

Big Data Competition and Market Power*

*Marco Gambaro***

ABSTRACT: Big data are considered at the same time a promising driver of economic development and a concern for possible manipulation and privacy intrusion. Data diffusion and their uncertain appropriability can make property rights regarding data less precise than those regarding traditional goods.

The article reviews some economic features of data. In many digital markets data can be considered a relevant input for production but hardly an essential facility.

Many data are collected in two-sided market platforms and on the one side, they are used to personalise services and to add quality, while on the other side of the platform they contribute to make advertising collection more efficient. So, the transfer of personal data can be considered an implicit price for many free information services. Consumers are usually unaware of subsequent pervasive use of their personal data, and therefore give them away easily. Big data can amplify competitive advantages and related dominant positions, leveraging on information asymmetries.

A dominant position obtained through collection and processing of big amounts of personal data allow practices such as first-degree price discrimination, personalised advertising, and artificial degradation of services that can sometimes be considered competitive abuse, but it is difficult that data alone allow to maintain a true dominant position.

KEYWORDS: big data, data protection, personal data, privacy, value of data, value of privacy, relevant market for data

* Date of reception: 30 July 2018. Date of acceptance: 15 August 2018.

** Marco Gambaro is an Associate Professor at the Department of Economics, Management and Quantitative Methods at the Università degli Studi di Milano, 20122 Milan, Italy. marco.gambaro@unimi.it.

1. Introduction

For some years, big data have been considered one of the most promising drivers of economic development; yet, more recently, they are a source of concern for the possible market power accumulated by those who collect large amounts of data and have the ability to process them.

Big data are at the centre of a cluster of innovations that interact with each other and which, together, drive the digitalisation of markets and society. All are based on technological innovation and cost reduction of processing and communication equipment. Cloud computing can in fact transform into variable costs what once were fixed costs of infrastructure, processing and computer memories. This drastically lowers the economies of scale for the birth of digital start-ups and the scalability of resources accompanies growth without requiring significant specific investments. Successful companies, such as Huber or Netflix, started using computing power or vector maps provided by Amazon and Google, among others.

Open data refer mostly to data generated by transactions and procedures of public administrations and are generally big (census, fines, access to offices), but sometimes small (opening shopping centres, homicides), and can sometimes be part of the Internet of Things (data on traffic lights). When these large amounts of data are made available to researchers and businesses, innovations can emerge that lead to new ideas and new services. Open data are therefore a potentially powerful driver of innovation.¹ But usually public administrations do not easily spoil themselves from the exclusive control of data, and in many countries these kind of data are not easily accessible.² Many big data consist of information generated by machines and equipment and are generally included in the Internet of Things.

By Internet of Things we generally mean the diffusion of sensors and regulators in devices and objects that in addition to performing a specific measurement or regulation action produce a large amount of data on the phenomena to which they are connected. The most reasonable estimates speak of 25-50 billion connected devices within the next 10 years. This is an order of magnitude greater than the number of devices currently

¹ Mark Huberty, "Awaiting the second big data revolution: From digital noise to value creation", *Journal of Industry Competition and Trade* 15 (2015): 35-47; Hal Varian, "Big data: New tricks for econometrics", *Journal of Economics Perspectives* 28, no. 2 (2014): 3-28.

² Ron Jarmin and Amy O'Hara, "Big data and the transformation of public policy analysis", *Journal of Policy Analysis and Management* 35, no. 3 (2016): 715-721.

connected to the network. Before data can be properly explored there are several problems that need to be addressed. First of all, devices need to be connected (Wi-Fi and mashup), and the same networks may have problems of saturation, not so much for the bandwidth employed, which is generally low, but mainly for latency and signal capabilities. A further problem can be the electric power supply needs. These data may be more useful if all these devices have an IP address that allows them to be managed as elements connected to the network, also for data collection and distribution.

Then, there are all the data of a more personal nature, including data generated by transactions (financial, administrative and commercial), those related to surfing the net, and all user-generated contents (email, social networks, images, video, web content, free text), which are generally in an unstructured form and require semantic analysis in order to be used.³ It is possible to find many optimistic forecasts for the economic values of big data that are expressed in very large numbers, but for now there are more assumptions than predictions. Generally, these forecasts are made with the logic: if the manufacturing costs of the entire economy fall by 1% then the value would be X. Thus, for example, Deutsche Bank hypothesises in the oil industry a 1% reduction in capital expenditure which would result in USD 90 billion of save costs; in smart utilities and aviation, a 1% reduction in fuel used translates, respectively, into USD 66 billion and USD 30 billion of cost reduction.

Automotive industry, utilities, connected houses, the industrial sector and healthcare are usually considered the most promising sectors. While the theoretical potential gains are spread horizontally, almost onto the entire economy, concrete developments will be much more localised and will depend on specific competitive conditions of single applications and markets. In all these potential areas for digital development, there are significant network economies that do not make it easy to start up and catch the potential benefits. In many cases, companies must first install a large number of devices, sensors or regulators in order to collect a sufficient amount of data, find some suitable algorithms to properly use the data, solve the technical problems of processing large amounts of data, and hope for the benefits in the end.

³ Marco Gambaro, "Some economics of new media content production and consumption, and strategic implication for media companies", in *Handbook of Social Media Management*, ed. Mike Friedrichsen and Wolfgang Mühl-Benninghaus (Berlin: Springer, 2013).

The possibility of fully internalising potential benefits is a powerful accelerator for efforts and investments, while in those cases where monitoring generates a better service for third parties or final customers, misalignment of incentives and potential opportunistic behaviour can slow down adoption. In many cases the prevalence of private benefits at the beginning can speed up the initial investments, while many positive externalities make it difficult to find resources. For example, if through the sensors installed in a car manufacturers are able to predict the possible breakdowns and schedule maintenance work, they will also have to convince customers that they are not suggesting these maintenance operations in order to maximise car repairs or spare parts budgets. New big data application will generate information asymmetries between buyer and seller and will allow the latter possible opportunistic behaviour.

