

Big Data as an essential facility: a perspective on big data as a crucial infrastructure in the data-driven economy

According to “The Economist”, with the emergence of a data-driven economy, «Conventional antitrust thinking is being disrupted from within»¹ and the inability to regulate such emerging market might have an impact on the dynamics of value creation within digital capitalism.²

As a matter of fact, in an environment dominated by few gatekeepers, antitrust theory seems not willing to consider big data as an input comparable to what was oil in the last century,³⁻⁴ therefore risking not to set the appropriate boundaries that are intended to guarantee fair access conditions to new players. Considering the importance that information assumes today, data collection and exploitation appear indeed as a new essential input for the entire data-driven industry. Such a development has taken place though without the Antitrust Authorities presiding over any data concentration processes (data concerning preferences, wishes, interests and habits of billions of people), thus assuming that the asset at stake would have not become key for any company wishing to profitably perform its own activities. Companies can indeed exploit big data analytics in order to develop new products or services, rethink their business processes and better manage their own supply chain, starting right from the analysis of their clients’ needs via big data techniques.

For this competition-related reason, but also in the light of the privacy-related considerations borne by big data in relation to individuals and their fundamental rights⁵, it would be desirable to define a data governance scheme able to take into consideration – at

¹ “A New School in Chicago,” *The Economist - Special Report - Fixing the Internet*, 2018.

² Viktor Mayer-Schönberger and Thomas Ramge, “Reinventing Capitalism in the Age of Big Data,” *Basic Books*, 2018.

³ “The World’s Most Valuable Resource Economist,” *The Economist*, 2017. : «A new commodity spawns a lucrative, fast-growing industry, prompting antitrust regulators to step in to restrain those who control its flow. A century ago, the resource in question was oil. Now similar concerns are being raised by the giants that deal in data, the oil of the digital era».

⁴ “Fuel of the Future,” *The Economist*, 2017. : «Data are to this century what oil was to the last one: a driver of growth and change. Flows of data have created new infrastructure, new businesses, new monopolies, new politics and—crucially—new economics. Digital information is unlike any previous resource; it is extracted, refined, valued, bought and sold in different ways. It changes the rules for markets and it demands new approaches from regulators. Many a battle will be fought over who should own, and benefit from, data».

⁵ Not only the right to a lawful treatment of their personal data, but also the right to information, impacted in the recent Cambridge Analytica case.

the same time – the need for a data market regulation and the protection of data referred to natural persons.

One of the key arguments of those who believe that the competitive dynamics in the big data market, according to the neoclassical view, should be left free to thrive (here is the clash between Hamiltonians and Jeffersonians),⁶⁻⁷ is focused on the economic nature of data as an input. Such docents question indeed the comparability between oil and data as essential inputs: while the first could not be considered as a public good because of its features (rivalry and excludability from consumption), the second one should be instead. This is why, still according to the neo-liberal theory, the data market would not need any kind of intervention. Nevertheless, although the simultaneous consumption of the same bunch of data by different subjects would not hinder its exploitation by others (non-rivalry in consumption), data cannot be considered as easily duplicable and therefore replicable by every competitor. In this regard, if it is true that some data are available in abundance and without any restrictions, others are inaccessible instead, unless subject to the payment of a fee (excludability from consumption).

From a neoliberal perspective, it would be possible to assume –because of the nature of data – that data-driven markets are characterized by the absence of entry barriers, due to the above-mentioned data features, including the negligibility of the marginal costs associated with their reproduction, which appears as a typical feature of information economy goods. However, this would be equivalent to affirm that web giants have not taken any advantage of their position, consolidated through collecting and analyzing data provided by the users of their services.

Such data appear instead as not replicable by a hypothetical newcomer: they are indeed not ubiquitous, nor they are easily available to each player on the market. Smaller competitors – due to their weaker position– could be therefore easily excluded from the market by dominant players because of their inability to access the same data collected by them.

