CAN DIGITAL DATA BE REPLACED?
DATA SUBSTITUTABILITY IS THE KEY

BY MARIATERESA MAGGIOLINO[1] & GIULIA FERRARI[2]

1 Associate Professor, Department of Law, and Director of the Master of Arts in Law, Bocconi University.

2 PhD Candidate in Legal Studies, Bocconi University.

Visit **www.competitionpolicyinternational.com** for access to these articles and more!

I. INTRODUCTION

"Just the facts ma'am: competition cases are all about facts."[3] That sentence has the merit of highlighting how antitrust law, aiming at ensuring the proper functioning of the market, cannot disregard the knowledge of the many and varied empirical elements that influence economic agents' decisions and their effects on consumer welfare. This is the reason why, for example, antitrust authorities and judges: (i) investigate the way the demand is oriented in order to understand consumers' preferences and spending constraints; (ii) identify the business models used by undertakings to grasp the economic rationality underlying their strategies; and, finally, (iii) describe the structural and institutional characteristics of markets in order to establish the empirical elements that affect businesses and consumers' choices.

Therefore, there cannot be a "one-size-fits-all" answer to the question of whether data constitute an essential resource (i.e. facility) or a barrier to entry. This question cannot be answered without taking into consideration the specific circumstances that characterize the particular case under scrutiny.

What instead does not vary depending on the facts of the case at stake is the analytical framework that those in the antitrust field use to develop their reasoning. Thus, let's try to reconstruct it.

In essence, when we say that a resource used in a given market to produce the goods there sold works as a barrier sheltering that market, we mean that an undertaking that wants to provide those goods must incur the costs necessary to use that resource: the higher the costs, the higher the barrier. With the same degree of approximation, when we say that a resource is essential, we mean both that, without that resource, a company will not be able to create its products and/or services (obviously to sell them in a given market), and that the said resource does not have economically reproducible substitutes. From this point of view, therefore, it could also be argued that an essential resource is an insurmountable barrier to entry. But this is not the point.

The point is that, in both cases, the interpreter's main task will be to understand whether and to what extent the resource in question has viable substitutes, which could possibly be cheaper. In the event that those substitutes are deemed to exist, the resource could certainly not be considered to be essential; it could be considered, at most, as a barrier to entry, depending on the cost of the available substitutes.

By applying this reasoning in relation to data, the crucial issue is thus to understand if and to what extent digital data have substitutes. In other words, what matters the most in the context of this competitive assessment is the existence (or not) of other "elements" (and mainly other data) that can be considered to be adequate substitutes for the resource in question, i.e. digital data.

3 Ian Forrester, *A Bush in Need of Pruning: the Luxuriant Growth of Light Judicial Review* 410, Claus-Dieter Ehlermann, Mel Marquis, *European Competition Law Annual 2009: Evaluation of Evidence and Its Judicial Review in Competition Cases* (2011).

II. DATA SUBSTITUTABILITY

Reading about the digital economy, it is common to come across the observation that digital data are non-rivalrous resources. The term refers to the idea – which is certainly shared here – that the use by Undertaking A of certain data would not prevent Undertaking B from using the same data at the same time and equally benefitting from them, since, like other intangible assets, digital data do not lose value with use.[4]

In addition, the Fourth Industrial Revolution has chosen binary code as the language for the representation of the world and human behaviors. As a result, many objects have become sources of digital data: not only personal computers and the Internet but also devices – i.e. intelligent objects – in everyday use and of modest cost.[5] Since nothing leads to the belief that this tendency towards the "dataization" of reality is likely to be interrupted, it is easy to understand why the concept of ubiquity is often attributed to digital data.[6] In the context of the digital economy, that expression refers to the fact that potentially all companies are in a position not only to generate, but also to receive and retain huge volumes of data. It is thus easy to understand why many scholars argue that companies that are interested in exploiting these data only must face the fixed costs of acquiring the infrastructure and technical skills needed to generate and collect them.[7]

However, these general statements, while correct and acceptable, do not help competition law scholars and practitioners to understand whether, in a specific case, a certain set of data could be considered substitutable with others.