In automotive big data application a general and potentially very significant problem of data ownership arises. If several sensors are installed in my car which monitor driving actions from the moment I purchase it from a manufacturer, are the data generated my property or are they available for the manufacturer? And if the manufacturer refuses to repair the car if I do not make the data available, is he abusing his position? If, in the same car, an insurance company installs a GPS that monitors the use of the vehicle, can the car's sensor data be used, and by whom? And who can control these data: the insurance company that installed the GPS, the client that owns the car, or the manufacturer that builds it? These are regulatory problems that are not easy to solve.

Big data, together with other digital innovations, are a prerequisite and a driver for the emergence of new markets. This happens, for instance, in the sharing economy, where, even if the starting point is the presence of underutilised capital goods (homes and cars)⁴, without the widespread use of smartphones, the availability of traffic data, high-quality photography technologies, low-cost mapping technologies, services like Huber or Airbnb would not even be thinkable.

However, the development of big data is not an easy trial and it is necessary to overcome considerable troubles both in the case of data produced by

⁴ Georgios Zervas, Davide Proserpio and John Byers, "The rise of the sharing economy: Estimating the impact of Airbnb on the hotel industry", *Journal of Marketing Research* 54, no. 5, (2017): 687-705.

objects and by personal data.⁵ When not used to improve the specific transaction but to predict the performance of markets or activities, big data do not escape the traditional statistical traps. Although data volumes are very large and low-cost, they are always samples, not the universe. Secondly, today's actions and information cannot be easily extrapolated to tomorrow and, finally, what happens online is not necessarily the same as what happens offline.⁶ In this context, the fear that such large amounts of data could be used in improper ways and that an asymmetry of power could emerge between the companies that collect this data and the consumers who would be at their mercy is also emerging. Some recent acquisitions such as that of WhatsApp by Facebook, or that of Double Click by Google, have exacerbated these concerns and pushed the requests for regulatory intervention both by the legislative bodies and by the different authorities involved in this area, mainly antitrust and privacy.

The multiple applications of big data bring many benefits to society. They provide companies with valuable information that can be used to improve products and services, while consumers have access to more and often very targeted or tailored information which can help them make better decisions. Examples of big data applications are online search engines, targeted advertising on social media, medical tools that intelligently combine different data sources for diagnostic purposes, and autonomous driving cars which rely on machine learning and a large volume of data.

Although the benefits of big data are clear, there are also concerns expressed in the public debate that companies that have access to data may become too powerful. For some authors, data and algorithms can facilitate collusion and the companies that have exclusive access will direct markets.⁷

This contribution is organised as follows: the next paragraph describes some economic characteristics of data, the third paragraph reviews the possibilities of antitrust authorities' intervention, the fourth discusses the definition of relevant market, and the fifth examines the problems of the ex-ante regulation and discusses some policy implications.

⁵ John Horton and Richard Zeckhauser, "Owning, using and renting: Some simple economics of the 'sharing economy'", *NBER Working Paper 22029* (2016).

⁶ Maurice Stucke and Allen Grunes, *Big Data and Competition Policy* (Oxford: Oxford University Press, 2016).

⁷ Ecorys, *Big Data and Competition* (Report for the Ministry of Economic Affairs, Rotterdam, The Netherlands, 2017).

2. Economic data characteristics

There are many possible definitions of big data that relate to different features of data and information. A well-established approach refers to the large dimension of the datasets and the need to use large scale computing power and nonstandard methods to extract value therefrom.⁸ According to De Mauro: “Big Data is the information asset characterized by such a high volume, velocity and variety to require specific technology and analytical methods for its transformation into value”.⁹

Sometimes it can be important to stress the difference between data and information in general in terms of volume, velocity, variety and value.¹⁰ The volume of data processed expands almost exponentially; many media, social and economic activities migrate on the internet and this generates tons of data every second. The velocity at which some firms access and analyse data is approaching real time. This kind of nowcasting can give the first movers a great advantage over newcomers. The variety of data is enabled by the spreading of digital technologies, which generate data as a spill over, and by capabilities of data fusion.

Data and information exhibit some specific economic features. First of all, they are not rivals nor inappropriable goods. Giving someone a piece of information does not preclude them from giving it to others nor reduce the subsequent possibilities of its use. The same information can be used several times. However, although data is non-rivalrous, it can be exclusive: parties can be practically or legally excluded from access to data. For example, in order to obtain access to data it may be necessary for a company to build a sufficiently large customer base. There can be economic barriers that prevent a company from obtaining users due to network and to experience effects as well as scale economies. Also, personal data are subject to special data protection rules, which limit the gathering, processing and usage of such data. Data protection rules also limit the possibilities for third parties (“data brokers”) to trade data. Companies can make the decision not to provide access to data or may be forced not to provide access to data. A certain degree of exclusivity is required for private actors to invest in the production of scarce goods. With regards to the production

⁸ OECD, *Big Data: Bringing Competition Policy to the Digital Era*.

⁹ Andrea de Mauro, Marco Greco and Michele Grimaldi, “A formal definition of big data based on its essential features”, *Library Review* 65, no. 3 (2016): 122-135.

