Giuseppe Colangelo* and Mariateresa Maggiolino**

Data access and AI: Antitrust vs. Regulation

The present contribution is submitted in response to the European Commission's call for contributions on "Shaping competition policy in the era of digitazion." We appreciate the opportunity to submit our work and commend the European Commission for its commitment to encourage reflections on the implications of digization for competition policy.

This contribution is based upon our recent publications on antitrust law, regulation, digital platforms, big data, and data protection. Namely, 'Data Accumulation and the Privacy-Antitrust Interface: Insights from the Facebook case', forthcoming in *International Data Privacy Law*; 'Fragile or Smart Consumers? Suggestions for the US from the EU', (2018) Stanford-Vienna TTLF Working Paper No. 36; 'Big Data as Misleading Facilities', (2017) 13 *European Competition Journal* 249; 'Data Protection in Attention Markets: Protecting Privacy Through Competition?', (2017) 8 *Journal of European Competition Law and Practice* 363. Further developments come from 'Data sharing and interoperability: fostering competition through APIs' (by G. Colangelo and O. Borgogno), (2018) mimeo; 'Open (Private) Data' (by A. Bertoni, M. Maggiolino, and M.L. Montagnani), forthcoming in P. Drahos, G. Ghidini, H. Ullrich (eds.), *Kritika: Essays on Intellectual Property*, Edward Elgar.

The contribution addresses the following issues: (*i*) Data access and Artificial Intelligence; (*ii*) Data access under competition law; (*iii*) The role of regulation; (*iv*) Personal data protection as a barrier to data sharing.

I. Data access and Artificial Intelligence (AI)

AI strategy requires a data strategy. AI environments are inherently dependent on data as an essential raw material, particularly with regards to deep learning. Since AI functioning is based on the identification of patterns in available datasets and the subsequent making of predictions and correlations able to solve technical problems, the presence of large amount of information to be processed is crucial to its functioning. In other words, provided that emerging technologies need a continuous access to streams of data from several sources, access to data and related data sharing practices are crucial factors to thrive innovation.

On the other hand, irrespective on any sophisticated discussion as to the existence of property rights on data, it is a fact that currently only some firms hold and control data portfolios that are sufficiently huge and diversified to be used to develop AI. Whether Internet platforms, IoT firms, or traditional businesses such as banks and insurance companies, these firms are the ones that, at the moment, can successfully exploit their

* Jean Monnet Professor of EU Innovation Policy; Associate Professor of Law and Economics, University of Basilicata; Adjunct Professor of Markets, Regulations and Law, and of Legal Issues in Marketing, LUISS Guido Carli; Adjunct Professor of Legal Issues in Marketing at Bocconi University; TTLF Fellow, Stanford University and University of Vienna; giuseppe.colangelo1975@gmail.com.

** Associate Professor of Business Law, Bocconi University; mariateresa.maggiolino@unibocconi.it .

first mover advantage in data collection and digital technologies to acquire an incomparable foothold in designing AI. Consequently, as the Commission suggests, the time is ripe to assess whether some legal instruments – ranging from competition law to regulation – could serve as a means of breaking away from this state of affairs and whether other legal rules, such as the provisions protecting personal data, prevent from granting all companies equal opportunities in the creation and improvement of the AI.

II. Data access under competition law

Competition policy makers have long been debating about the role of antitrust in facilitating data sharing in order to ensure a level playing field between undertakings. Access to data under competition law can be obtained only in exceptional circumstances, notably those referred to the essential facility doctrine (EFD).

The EFD belongs to the framework of refusal to deal and is based on the idea that a firm which is a monopolist has a duty to share its facilities with everyone asking for access, including competitors. As it provides for an exception to the general rule which states that firms, even monopolistic ones, are free to contract by choosing whether and with whom to make a deal, the EFD represents one of the most controversial antitrust issues. Indeed, the provision of a duty to share is likely to come along with counterincentives to invest due to the limited possibility of securing returns. The EFD, originally developed by US courts throughout the '80s and then gradually repudiated, has gained an increasing success in the EU as it represents the main antitrust instrument for addressing intellectual property issues from an antitrust perspective.

