers responding to the market investigation highlight the fact that Google's query volume is one of the main reasons why Google is a 'must have'" (§157).

¹⁰² Newman, Nathan, Search, Antitrust and the Economics of the Control of User Data, Yale Journal on Regulation, Vol. 30, No. 3, 2014, also available at SSRN: <http://ssrn.com/abstract=2309547> or <http://dx.doi.org/10.2139/ssrn.2309547>.

While each of these aspects can separately be of interest for an advertiser or any sales company, their combination may result in a better knowledge of each user, in order, for instance, to better infer the probability that a user is likely to be interested in buying a given product at a given point in time.

The value of the scope of data has been highlighted in various reports. For instance, the report to the President of the United States indicates that “[the combination of data from different sources] *may uncover new meanings. In particular, data fusion can result in the identification of individual people, the creation of profiles of an individual, and the tracking of an individual’s activities*”.¹⁰² In its Google/DoubleClick merger decision, the European Commission similarly stated that “*competition based on the quality of collected data thus is not only decided by virtue of the sheer size of the respective databases, but also determined by the different types of data the competitors have access to and the question which type eventually will prove to be the most useful for internet advertising purposes*” (§ 273). Likewise, the Information Commissioners’ Office, i.e., the UK’s independent authority in charge of upholding information rights in the public interest, stated that, according to some practitioners, “*of the ‘three Vs’, variety is the most important characteristic of big data*”.¹⁰³

To conclude, the advantage associated with an access to a larger volume of data may be quite different from one market to another. That calls for a case-by-case assessment.

V. Conclusion

The use of data is not a new phenomenon, as past assessments by competition authorities of the “data advantage” in non-digital markets illustrate. Maintaining a customer database, conducting consumer surveys and market research have long been staples of every business activity. However, technical progress and the

¹⁰² Newman, Nathan, Search, Antitrust and the Economics of the Control of User Data, Yale Journal on Regulation, Vol. 30, No. 3, 2014, also available at SSRN: <http://ssrn.com/abstract=2309547> or <http://dx.doi.org/10.2139/ssrn.2309547> .

¹⁰² Executive Office of the President, President’s Council of Advisors on Science and Technology, , Report to the President - Big Data and Privacy: a Technological Perspective (2014), https://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_big_data_and_privacy_-_may_2014.pdf .

¹⁰³ Information Commissioner’s Office, Big data and data protection (2014), <https://ico.org.uk/media/for-organisations/documents/1541/big-data-and-data-protection.pdf> , § 25. Also see Networked European Software and Services Initiative (NESSI), Big Data: A New World of Opportunities, White Paper (2012), http://www.nessi-europe.com/Files/Private/NESSI_WhitePaper_BigData.pdf .

digitalization of the economy have expanded the nature (e.g., real-time location data fed by smartphones), sources (e.g., cross-device tracking of a user's web journey), applications (e.g., machine decision-making and learning) and volume of data.

The economic relevance of data as an input has already been considered and, in several instances, acknowledged by competition authorities. To date, the risk of foreclosure associated with the concentration of data in digital industries has mostly been looked at in the context of merger control. This does not exclude the use of antitrust enforcement tools to tackle behaviour related to the collection and processing of data, similarly to what has already occurred in some non-digital markets. There are several possible "data-based" conducts, whether exclusionary or exploitative, which could, depending on the circumstances of the case, lead to enforcement action.

However, the theories of harm underlying the prohibition of such conducts are premised, for the most part, on the capacity for a firm to derive market power from its ability to sustain a data trove unmatched by its competitors. A case-specific assessment of the reality and extent of the "data advantage" needs to be undertaken to bear out or reject this premise. In doing so, consideration should be given at the outset to the features which are particularly found in online markets (network effects, multi-homing, and market dynamics) which may or may not be conducive to market power, before proceeding to determine whether data contributes to the creation or strengthening of market power.

Two aspects of particular relevance when looking at data's contribution to market power can be identified: the scarcity of data or ease of replicability, on the one hand; whether the scale/scope of data collection matters, on the other.

Data is "non rivalrous" in the sense that access to data by an operator does not, in and of itself, preclude access by other operators. Multi-homing by customers as well as the diversification of services offered by a single firm provides opportunities for the concurrent collection of user-specific data. However, accessing this data in the first place may be conditioned on the capacity for the firm to build a sufficiently large customer base, which in turn depends on the extent to which network and experience effects as well as scale economies act as barriers to entry. The availability of data from third parties, such as data brokers, can counteract data accessibility concerns, but the impact of such external sourcing may vary depending on the nature of the

data concerned, applicable rules – contractual or regulatory – protecting user privacy and the general reluctance of firms to share their “data advantage” with competitors.

The ongoing digitalization of the economy and the deployment of connected devices capturing offline activities have dramatically increased the amount of data available for collection. Accordingly, this has led to the assertion that “data is everywhere” and therefore cannot constitute an input prone to hoarding and foreclosure. However, this can hold only if the said data is, in fact, accessible (see previous paragraph). This assertion also depends crucially on whether different types of data are substitutable – and absent substitutability, whether each category of data is widely available for collection. Empirical findings and previous analysis of data substitutability by competition authorities point to differences which must be taken into account on the basis of their significance for the individual case at hand (e.g., between mobile and static data; between data retrieved from search queries and data retrieved from social networks; between transactional data and data evidencing purchasing intentions which did not materialise).

Finally, the scale and scope of data required must be ascertained. The relevance of data as a strategic input and the opportunities for foreclosure depend in part on the volume levels: (i) at which a firm can reap the economic benefits of data; (ii) beyond which these benefits decline or cease to exist altogether. These levels will vary, depending on the type and purpose of the data. For instance, the marginal value of customer contact and socio-demographic information may tend, in several industries, to remain somewhat constant (subject to updating requirements), whereas the value of data used for inference purposes will tend to decrease, beyond a threshold, in tune with the rate at which statistical sampling error declines with every extra data. However, a significant and recurring volume of data may need to be collected before attaining this threshold, for instance if an activity must account, to remain competitive, for high pace changes and “long tail” occurrences. The scope of data may likewise prove as important as scale and can warrant, depending on the market conditions and the case at hand, further scrutiny.