¹⁰ Stucke and Grunes, *Big Data and Competition Policy*.

of knowledge and information, exclusivity is sometimes arranged legally via patents or copyrights.¹¹

Secondly, data and information are assets with high fixed costs and low or no variable costs. The related markets are therefore characterised by strong scale economies.¹²

Thirdly, in markets with intensive data use, these data can usually be considered a relevant input, but certainly not the only one. To produce knowledge and value we also need other resources including computing skills, specialised processing software and qualified human resources to read and design results, computer scientists and data analysts.¹³ But often the most important barriers to a profitable use of the available data are those related to consolidated managerial models and the scarce capacity to make organisational processes flexible. This is why digital organisations that consolidated themselves with a culture of data use often have a flexibility advantage over established companies that have to undertake important and difficult organisational processes of change. From an economic point of view, a relevant aspect of data and information is the possibility of replacing other expensive resources in production and exchange processes. The examples are many and concern both the substitution of human resources and other types of input. Within the sales department, recommendation systems can replace physical stores and clerks, while in logistics the digital monitoring and planning of the routes allow to save fuel; in the oil industry, preliminary analyses of the extractive potentials save much drilling.

There is therefore a greater efficiency of many physical production processes and at the same time a better design of products and services offered to consumers.¹⁴

With respect to traditional data, big data are more pervasive and easier to collect and process in large amounts, thus lowering collection costs. In fact, since many human activities are carried out with the support of digital applications, the collection of information requires very few additional monitoring costs. In many cases, information gathering is a by-product

¹¹ Ecorys, *Big Data and Competition*.

¹² Gambaro, "Some economics of new media content".

¹³ David Balto and Matthew Lane, "Monopolizing water in a tsunami: Finding sensible antitrust rules for big data" (2016), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2753249.

¹⁴ Michele Polo, "Consumer's search in the era of big data", *Competition Policy International Antitrust Chronicle*, 2017, <https://www.competitionpolicyinternational.com/wp-content/uploads/2017/05/CPI-Polo.pdf>.

of a main separate activity. Many personal data are collected through two-sided platforms which, on the one hand, use the data to create and customise the service offered to customers for free or strongly subsidised and, on the other hand, use them to make advertising more effective, more productive and therefore more profitable.

Baxter presented one of the first formal analyses of a two-sided market, showing that the payment card provided a service only if both cardholders and merchants agreed to use a card for a transaction and explored the consequences of this joint demand.¹⁵ Rochet and Tirole showed that businesses such as media, dating clubs, shopping malls and videogames are all based on a two-sided platform framework. A market can be considered two-sided if the platform can affect the volume of transaction by charging one side of the market more and reducing the price paid by the other side by an equal amount. Price structure is a crucial feature of this business model and platforms must design it in order to bring both sides on board.¹⁶

In two-sided markets, profit maximising prices do not necessarily take into account marginal costs for both groups of customers, and prices below marginal costs can be non-predatory. Profit maximizing prices in two-sided markets depend on the price elasticities of demand on both sides, the magnitude of indirect network externalities and the marginal cost for both sides¹⁷. It is perfectly competitive for a newspaper, a television or a search engine to subsidise the information side in order to collect audience for the advertising side and this cannot be considered predatory pricing.

A fundamental feature of platforms is the presence of network effects: platforms become more valuable as more consumers use them. As more users engage with the platform, the platform becomes more attractive to potential new users. This goes a long way toward explaining why some platforms have had viral growth. There are two kinds of network effects: direct network effects (where more users beget more users, as in more Facebook users will beget more Facebook users); and indirect network effects, where more users of one side of the platform (for example, video game users) attract more users on the other side of the platform (in this

¹⁵ William Baxter, "Bank exchange of transactional paper: Legal and economic perspectives", *Journal of Law and Economics* 34, no. 2 (1983): 309-328.

¹⁶ Jean-Charles Rochet and Jean Tirole, "Platform competition in two-sided markets", *Journal of the European Economic Association* 1, no. 4 (2003): 990-1029.

¹⁷ David Evans and Richard Schmalensee, "The industrial organization of markets with two sided platforms", *NBER Working Paper* no. 11603 (2005).

example, video game developers). It is important to understand that with platforms, scale is both the outcome of initial success and the engine for further growth.¹⁸

In this context, data could be considered an implicit price for the offer of those subsidised services. And for some scholars, data could even represent a new currency.¹⁹ First of all, it should be pointed out that in two-sided platforms, the implicit price is mainly represented by the attention paid to advertising,²⁰ and data are indeed an accessory input to make advertising more effective through profiling and segmentation. Secondly, data cannot be considered a real currency in the economic sense because they are neither a scarce resource nor do they have a uniform value recognised by all the participants to the market. Of course, the loss of privacy generated by the transfer of data is a disutility for the consumer, but not enough to define them as a currency.

The value of data is very contingent and depends on different factors. Single data hardly have much value, while a large collection of data can be valuable. The timeliness of data can be important because often, especially for operational uses, the usefulness and value of data decrease over time. Consumers' willingness to sell data changes according to the trust in the counterpart and its credibility. The same user can evaluate the transfer of their data in different ways according to the interlocutors and the context. Many signals seem to indicate how the contingent value of own personal data is fairly low for consumers, also because the transfer of data is never exclusive. However, there are no structured assessments on the size of the value and on the determinants of the changes.

Much research shows that consumers are concerned about the use of their personal data. According to Pew Research (2016), 71% of American adults think that consumers have lost control over how their data are collected and used by companies and few express the belief that their data remain private and secure once they have been collected.²¹

¹⁸ Annabelle Gawell, "Big data: Bringing competition policy in the digital era", *Working Paper*, OECD (2016).

¹⁹ David Evans and Richard Schmalensee, "The antitrust analysis of multisided platform business", *Coase Sandor Working Paper Series*, University of Chicago Law School (2012).

²⁰ Marco Gambaro, "Concorrenza e pluralismo nel mercato di Internet: La prospettiva economica", in *Le Libertà Fondamentali nell'Era di Internet*, ed. Marco Bassani, Oreste Pollicino and Tommaso Frosini (Milano: Mondadori, 2017).