In competition law, indeed, the substitutability of any good, be it an input or a product (whether intermediate or final) must be assessed in relative – rather than absolute – terms. The assessment exercise must be accomplished by understanding the use that is made of that particular resource, which in turn can be identified based on the needs and desires that it satisfies. Therefore, in order to understand whether data – or, *rectius*, datasets – have substitutes or not, it is essential to understand what those data are intended for.

Suppose, for example, that a businesswoman active in the tourism sector of the Riviera Romagnola in Italy wants to know the eating habits of her potential customers, as this information could enable her to personalize the menus of the chain of hotels and guesthouses she manages. To this end, already equipped with the software and skills necessary to develop inferential analysis, she strives to obtain as many useful data as possible on the eating habits of those visiting Romagna so as to map in the most accurate way the target she wishes to please with her services. To do this she probably intends to merge those data with those that she already has, having presumably collected them during the years of activity of her hotels and/or obtained from her trade association.

In order to obtain such data, she can, for instance, turn to Google Search, since the American company certainly has traces of the food and wine searches made by users who, during the summer, connected to the Internet from Romagna. Similarly, she may seek to obtain the data that platforms such as TripAdvisor and The Fork certainly have collected when their users consulted them between May and September to find out about the restaurant and food services available in the Romagna provinces. Furthermore, the businesswoman can address insurance companies, companies offering navigation systems and/or car rentals in order to find out which restaurants and culinary establishments those who travelled the roads of Romagna in the summer visited.

In short, since there are several sources of data capable of revealing the information required (in our example the eating habits of those who visit the Romagna region), all the datasets above mentioned must be considered as belonging to the same market – i.e. the market of the data from which the eating habits of customers on the Riviera Romagnola are inferred – and, thus, all the firms offering them must consider themselves to be competitors for the supply of such data.

Many other similar examples can be made: let's think for instance of the reading, music and movie preferences of a particular group of

---

4 As evidence of this, it is noted that undertakings interested in using digital data may choose to acquire such resources from the market. Datasets made available on the market, either free of charge or for a fee, may simultaneously be the subject of several contracts and thus multiple uses by different undertakings without reducing their value, i.e. the possibility for undertakings "buying" such data to use them profitably. See Anja Lambrecht & Catherine Tucker, *Can Big Data Protect a Firm from Competition?* 4-5 (2015), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2705530, and Alessandro Acquisti & Hal Varian, *Conditioning Prices on Purchase History*, 24 Marketing Sci. 367 (2005).

5 Reference is made to the wide variety of "smart devices" (such as virtual assistants based on AI technology) and to the so-called Internet of Things (IoT).

6 See Catherine Tucker, *The Implications of improved attribution and measurability for antitrust and privacy in online advertising markets*, 20 *Geo. Mason L. Rev.* 1025, 1030 (2013) and Manuel Castells, *La nascita della società in rete* (2014).

7 Carl Shapiro & Hal R. Varian, *Information rules: A strategic guide to the network economy* 24 (1999).

users (such as, young people between 20 and 30 years old living in Europe) that some undertakings may want to know in order to develop successful "products" (such as, reading apps, music and streaming apps, but also songs and movies/tv shows specifically targeting the preferences of this segment of the public). This information may be available both to search engine providers by way of analyzing the search queries that these users have entered over a given period of time, while a social network provider may also be able to gain the same knowledge by looking at the profile information that users have shared on its platforms. In addition, the same information may be held, respectively, by platforms offering reading services (such as, to name a few, Blinkist, Instaread, or Amazon Kindle), music streaming platforms (such as Spotify or Apple Music), as well as streaming platforms (such as Netflix, Sky Go, or Amazon Prime Video).

Again by way of example, an undertaking wishing to develop a word recognition program based on artificial intelligence ("AI") technology would need to have digital versions of texts – if necessary translated into several languages – for its machines to read, so that it can design and test the algorithm which, over time, will learn to identify words, assign them meanings and translate them.[8] Even in this scenario, the undertaking may get access to such texts through a contract with libraries, but those texts can be also replaced by other texts held by the same or other libraries.