The case law of the CJEU has defined a framework of exceptional circumstances under which a refusal to deal might involve an anti-competitive conduct. According to the leading case *Magill* (Joint Cases C-241/91 P and 242/91 P), an undertaking holding an exclusive right may engage in an abusive conduct if the following conditions are met: (i) the input protected is indispensable due to the lack of actual or potential substitutes, (ii) the lack of an objective justification for a refusal to share, (iii) the possibility of the facility owner reserving for itself a secondary market through its conduct and (iv) the possibility of such a refusal preventing the appearance of a new product which the intellectual property right owner does not offer and for which there is a potential consumer demand. Further, in *Bronner* (Case C-7/97) the CJEU clarified that the first circumstance (i.e. indispensability) involves the existence of legal, technical or economic obstacles so serious that any duplication of the facility is nearly impossible or not viable. Subsequent case law has gradually dismantled both the secondary market (since in *IMS,* Case C-418/01, the CJEU considered the requirement to be met even if that market was just potential or hypothetical) and the new product requirements (since in *Microsoft*, Case T-201/04, it has been argued that this condition is fulfilled also by a follow-on innovation).

According to the European Commission ('The free flow of data and emerging issues of the European data economy', 2017), there is nothing to prevent competition authorities from applying the EFD in the context of data-driven markets. However, the exceptional circumstances test appears inherently ill-suited to tackle consistently competitive concerns involving data-driven markets. Indeed, as for the first condition, there is no agreement among scholars whether data may be considered as indispensable asset according to *Bronner*. While some contributions maintain that accessible data (i.e. open data and those who can be collected with the help of data brokers) should never be

considered indispensable, others stress that a vast array of obstacles may render impossible the replicability of specific datasets to new entrants. Namely, data generated by machines or processes stemming from IoT devices should be considered inaccessible to the user data of digital platforms. However, even with reference to this last scenario, if the data are input for the functioning of the AI or for obtaining information, even a dataset that cannot be replicated could find substitutes suitable for making the AI work or for the production of the desired information. In other words, it should not be considered essential.

Additional practical issues are raised by the third condition, namely the exclusion of effective competition in a secondary market requirement. This circumstance is met only when the undertaking holding the essential input is already marketing in the downstream market and, by denying access, forecloses that market to potential new entrants. Such a condition, however, is absent in many cases of refusal to share data.

Moving to the fourth requirement, i.e. the prevention to the appearance of a new product, its fulfilment in data contexts is not straightforward. Indeed, usually in data-driven markets firms do not know the products or service they are going to design by using those data before getting access to them.

Moreover, even if the EFD requirements were met, compulsory licences regarding data would be difficult to manage for several reasons, namely the scope of the duty to share in terms of subject matter (i.e. the identification of a well-defined set of data) and time horizon, the definition of terms and conditions for the licence, the compliance with data protection law.

More in general, setting aside the above-mentioned hurdles to apply EFD, it shall be considered that the antitrust toolbox scope is limited by its inherent case-by-case approach. Thus, regulatory interventions seem better-suited to tackle data-driven economy core issues. Since each industrial sector presents specific and dynamic needs which require to be duly addressed, regulation may readily be tailored on such peculiarities in order to accomplish coherent forms of data access.

III. The role of regulation

In recent years, the European Commission has started to tackle issues related to data access and sharing with a broad array of different legislative initiatives. While the General Data Protection Regulation (GDPR) introduced a general scope data portability right, the Second Payment Service Directive (PSD2) enshrined a sector-specific access to account data rule. Moreover, the Commission has tabled two proposals respectively aimed at removing obstacles to the free movement of non-personal data and at promoting the re-use of government data.

Namely, pursuant to Article 20 of the GDPR, each person has the right to have returned to them personal data they have provided to a company or organization on the basis of consent or contract and has the right to have that data transmitted without hindrance from one controller to another (even directly where technically feasible). Thus, owing to the data portability right, internet users are allowed to choose how to manage their data: they can transfer data between online providers; they are able to give their profiles, such as their past search history, to whoever will use them to offer value-added personalized

services; or they are capable of exerting influence over the trading and commercialization of their data.

Together with the GDPR, the European Commission has also targeted the free flow of non-personal data through a specific regulation proposal. Since "the ability to port data without hindrance is a key facilitator of user choice and effective competition," the proposal entrusts the European Commission with the task to "encourage and facilitate the development of self-regulatory codes of conduct at Union level, in order to define guidelines on best practices in facilitating the switching of providers and to ensure that they provide professional users with sufficiently detailed, clear and transparent information before a contract for data storage and processing is concluded."

Alongside with these general-purpose data portability rights, a sector-specific form of data portability has emerged in the field of payment services, that is the access to account rule enshrined in the PSD2. Pursuant to this new regulatory mechanism, account servicing payment service providers, such as banks, shall allow third parties to obtain real-time data relating to customers' accounts as well as provide access to such accounts by executing payment orders initiated through digital interfaces, on condition that customers give their explicit consent and that the account is accessible online. Furthermore, banks are under obligation to grant such access on a non-discriminatory basis both to payment initiation services and account information services.