²¹ Pew Research, "The state of privacy in post Snowden America", Pew Research, 2016, <http://www.pewresearch.org/fact-tank/2016/09/21/the-state-of-privacy-in-america/>.

Naturally, the consumer might be able to choose which data to sell and in which contexts, and therefore the situation that is observed is simply the representation of consumer preferences. But this kind of market discipline and this possible role of the consumer require two important conditions: the availability of complete information on the various contingencies and the presence of real and possible alternatives that can trigger a competitive market. In many situations these two conditions are not verified. Many empirical researches, as well as daily experience show how consumers are not informed about the implications of the transfer of their personal data, either because reading very long written disclaimers in very small characters is practically impossible for anyone, or because producers do not clarify the different transfers and processing of the data and the possible subsequent uses.²²

Even from an economic point of view, consumer concern is not irrational because, if information is power, the information asymmetry created with companies can shift the balance of power in future transactions. If sellers are able to tell how little I know about the products I want to buy, they will try to direct me to a product that is more convenient for them than for me. Indeed, companies do not have socially optimal incentives to aim for a correct matching between products and customers.²³

The transfer of data by consumers generally takes place in exchange for an immediate benefit (a service or a product) and involves uncertain costs, spread in a future and little-known period. It is therefore not clear that the consumer is able to correctly evaluate the trade-offs associated with the transfer of information.

In recent years, in particular, concerns have emerged regarding the collection and use of personal information, a problem which, as we have seen, has an economic dimension, but which extends to other areas of social interaction. In fact, the absence of relevant information asymmetries and the availability of reliable and reasonably objective information are a prerequisite for the proper running of the political decision-making system, of the financial market and of course for the operating of the markets of goods and services.

And so, if a large part of social interactions are mediated by platforms that collect personal data and monitor consumer behaviour and

²² Alessandro Acquisti, Curtis Taylor and Liad Wagman, "The economics of privacy", *Journal of Economic Literature* 54, no.2 (2016): 442-492.

²³ Acquisti, Taylor and Wagman, "The economics of privacy", 442-492.

information choices, concerns arise precisely regarding all these inequalities and information asymmetries. Moreover, it is difficult to maintain that the information collected is relative and proportional to the services provided. Applications on mobile phones provided by the largest digital platforms require, among other things, the possibility to modify the address book and the contents of the device, the possibility to activate the microphone and the camera, to read all the contents of the memories, download files without notification, and send communications without the owners' knowledge. This information collection represents a particularly high degree of potential control whose results are beginning to be seen. A man was talking to a friend face to face complaining about plant pests and, shortly after, he received the personalised advertising of an insecticide. The conversation had evidently been intercepted by the phone's microphone and some robot had recognised the relevant keywords.

Large digital platforms, based on the analysis of traffic and interaction data, are able to send personalised content, capable of influencing the mood and orientation of users. But what happens if who controls a platform decides to run for a public office or decides to support any candidate?

The collection, analysis and use of large amounts of personal data bring out problems that simultaneously affect the economic behaviour of markets, generate inequality and ease intrusion into the personal sphere.

When a car that is now equipped with a digital control unit and a GPS collects data on wear and tear of different components and on the driving style, how one can be sure that the suggested maintenance interventions are based on the real need and not to maximise the revenues of its repair services centres? And who is the owner of these data: the manufacturer, the insurance company, or the owner of the car? Privacy and data sharing therefore represent a terrain at the intersection of different disciplines and regarding which it is possible to think of interventions of a different nature that may concern the promotion of competition, the defence of intellectual property of personal data, consumer protection or defence of privacy as a fundamental right.²⁴

²⁴ Wolfgang Kerber, "Digital markets, data, and privacy: Competition law, consumer law, and data protection", *Joint Discussion Paper Series in Economics* 14 (2016).

3. Range of antitrust action

Data can generate market power in several ways. The presence of direct and indirect network effects can lead to a snowball effect where the markets tip in favour of a small number of players. The learning curves are strong and steep with the growing importance of self-learning algorithms which may substantially increase the first-mover advantage. The multi-sided relationship between different markets may ease the leverage of a dominant position from one market to the other.²⁵

The German Competition Authority initiated an investigation in March 2016 into the suspicions that Facebook's specific terms of service on the utilization of user data could represent an abuse of its possibly dominant position in the market for social networks, and, in December 2017, issued a statement of objections where it is assumed that Facebook is dominant on the German market for social networks, and that the undertaking is abusing its dominant position by making the use of its social network conditional and being allowed to limitlessly amass every kind of data generated by using third-party websites and merge them with users' Facebook accounts.

Access to data may be one of the reasons why a company has a competitive advantage over its rivals. This advantage can be persistent if data result in entry barriers. This is the case 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. Data can be an important input contributing to market power but, as follows from the defined characteristics, there can be other important inputs as well. Besides, in many cases, data can be non-rivalrous, ubiquitous, and with low barriers to entry. In other cases, data are at least to some extent exclusive and without substitutes. This makes a case by case analysis necessary whereby the characteristics of the data and the business model in which they are used need to be considered. Companies with market power can behave in ways that harm consumer welfare. However, market power can also bring benefits. Companies in industries with economies of scale, for example, have some degree of market power which is efficient from a welfare perspective, since costs per product or service are lowered. More generally, in many markets, companies have a certain amount of market power, and entrepreneurial

²⁵ Marc Burreau, Alexandre de Streel and Inge Graef, "Big data and competition policy: Market power, personalized prices and advertising", *Project report, Centre on Regulation in Europe* (2017).

activity often corresponds to developing a unique business proposition which temporarily results in market power. Yet, problems may occur if market power becomes substantial or is sustained for too long due to entry barriers. In such situations, companies may become dominant, which, in turn, may give rise to abuse of dominance. Assessing harm arising from market power has, however, become more difficult in data-driven markets, which are often multisided markets. This is because in multisided markets, prices reflect ways to generate network effects, possibly irrespective of underlying marginal cost levels. As such, it has become more difficult to assess if a price is above the competitive price, because such a reference point no longer exists in these markets.²⁶

Some recent acquisitions by large platforms, such as WhatsApp by Facebook or DoubleClick by Google, have generated much debate on the possible role of Antitrust in the control of possible abuse of dominant positions by the platforms that collect large amounts of data.