Therefore, depending on the particular information at issue, all the said companies/libraries can be considered to be competitors as they may all be in the position to "answer" the same needs and uses. In each of the examples above, indeed, the competitive analysis should aim at understanding if and how much the resources in question (data about the eating habits, reading/music/movie preferences or texts) can be replaced by other data held by different subjects.

Out of these examples, it must be noted that typically any information can be inferred by querying different and distinct datasets, not necessarily controlled by the same subject. Since digital data are nothing more than the representation in sequences of ones and zeroes of real-world facts and related human behaviors, and since, due to the Fourth Industrial Revolution, digital data can be obtained either by transforming analog data into digital data, or gathering the inputs given by users not only on the Internet or the IT structures of companies, but also offline through intelligent objects, information such as the eating habits of those who visit the shores of Romagna (or the reading/music/movie preferences of young European or texts) can be inferred from a number of various sources. Moreover, since under antitrust law, all goods responding to the same needs and satisfying the same desires are deemed to belong to the same market, then all these datasets must be deemed to belong to the same market.

Of course, going back to our examples, it could be argued that not all those datasets are perfectly equivalent to each other in terms of quantity, quality and accuracy. Nevertheless, in the everyday life of market analysts, it rarely happens that products and services belonging to the same market are *perfectly* interchangeable. A certain degree of heterogeneity is indeed inherent in the contemporary economy, especially in all the markets that are not oligopolistic.

Moreover, it is doubtful that the hotelier from Romagna or the company wishing to develop or ameliorate new products/services targeting young European or AI software would be refused almost all the requests made to the subjects owning the desired data. This possibility (i.e. the refusal to provide the data) – which often attracts the attention of commentators whose intent is to (uncritically) discuss the indispensable nature of "big data" – sets the way, though, to discuss another important question, i.e. the access to data. Discussing of data access, however, reinforces the importance - as discussed so far - of analyzing the factual elements that serve to identify the relevant markets.

In other words, stating that several datasets similarly capable of generating information belong to the same relevant market is a separate (and logically prior) matter to the question of whether access to some of those datasets is prevented by the companies that control them. Indeed, while the first consideration concerns the description of the markets, the second concerns the behavior of the undertakings. Specifically, the latter concerns a possible refusal to deal which the antitrust authorities would be able to pursue – whether successfully or not is still a different issue – if it resulted from a concerted action or, more likely, from a unilateral act by an undertaking holding a dominant position.

Finally, it could be argued that the decision to identify markets for data on the basis of the use that is made of them does not allow *per se* to achieve a result that some commentators hope to reach: namely having a perfect coincidence between the resource named "dataset α" and the relevant market. This coincidence is what in fact would allow antitrust law enforcement agencies to nimbly show the dominance of a given company on the market for the sale of data α.

---

8 This is for instance the case of the Google Books project: through a set of "digitalization agreements" entered into with a number of libraries starting from 2002 the US company was able to obtain a digitalized copy of the books included in their catalogues. By doing so Google was also able to achieve at least two different objectives: (i) refine its search and data mining activity within the digitized texts and (ii) develop further products and services based on that data and the information extracted from it, such as the translation service Google Translate.

In this regard, it is necessary to focus on a cornerstone of antitrust analysis which can be recalled as follows: it may happen that a resource constitutes its own market, i.e. that a resource does not have substitutes. Nevertheless, this possibility – which may arise more or less frequently – must be ascertained on a case-by-case basis, i.e. after having assessed and ruled out the existence of possible substitutes based on factual elements. In more sophisticated terms, the possibility that a resource constitutes a market in and of itself is an issue that may prove empirically grounded in an individual case, but conceptually is not always true.

It is not by chance that opposite scenarios are found in the decision-making practice.

For example, the European Commission chose to clear the *Facebook-WhatsApp* merger as the empirical analysis allowed the authority to verify that despite the merger between the datasets owned by the two companies, the resulting datasets continued to have substitutes on the market; in other words, the Commission ruled out that Facebook would have become the sole gatekeeper of users' digital data as a result of the acquisition of WhatsApp.