Finally, given the potential of public and publicly funded data, the European Commission has decided to encourage data re-use and access of public sector information (PSI) through the review of the Directive 2003/98/EC. In its very essence, the proposed changes to the Directive aim at speeding up the transition of public sector bodies towards digitally-enabled functionalities and contributing to the creation of valuable ecosystem around data assets.

Despite different aims and scopes, all these regulatory interventions share a common reliance on application programming interfaces (APIs) as a crucial element for the flourishing of a common European data space. Indeed, the Commission ('Building a European Data Economy', 2017) has envisaged the adoption of a "broader use of open, standardized and well-documented APIs … through technical guidance, including identification and spreading of best practice for companies and public sector bodies." Moreover, the Commission ('Guidance on sharing private sector data in the European data economy', 2018) has launched an assessment process aimed at deciding how to better nudge undertakings to adopt "open, standardized and well-documented APIs."

However, any regulatory initiative is called to solve two main thorny issues. First, the effectiveness of data sharing regulatory interventions is linked to the technical implementation process. Second, if access to datasets has to be provided for, then it is equally necessary to establish appropriate compensation schemes able to strike a balance between the interests of data holders and access seekers. While, with regards to the former, a clear view as to who and how define APIs is still lacking, the latter has been addressed by relying on fair, reasonable and non-discriminatory (FRAND) terms as a possible way to set remuneration rules for the data accessed by third parties (European Commission, 'Building a European Data Economy', 2017; 'FinTech Action plan: For a more competitive and innovative European financial sector', 2018).

In this scenario, competition law enforcement might play a residual role by overseeing the transition towards a European data common space driven by the regulatory

intervention. Notably, competition authorities should fill the gaps which are likely to emerge from sector-specific frameworks as well as general-scope regulations. Since the implementation process of data sharing regimes is inherently complex and dwelling, all the authorities involved depending on the industry at stake are called to prevent subtle forms of anti-competitive practices which risk to frustrate the economic potential of data portability regimes (see, e.g., Bundeskartellamt, Case B 4 – 71/10, regarding certain rules of the Banking Industry Committee that prohibited the use by online banking customers of personal identification number and transaction authentication number to allow access to third party systems).

IV. Personal data protection as a barrier to data sharing

Personal data – that is, information relating to identified or identifiable natural people – represent a good share of firms' data booties. Therefore, any possible obligation imposing data access or data sharing as well as any private business strategy to open or pool data must be reconciled with the new EU rules set forth by the GDPR to govern data processing. Indeed, according to the Art. 4 of the GDPR, a firm processes personal data also when it transmits, disseminates, or makes them available to third parties or to the public in general.

However, the provisions of the GDPR have been designed to govern one-to-many relationships, i.e. cases in which one single company is in control of the personal data of several people. Differently, except for the cases where data transfer results from the individual choices of consumers as it happens with the data portability right of the GDPR or the access to account rule of the PSD2, the scenario of data sharing – whether imposed or chosen – refers to many-to-many relationships, i.e. cases in which many companies are in control of the personal data of many people. As a result, compliance with the rules of the GDPR risks being cumbersome, expensive, and problematic.

For example, consider that in order to have a fair, lawful, and transparent act of data sharing compliant with the purpose limitation principle, all the data subjects whose personal data are going to be shared should agree to the sharing, either from the outset or subsequently, but before the sharing takes place (see GDPR, Artt. 5(a), 5(b), 6(a), and 13(3)). Therefore, firms intending to open their datasets or to put them in common should specify this intention when they collect personal data, or clarify it later on, via other specific communications requesting permission. And this, unless data sharing resulted from a specific legal obligation, because any processing is deemed lawful when necessary for compliance with a legal obligation to which the controller is subject (see Art. 6 (c)).

Likewise, consider that among the big data held by firms there may also be sensitive data. As their sharing is strictly prohibited, no firm should be allowed to open those data or to put them in common with other firms, unless the sharing of sensitive data were deemed necessary for reasons of substantial public interest, on the basis of Union or Member State law which shall be proportionate to the aim pursued, respect the essence of the right to data protection and provide for suitable and specific measures to safeguard the fundamental rights and the interests of the data subject (see Artt. 9(1) and 9(2)(g)).

Overall, the data protection rules about individuals' consent do not permit firms to freely share their datasets for the sake of AI's development, unless such a strategy was imposed by the law. In other words, in order to be fair, lawful, transparent and not very

cumbersome and expensive, data sharing should be imposed by public authorities and not chosen by firms.