The underlying reasoning roughly assumes that the fundamental mechanisms of the internet lead to a progressive concentration of the market mainly due to personal data collected which, beyond a certain threshold, would make it impossible for a new entrant to replicate the skills and data of the leaders.²⁷

In this view, the collection of personal data is characterised by economies of scale in the sense that the larger the collection of data, the greater the value one can extract. According to this line of reasoning, data could be considered as an essential facility in the future. In the value chain of big data, which includes collection, storage and analysis, there are many direct and indirect network effects that can be identified by the antitrust authorities. This requires an understanding of the interactions and possible feedback loops between data and other inputs in a sense of looking at the whole supply chain rather than analysing single parts separately²⁸. Looking at market shares is often misleading because this kind of competition is mainly dynamic and for the market.²⁹ The recent history of the internet is full of cases in which a strong control over data relevant to that

²⁶ Ecorys, *Big Data and Competition*.

²⁷ Nathan Newman, "Search, antitrust and the economics of the control of user data", *Yale Journal of Regulation* 31, no. 2 (2016): 401-454.

²⁸ Bureau, Streef and Graef, "Big data and competition policy".

²⁹ Giuseppe Colangelo, "Big data, piattaforme digitali e antitrust", *Mercato Concorrenza e Regole* 18, no. 3 (2016): 425-460.

particular market could not defend the position of an incumbent against a competitor able to have a better business idea. For instance, the accumulation of search data by the former leader Yahoo! did not prevent Google from prevailing thanks to a better idea for site ranking. Facebook outperformed both Friendster and Myspace because it organised the social network more effectively for users. Dating sites are another example of market share instability. This is one of the oldest online sectors where the use of personal data has always been intense to feed the matching algorithms. In 2006 Yahoo! Personal was the undisputed leader with a user base twice as large as any competitors'. But instead of maintaining permanent dominance, it abandoned the market less than four years later.³⁰ According to a survey by Pew Research, 6 of the 10 most popular dating sites in 2005 no longer existed in 2013. Today the market is headed by a few dominant sites, but there are over 2500 online dating sites only in the United States. The network effects that could push toward concentration have been superseded by users' multihoming, as many customers maintain their profile on several sites, and by specialisation towards a specific audience segment. Of course these examples cannot be generalised, and it is possible that in many situations the exclusive control of data can lead to exclusion of competitors and monopolisation of the market.

When the search results for publisher sites disappeared from the top positions of the search engine in the course of a trade dispute between Google and Hachette, many observers suspected that the natural search results could also be driven by strategic interests. Therefore, the information offered to consumers can be slanted to accommodate profit maximisation aims of search engines and become less credible.

From the review of the literature and the analysis of the different markets involved, it emerges that the assessment of competitive effects of the control of big data is contingent and depends on many factors, including the specific network effects and the interactions between online and offline. It is therefore necessary to carefully evaluate eventual market power on a case-by-case basis and assess the possible role of the controlled and whether these constitute an obstacle to competition. But for a market leader to maintain a sustainable competitive advantage over time, potential

³⁰ David O'Connor, "Understanding online platform competition: Common misunderstandings", *Internet Competition and Regulation of Online Platforms, Competition Policy International* (2016).

competitors need to be unable to realistically replicate the benefits of a specific strategy or input.³¹

This is only rarely achieved in the case of data. The fact that data are a non-rival good means that even if the consumer has sold them to a company, the company can easily give them to others. In addition, there is a flourishing intermediate market for aggregators who gather data from various sources and resell them to businesses. Although from consumers' point of view this market can be questionable and problematic, its presence makes it difficult to implement exclusionary practices through data. The use of data requires the development of specific algorithms that start to be self-learning and which can sometimes reinforce an advantage of the first move. Some of these algorithms can be obtained from third parties, as happens with cloud computing for computing power, transforming a fixed cost into a variable cost and making this type of activity scalable. A specific problem concerns the value of data, both for consumers and for companies that collect them. From the few works done on this topic, the value for the consumer is often very contingent and depends on the trust in the counterpart, on the expectations regarding the future use of data and on the frame surrounding the decision.³² Many factors seem to indicate that the implicit evaluation is generally quite low, given the ease with which data are transferred.³³ At the same time, however, the transfer of data, as we have seen, occurs without the consumer being aware of subsequent uses and consequences to himself.

A recent experiment shows how the willingness to provide sensitive data is drastically reduced when the term privacy is made salient, even in a positive context.³⁴ This would seem to indicate that people are often inattentive and that hidden privacy concerns arise only when consumers are asked a specific question.

As regards personal data, consumers display considerable inertia and tolerance. Only 2-4% of customers terminate their contractual relationship after receiving a data breach notification. One explanation could

³¹ Anja Lambrecht and Catherine Tucker, "Can big data protect a firm from competition?", 2016, <https://ssrn.com/abstract=2705530>.

³² Francesco Feri, Caterina Giannetti and Nicola Jentzch, "Disclosure of personal information under risk of privacy shocks", *Journal of Economic Behavior & Organization* 123 (2016): 136-148.

³³ Simeon Schudy and Verena Utikal, "You must know about me – On the willingness to share personal data", *Journal of Economic Behaviour & Organization* 141 (2017): 1-13.