Indeed, the Commission clearly stated that, even after the transaction, "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"[9]. This supported the argument that, post-merger, the parties' competitors, including telephone companies and other digital platforms, were still able to access alternative sources of commercially useful Internet users data.

On the contrary, in the recent Italian *Enel* and *Acea* cases,[10] the Italian Antitrust Authority (Autorità Garante della Concorrenza e del Mercato) ascertained the non-substitutable nature of the lists of customers to which the two companies had exclusive access. The lists contained the name of the clients served by the two energy companies on the regulated market, who had given their (privacy) consent to be contacted for commercial purposes for the supply of liberalized energy services.[11]

Specifically, the Authority maintained that such lists of customers could not be replicated by other companies that, unlike Enel and Acea, were not vertically integrated and thus could not serve customers on the regulated market.[12] It consequently concluded that the exclusive control of these customers' data by Enel and Acea could produce an abusive form of anti-competitive foreclosure.[13]

III. CONCLUSIONS

While this article does not set out to *a priori* exclude that the datasets held by certain online platforms (such as the so-called GAFA[14] companies) could give rise to a competitive advantage for incumbents and an entry barrier for potential competitors due to the quantity, quality and accuracy of the data concerned, we want to highlight that an empirical analysis of the factual circumstances on a case-by-case basis should always be carried out to assess whether the same data (i.e. the data responding to the same needs) could be obtained elsewhere on the market.

---

9 European Commission, 3 October 2014, case COMP/M.7217, §§ 188-189. See also the case *TomTom/Tele Atlas* (European Commission, 14 May 2008, case COMP/M.4854) and the case *Google/DoubleClick* (European Commission, 11 March 2008, case COMP/M.4731) for other hypothesis in which the Commission denied the occurrence of an anticompetitive foreclosure.

10 Autorità Garante della Concorrenza e del Mercato (AGCM), Decision no. 27494 of 20 December 2018, case A511 – *ENEL - condotte anticoncorrenziali nel mercato della vendita di energia elettrica*, and Decision no. 27496 of 20 December 2018 - *A513 – ACEA - condotte anticoncorrenziali nel mercato della vendita di energia elettrica*. It should be noted that the paper focused on the *Enel* case as it is considered to be more explanatory of the critical competitive issues related to the collection and use of data, and because this decision has passed the Lazio Regional Administrative Court's scrutiny, which instead annulled the AGCM's Decision of the *Acea* case. On the other hand, the Regional Administrative Tribunal ("TAR") has imposed a recalculation of the fine inflicted on Enel on the basis of the parameters identified by the Administrative Judge in the ruling. See Tar Lazio, Judgment no. 11976/2019 of 17 October 2019 (*Acea* case) and Tar Lazio, Judgment No. 11958/2019 of 17 October 2019 (*Enel* case).

11 Pursuant to Law 124/2017 the Italian electricity market should have been fully liberalized as of July 1, 2020 when the regulated regime of "greater protection" (according to which domestic and small users can be provided with services at a regulated price) would have been totally replaced by the free market. However, in December 2019 the Italian Parliament passed an amendment according to which the full liberation will take place starting from January 2022.

12 Indeed, only the companies acting as distributors of electricity can provide the regulated energy services. See §§ 87 and 226 of the Decision in the *Enel* case.

13 More in detail, Enel has managed to acquire such data by offering their customer base the possibility to provide their consent to the processing of their personal data for commercial and marketing purposes separately to the companies of the Enel group and to third parties, although there was no obligation in this regard under the legislation on the protection of personal data. As a result of such setting, it was found that on average 70 percent of customers gave their consent to be contacted exclusively by companies of the Enel group, while only the remaining 30 percent also gave their consent to the processing of their data by third parties.