But the issues connected to data subjects' consent do not exhaust the problems that privacy compliance implies. For example, the GDPR requires that firms be precise in giving the information that data subjects need to freely express their independent and unambiguous consent. As a consequence, should the firms requesting permission to share personal data also indicate the purposes of data sharing? And, in the affirmative, how could they do it, if the many firms that will access the shared data will use them differently? Furthermore, the GDPR establishes that, where personal data have not been obtained from the data subject, the controller shall provide the data subject with several pieces of information, included the public interests pursued via data sharing (see Art. 14). Therefore, it is reasonable to expect that also firms receiving data be obliged to inform data subjects of the sharing and of the other information that Artt. 13 and 14 require, unless the data subject already has that information (see Art. 14(5)(a)).

In addition, according to the GDPR, firms should warrant that data sharing happen in a manner that ensures appropriate security of the personal data, including protection against unauthorized or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organizational measures (see Art. 5(e)). However, data integrity and confidentiality could be seriously undermined in a context where data sharing could involve thousands and thousands of firms with their own information systems (and information bugs).

Finally, and even more importantly, it is hard to understand how data subjects could exercise their rights – such as the right of access, the right to rectification, the right to erasure, the right to withdraw consent, the right to object, or the right to explanation – when many firms hold and exchange their personal data. For instance, to exercise their rights effectively and simultaneously, data subjects should be enabled not only to track the movements of their personal data, but also to "talk" to all (!) the controllers involved in the sharing.

To be sure, as the same GDPR suggests in connection to the act of archiving for public interest (see, e.g., Artt. 5(b), 5(e), and 89), anonymization and pseudonymisation could serve to work around these issues that may prevent data sharing from happening or make it cumbersome, expensive, and problematic. Indeed, notwithstanding the many doubts on the efficacy of anonymization and pseudonymisation, the GDPR relies on them to ensure that stored confidential data are not accessed inappropriately. After all, AI may need a huge and diversified amount of data, but it does not necessarily want to understand who individuals are and what they do: not necessarily AI needs data that serve to identify single persons or that make them identifiable. AI can work also with anonymized and psedunymized data. To be sure, this solution would not rule out the possibility of some perverse usages of these encrypted data. Though, this could be a good compromise.

V. Concluding remarks

Guaranteeing equal opportunities for companies in the development of AI is a laudable goal of industrial policy. However, two sets of issues come with it, even when all the available data have been anonimyzed or pseudoanonimyzed to overcome any data protection problem.

First, policy makers need to indicate the tools to be used to pursue this goal. In this regard, antitrust law and its EFD doctrine are ill equipped. Not only, rarely apply the liability conditions of the EFD. Also, any antitrust action regards specific cases that, by definition, would not solve the general problem of making all the available data freely usable by whoever could be interested in developing AI. On the other hand, regulation can tackle this issue directly, because it offers universal and general solutions, which may address whole industries, sectors, and markets. For example, there are sector-specific rules that impose sharing obligations on the results of tests regarding some chemicals (Regulation (EC) No 1907/2006) or in favor of the producers of spare parts for automobiles (Regulation (EC) No 715/2007, as amended by Regulation (EU) No 459/2012). Likewise, in 2016, the French government has opted for a more general provision that cuts across industries and obliges companies holding data of public interest, such as data on tenders, real estate sales, and consumption of gas and electricity, to make them freely usable and reusable by third parties (Loi n° 2016-1321 pour une République numérique).

The problem with regulatory solutions imposing sharing leads to the second issue that policy maker should consider, that is: the costs of any measure meant to realize data sharing in a context where private firms hold those data. Not only, such top-down decisions may discourage investments in data generation and collection. Also, they can undermine the principles and ideals underpinning free market economies: the need to guarantee equal opportunities to firms that intends to develop AI should not be satisfied at the expenses of other entrepreneurs that legitimately have previously invested in data gathering. This would ultimately mean thinking about digital data as public resources, freely appropriated by anyone regardless of who generated and created them.

As a result, it seems reasonable to proceed in another way.

While measures for free circulation of public sector data and for supporting private companies to opt for data sharing should be promoted, data sharing obligations should be applied cautiously sector-by-sector and only in relation to certain categories of data (see the recent Australian Government's proposal for a new Consumer Data Right). Since types of data may vary between sectors, there should be an industry data-specification process that enables the relevant industry to agree on the types of data that will be covered, as well as mechanisms for transfer and security protocols. Moreover, solutions for the standardization of both APIs and contractual clauses to share data and manage personal data would be crucial.