³⁴ Helia Marrienos *et al.*, "Now that you mention it'. A survey experiment in information, inattention and online privacy", *Journal of Behavior & Organization* 140 (2017): 1-17.

be that customers do not regard the resulting damage as great enough to change their behaviour. Another is that consumers numb considering the frequency and the number of breaches reported in the news. In a controlled laboratory experiment, Feri, Giannetti and Jentzsch show that breach notifications change the individual propensity to provide sensitive personal information. Notifications induce the sub group of individuals with personally sensitive information to reduce the disclosure of information with firms.³⁵

Moreover, there is no objective parameter to identify the value of the data. The reference should be the price that consumers accept to release their data, the price they would pay to protect them, the costs in which they expect to incur in case of incorrect use of the data, or the expected profits generated by use of their personal data.

A dominant position obtained through the control of personal data can facilitate some potential abuses: first-degree price discrimination, personalised advertising and the degradation of service quality. Naturally, in evaluating these possible behaviours, it is also necessary to take into account the advantages of efficiency and improvements of the products obtained with the use of big data.

In the first-degree price discrimination, sellers with monopoly power are able to know buyers' reserve price and, if they can prevent arbitrage, they are able to fix for each consumer a price equal to their reserve price. The economic effect is that the seller can appropriate the full consumer surplus generated by the transaction. The welfare effects of this price discrimination are ambiguous because, if sellers can prevent arbitrage, then they can serve all the demand willing to pay a price higher than the marginal cost, which is equal to the demand served in a competitive market. With a first-degree price discrimination, a monopolist produces the same quantity than in effect competition and therefore there is no allocative inefficiency. Yet, the entire consumer surplus is forfeited by the seller. This market outcome does not directly pose welfare or competitive problems, but certainly questions the distribution of value. On the other hand, first degree price discrimination implies an expansion of the market as consumers with low willingness to pay are served. Although in theory a deep knowledge of consumers and the lack of transparency of the market (consumers are unaware of the prices proposed to others) could allow the application of

³⁵ Feri, Gianetti, and Jentzsch, "Disclosure of personal information", 138-148.

personalised prices, in practice and for the moment this tool does not seem to be used much because it clashes with a strong annoyance of consumers. Experiments are often done in secrecy. When Amazon was discovered to offer different prices depending on consumers, it had to quickly reverse. Generally, simpler systems are used, such as time discrimination, in the case of airlines, or special pricing for large groups.

To overcome the opposition of consumers, personalised prices can be implemented with less recognisable design, such as a uniform price and personalised discounts, and it is possible to show different products to consumers of different groups depending on their willingness to pay. Naturally, complex pricing systems can loosen competition and raise consumer search costs.

Personalised advertising consists of showing a banner or a sponsored search result to a specific audience based on personal characteristics or interests.³⁶ From advertisers' perspective, the potential benefits are that the ad is shown to potentially interested customers and this reduces both the useless messages and the annoyance of the consumer and, secondly, that the ad content can be customised, potentially increasing the effectiveness of the advertising message³⁷. If personalised advertising is an improvement in efficiency, and if not all competitors are able to achieve it at the same level, then one can expect an increase in concentration and market power, with a simultaneous price increase.³⁸ The question of pricing, however, is a bit controversial because normally media sell targeted advertising at higher prices than traditional advertising, depending on how much the effectiveness of advertising increases. So, potential efficiency gains are usually captured by media instead of advertisers, who can anyway avoid useless advertising. From the point of view of consumers, the welfare effects can be different: on the one hand, they receive advertisements closer to their interests and therefore potentially more useful, but on the other hand seem to be annoyed because they feel under control. Often platforms do not allow advertisers very personal customisations because of this fear.

³⁶ Marco Gambaro and Riccardo Puglisi, "What do ads buy? Daily coverage of listed companies on the Italian press", *European Journal of Political Economy* 39 (2015): 41-57.

³⁷ Avi Goldfarb, "What is different about online advertising", *Review of Industrial Organization* 44, no. 2 (2014): 366-376.

³⁸ Burreau, Streel and Graef, "Big data and competition policy".

4. Definition of relevant market

The definition of a relevant market in the big data value chain presents several problems due to the complex vertical articulation of the activity and the numerous players involved, where each company plays several roles at the same time.

Data can have multiple roles in the economy of digital markets. They can be a product in itself, with a specific price, as happens in the intermediate market chain – operators and brokers exchange data to complete their collections – or they can be an input, which helps to define the characteristics and needs of customers or to reduce the risk, as is the case in many financial transactions.³⁹

Therefore, the definition of the relevant market can specifically focus only on the data when these are actually the main product that is exchanged. In this context, the intermediate market for data seems to operate reasonably competitively with many operators, although the issue of vertical integration, that is, operators operating in multiple layers of the value chain, can be relevant.

In digital markets the simple analysis of market shares does not say much about the consolidated market power of the main vendors because competition has a strong dynamic component and often takes place for the market rather than in the market. In the first decades of internet diffusion life cycles have been extremely short with established leaders who get knocked out within a few years. Potential competition often emerges from new entrants able to redefine business models with the addition of ingredients which the incumbent did not think about before, as in the case of Huber, or which simply redefine the production function with a different combination of inputs, as in the case of Google. Digital environment makes this flexibility easy both through the development of numerous intermediate markets for specific services, such as payment services, data centres, or specialised software, and through a general reduction in data processing and communication costs, which together have the effect of lowering economies of scale.

This flexibility and this vertical articulation mean that many companies perform multiple roles both vertically and horizontally at the same time, and that their strategic location evolves rapidly. A typical example is Amazon, which few would have thought ten years ago to be a competitor

³⁹ OECD, *Big Data: Bringing Competition Policy to the Digital Era*.

of Microsoft or Google, but which playing on the flexibility of its vertical integration entered into intermediate markets, such as cloud computing, becoming one of the leaders. Or Apple, which is simultaneously a multi-sided platform, a technology vendor, an IT service provider and a multi-media content operator.