14 The acronym refers to the tech companies Google, Amazon, Facebook, and Apple.

This empirical analysis, indeed, is the first and essential step to establish – on the basis of solid grounds – whether such datasets in fact constitute an entry barrier and, if so, the "height" of any such barrier. At the same time, the analysis of the same circumstances will be crucial to identify whether the datasets can be deemed to amount to an essential resource. This is the case, of course, where it is ascertained that the data in question do not have any plausible substitutes. In this scenario, then, the task of the antitrust interpreters would be that of assessing the existence of the behavioral elements of the violation, i.e. a refusal to access the essential resource, should the other building blocks of the antitrust breach be deemed to arise – whether in the form of a concerted practice or an abuse of dominance.

To this end, data substitutability should always be assessed in relation to the kind of information that the subject(s) who seek(s) it want(s) to infer from those data. Lacking a "one-size-fits-all" answer to data substitutability, a correct analysis of whether or not certain data have substitutes can only take place after having thoughtfully established the very needs and uses that the "information seeker" wants to accomplish through such data. Assessing these preferences is what helps the antitrust interpreter to identify the kind of information an undertaking wants to infer from the data it seeks, and thus their potential substitutes.

However, there is a "but."

While data substitutability is key to establish whether data constitute a barrier to entry or an essential resource, another different data-related issue arises as a specific aspect of the Fourth Industrial Revolution. Big data companies do not always know in advance the kind of information they want to seek by analyzing data, but they will nevertheless find it very rewarding to analyze large volumes of different datasets to hunt for useful information. This process may indeed grant the undertakings with access to such massive amount of data (such as GAFA) an incomparable competitive advantage in that they may be able to find new business opportunities before and better than their competitors. As such, they may discern the opportunity to profitably expand to a collateral industry, and even be able to shorten learning times and thus understand not only in which markets it is worth investing, but also what to do to quickly develop a new business activity.[15]

In such circumstances, however, there is no need to assess whether these data constitute a barrier to entry or an essential resource and, as a consequence, assessing data substitutability cannot add much to the antitrust analysis, mainly because in such a case all data are equally important. In other words, one of the competitive dangers inherent in the Fourth Industrial Revolution relates to the possibility that many undertakings may not be able to form and use – for purposes *ex ante* unknown, but that only the analysis of such data will reveal – the knowledge of the world and of human behaviors that is already available to a few companies, and which grants them a precious competitive advantage.

That said, contemporary antitrust law can do little to face this competitive risk because it assumes by its very nature a position of partial economic equilibrium, as it proceeds "market by market," ascertaining the power that a company enjoys in a given market segment and verifying that the conduct of that company does not worsen, to the detriment of consumers, the balance achieved between supply and demand in that same market.[16] Indeed, because of the flexibility of digital firms, contemporary competitive analysis should instead focus on the economic power of firms and the effects of their behavior on several markets at the same time. At the moment, however, antitrust law interpreters have not yet succeeded in conceiving new categories of analysis that would be useful for this purpose. In order to overcome the problem of equal opportunities with regard to the knowledge of the world and human behaviors, a regulatory solution is seen so far as the most suitable, that is, the opening up of all data by whoever controls them, so that all companies can start from this common basis to develop products and services.[17]

---

15 This is the case of undertakings that have from the outset an ongoing, vast and varied set of updated data.

16 And not for other reasons, such as the fact that antitrust law does not protect the proper functioning of the market, instead of a fair distribution of wealth or other values of a more political nature, nor because it does not establish automatisms, such as that according to which an undertaking in a dominant position which infringes data or consumer protection rules should, for that very reason, be considered as abusing its dominant position within the meaning of 102 TFEU.

17 On this point, see Michal Gal & Daniel Rubinfeld, *Data Standardization*, *NYU Law and Economics Research Paper*, No. 19-17 (2019), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3326377, as well as the contribution published on the website of the European Data Portal (portal established under Directive 2003/98/EC on the re-use of public sector information), *AI and Open Data: a crucial combination*, 4 July 2018, available at the following link: https://www.europeandataportal.eu/en/highlights/ai-and-open-data-crucial-combination.

![CPI Competition Policy International logo]

CPI Subscriptions

CPI reaches more than 35,000 readers in over 150 countries every day. Our online library houses over 23,000 papers, articles and interviews.

Visit competitionpolicyinternational.com today to see our available plans and join CPI's global community of antitrust experts.