In all these activities some scope economies frequently emerge which are sometimes company specific. The definition of the relevant market must therefore be made according to the specific case and it is not always possible to generalise. Many digital markets are characterised by a multisided structure with both informational and transactional platforms. Although the concept of two-sided market has now entered the standard equipment of the antitrust authorities, the analysis of cases concerning it are not always carried out fully considering the implications of this specific feature.⁴⁰

The key element of the two-sided platforms are the externalities that a transaction on one side of the market exerts on operators and transactions on other side. In the traditional example of the media market, a hypothetical increase in prices on the editorial side rises the revenues and lowers the number of readers, but at the same time this makes the platform less attractive for advertising investors, making the revenues' decrease bigger.

The identification of the relevant markets on different sides can be complicated when platforms supply free services in exchange for data that are mainly used in advertising. Equally complex is the definition when the different firms considered have a portfolio of different multi-sided activities that overlap only partially which achieve specific economies of scope and are characterised by different reciprocal externalities. Through the use of personal data, platforms can offer consumers a variety of personalised, subsidised and sometimes totally free services, which makes it difficult to apply the SSNIP (small but significant non-transitory increase in price) test, which remains one of the standard tools for defining the boundaries of the market.

The solution may be to evaluate the possible quality degradation (i.e. a reduction in production costs) that an operator with market power can implement through a modified test such as the SSQNDQ (small but significant non-transitory decrease in quality).

⁴⁰ David Evans, "The online advertising industry: Economics evolution and privacy", *Journal of Economics Perspectives* 23, no. 3 (2009): 404-450.

The problem is that the degradation of quality can also be implemented for other reasons, some of which can be sanctioned from the point of view of competition law.

Degradation can occur selectively on some groups of consumers to support price discrimination. To push users with high willingness to pay to take the most expensive version of a service, it is possible to degrade the quality of service for other users, which is relatively simple and inexpensive in the digital environment.

Or a firm can implement a degradation of service to customers who use the services of competitors in other sides of the market. For example, a search engine could eliminate the natural search results of content providers, with which it does not have a revenue sharing agreement, from the top positions, but this would be a case of exclusion that is well managed with traditional tools of the antitrust authority.

5. The space for regulation policies

Big data can be considered amplifiers of other competitive advantages and pose different problems in the relationship between citizens and firms and between citizens and public authorities. They can contribute to the creation of information asymmetries that can have negative consequences on the functioning of the society.

Given the multiform nature of the issue of big data, it is not so clear if new regulation systems are needed and which the appropriate contexts to build them are. In particular, it is important to consider whether competition law should be expanded to pursue public policy objectives concerning consumer protection, economic justice or fundamental rights such as privacy.

I think this is an unpromising path, because it would mean distorting both antitrust instruments and objectives, and because antitrust intervenes *ex post* on specific cases and this approach is not suitable to manage policy choices that can be more general in nature.

Many issues related to big data are better managed through consumer protection or Intellectual Property Rights regulation. But policies that are intended to protect consumers and privacy may have the side-effect that they make entry into a market more difficult. Such entry barriers can have a negative effect on dynamic competition, incumbents are not challenged by new disruptive business models as much as they would with be without those barriers. A new entrant can for example not buy personal data

on a data-brokering market as this is not allowed without user consent. Compliance with consumer and privacy protection legislation requires investments and can make it difficult for small companies to operate on a market and to compete effectively with companies with a larger scale. It is desirable that in the design of consumer and privacy protection rules the effects on competition and specifically entry barriers are considered to maximise dynamic efficiency.⁴¹

In May 2018, after two years of transition, the General Data Protection Regulation of the European Parliament became active. The EU GDPR is a wide ranging personal data protection regime of far greater magnitude than any similar previous regulation in the EU, or elsewhere. Fostering individual rights to access, rectify, transfer and request the deletion of personal data, the regulation essentially gives back data ownership to more than 500 million data subjects within the EU rather than leaving them in the hand of the companies that store and process users' data.

Both the sharing⁴² and the protecting of personal data can have positive and negative consequences at both the individual and societal levels. On the one hand, personal information has both private and commercial value, and the sharing of data may reduce frictions in the market and facilitate transactions. But on the other side the use of personal data may subject an individual to a variety of costly practices, including price discrimination in retail markets, quantity discrimination in insurance and credit markets, spam, and risk of identity theft, in addition to the disutility inherent in just not knowing who knows what or how they will use the information in the future.

Many of the transactions with digital operators are characterised by information asymmetries, and consumers are not always aware of the value of their personal data and of the use that will be made thereafter.

Moreover, in recent years, with electronic commerce, consumers' contractual position has steadily worsened. Usually consumers pay before receiving the product and bear on their shoulders a large part of the risk associated with any economic transaction. The wave of consumer protection regulations that happened in the 70s and 80s developed mandatory information disclosure, warranty rights and withdrawal rights, somewhat limiting the arbitrary bargaining power of companies. In traditional

⁴¹ Ecorys, *Big Data and Competition*.

⁴² Acquisti, Taylor and Wagman, "The economics of privacy", 442-492.

commerce, the consumer selects the product in the shop, tries it, takes possession of it and at that point pays it; sometimes, with consumer credit, they pay it in instalments later.

Advanced payment has become a standard practice in e-commerce and the consumer suffers the risks of non-delivery, defective products, non-compliant products, and all possible delays. The fulfilment of a purchase starts only when the payment is made. It is a very important shift in power that is not always adequately emphasised. With the massive collection of personal information, contractual balance can move further in favour of sellers, who can choose whether to discriminate prices, while the consumer unconsciously provides private information on preferences, or reserve prices, which will then be exploited to extract the entire generated surplus from the exchange. If firms have all the bargaining power, the only possibility for the consumer is to accept terms and conditions or not to use services that are often almost indispensable for social relations.

I believe that in the next few years a new wave of consumer protection regulations will be deployed which will tend to slightly rebalance information and power asymmetries between firms and consumers, taking into account the new digital landscape.

Bibliography

- Acquisti, Alessandro, Curtis Taylor and Liad Wagman. "The economics of privacy". *Journal of Economic Literature* 54, no. 2 (2016): 442-292.
- Balto, David and Matthew Lane. "Monopolizing water in a tsunami: Finding sensible antitrust rules for big data". *Working Paper* (2016). <https://ssrn.com/abstract=2753249> or <http://dx.doi.org/10.2139/ssrn.2753249>.
- Baxter, William. "Bank exchange of transactional paper: Legal and economic perspectives". *Journal of Law and Economics* 34, no. 2 (1983): 309-328.
- Burreau, Marc, Alexandre de Streel and Inge Graef. "Big data and competition policy: market power, personalized prices and advertising". *Project report, Centre on Regulation in Europe*. 2017.
- Colangelo, Giuseppe. "Big data, piattaforme digitali e antitrust", *Mercato Concorrenza e Regole* 18, no. 3 (2016): 425-460.
- Evans, David. "The online advertising industry: Economics evolution and privacy". *Journal of Economics Perspectives* 23, no. 3 (2009): 404-450.
- Evans, David and Richard Schmalensee. "The antitrust analysis of multisided platform business". *Coase Sandor Working Paper Series, University of Chicago Law School* (2012).

- Evans, David and Richard Schmalensee. "The industrial organization of markets with two sided platforms". *NBER Working Paper No. 11603* (2005).
- Feri, Francesco, Caterina Giannetti and Nicola Jentzsch. "Disclosure of personal information under risk of privacy shocks". *Journal of Economic Behavior & Organization* 123 (2016): 136-148.
- Gambaro, Marco. "Concorrenza e pluralismo nel mercato di Internet: La prospettiva economica". In *Le Libertà Fondamentali nell'Era di Internet*, edited by Marco Bassani, Oreste Pollicino and Tommaso Frosini. Milan: Mondadori, 2017.
- Gambaro, Marco. "Some economics of new media content production and consumption, and strategic implication for media companies". In *Handbook of Social Media Management*, edited by Mike Friedrichsen and Wolfgang Mühl-Benninghaus, Berlin: Springer, 2013.
- Gambaro, Marco and Riccardo Puglisi. "What do ads buy? Daily coverage of listed companies on the Italian press". *European Journal of Political Economy* 39 (2015).
- Gawell, Annabelle. "Big data: Bringing competition policy in the digital era". *Working Paper*, OECD (2016).
- Goldfarb, Avi. "What is different about online advertising". *Review of Industrial Organization* 44, no. 2 (2014): 366-376.
- Henten, Anders Hansen and Iwona Maria Windekilde. "Transaction costs and the sharing economy". *26th Conference of the International Telecommunication Society, San Lorenzo de El Escorial 24-27 June 2015* (2015).
- Horton, John and Richard Zeckhauser. "Owning, using and renting: some simple economics of the 'sharing economy'". *NBER Working Paper 22029* (2016).
- Huberty, Mark. "Awaiting the second big data revolution: From digital noise to value creation". *Journal of Industry Competition and Trade* 15 (2015): 35-47.
- Jarmin, Ron and Amy O'Hara. "Big data and the transformation of public policy analysis". *Journal of Policy Analysis and Management* 35, no. 3 (2015): 715-721.
- Kerber, Wolfgang. "Digital markets, data, and privacy: Competition law, consumer law, and data protection". *Joint Discussion Paper Series in Economics* 14 (2016): 639-647.
- Marrienos, Helia, Mirco Tonin, Michael Vlassopoulos and M. C. Schraefel. "Now that you mention it'. A survey experiment in information, inattention and online privacy". *Journal of Behaviour & Organization* 140 (2017): 1-17.
- Newman, Nathan. "Search, antitrust and the economics of the control of user data". *Yale Journal of Regulation* 31, no. 2 (2016): 401-454.
- O'Connor, Daniel. "Understanding online platform competition: Common misunderstandings". *Internet Competition and Regulation of Online Platforms, Competition Policy International* (2016), <https://ssrn.com/abstract=2760061> or <http://dx.doi.org/10.2139/ssrn.2760061>.

- OECD. *Big Data: Bringing Competition Policy to the Digital Era* (Paris: OECD, 2016).
- Pew Research. "The state of privacy in post Snowden America", Pew Research, 2016, <http://www.pewresearch.org/fact-tank/2016/09/21/the-state-of-privacy-in-america/>.
- Polo, Michele. "Consumer's search in the era of big data". *Competition Policy International*, 2017, <https://www.competitionpolicyinternational.com/wp-content/uploads/2017/05/CPI-Polo.pdf>.
- Rochet, Jean-Charles and Jean Tirole. "Platform competition in two sided markets". *Journal of the European Economic Association* 1, no. 4 (2003): 990-1029.
- Schudy, Simeon and Verena Utikal. "You must know about me – On the willingness to share personal data". *Journal of Economic Behaviour & Organization* 141 (2017): 1-13.
- Sokol, Daniel and Roisin Comereford. "Does antitrust have a role in regulating big data?". *George Mason Law Review* 23, no. 5 (2016): 1129-1161.
- Varian, Hal. "Big data: new tricks for econometrics". *Journal of Economic Perspectives* 28, no. 2 (2014): 3-28.
- Zerbas, Georgios, Davide Proserpio and John Byers. "The rise of the sharing economy: Estimating the impact of Airbnb on the hotel industry". *Journal of Marketing Research* 54, no. 5 (2017): 687-705.