This implies a difficulty for any market player to challenge the GAFAs position in the market: the size of the “web giants”, combined with the out-turns of network and snowball effects⁸, make it practically impossible for any companies to try and reduce their market power in

⁶ “A New School in Chicago.”

⁷ Lina M. Khan, “Amazon’s Antitrust Paradox,” *Yale Law Journal* 126, no. 3 (2017): 564–907, <https://www.yalelawjournal.org/note/amazons-antitrust-paradox>.

⁸ For more info, see: <https://www.oxera.com/agenda/snowball-effects-competition-in-markets-with-network-externalities/>

the data-driven ecosystem. To this end, any competitor should indeed be in a position to access the same data collected by big internet companies and to create a platform capable to provide the same services so as to attract the relevant data volumes managed by market leaders (and provide innovative services for free, which helps trigger a “feedback loop” feeding this mechanism). However, on the one hand, the mentioned feedback loop allows dominant firms to attract more and more customers, thus improving over time the quality of their services and introducing brand new ones; on the other hand, network effects sharpen in the data- driven industry, which results into entry barriers.

Still on the accessibility of data in the big data market, imagining for a moment that big players wanted to sell their data in an unregulated market, it would be difficult to figure out how the creation of a “data market” could allow the identification of an univocal value for data (i.e. structured/unstructured), when this differs in function of the specific contexts and in relation to the specific applications. Furthermore, it would be equally unreasonable to think that monopolizing the data market should be permitted because dominant firms are offering their products for free (email accounts and cloud computing services), without tangible harm to the individuals, i.e. no reduction in consumer’s surplus (as the dominant antitrust orientations affirm): digital platforms are indeed more efficient as they grow in size as well as in terms of variety of traditional markets that they serve, correspondingly increasing the amount of data collected and thus progressively and constantly improving the effectiveness of the algorithms through which they provide their services.

Both the orthodox antitrust discipline and individuals do not to sufficiently consider the hidden price paid by users in terms of privacy loss, disregarding how by accepting such services for free, individuals provide data to internet giants that will use them as an exchange good in the wide digital context.

Besides not understanding when or how their data are gathered to be sold to third parties⁹ (which has to do with data acquisition as a free individual’s choice), users are subjected to a wider lock-in phenomenon, stemming from the impossibility, for an individual not willing to disclose his own data, to access essential services provided exclusively by GAFAs. Due to economies of scale and scope, such services are indeed offered only by a few dominant gatekeepers, in markets featured by high switching costs that, in turn, can be explained by a low propension of users to provide their personal data to new entrants and to insufficient incentives to this end.

⁹ Information asymmetry and context-dependence problems.

High switching costs, together with economies of scale and scope, lower the average production costs borne by multisided platforms, allowing them to offer an ever greater range of services via the same production input: as productive factors, data show indeed no rivalry in consumption and can therefore be reused to amplify the positive feedback loop both on the demand and on the supply side; this progressively widens data collection activities, allowing, in turns, the supply of an ever-increasing set of services at decreasing costs.

Such a scenario might lead dominant players to leverage their position from data markets to other markets, through product improvements made possible by elaborations on their own data (think about the relationship between the maps market and that of driver-less cars) and data ownership might therefore result in big players conquering adjacent and non-adjacent markets.

Moreover, it would be important to assess the hidden costs of services offered for free by GAFA i) in order to effectively evaluate the consumer's welfare loss and the consequential consumer harm and ii) to better understand the nature of new multi-sided platforms where a single subject (the platform) acts as the intermediary between different markets and different players in the same market, somehow turning the platform itself into the "relevant market"¹⁰: if we analyze the relevant market in which a conglomerate platform such as Amazon operates, it appears to coincide with the platform itself. If we analyze the services provided by Google or Amazon, it appears indeed that nothing can be considered as a perfect substitute to their services. Consequently, looking at the relevant market in which Amazon operates, it ends up with being the platform itself, as no other existing marketplaces can be considered as perfect substitutes.

This is why, starting from the premise that data are a not an accessible essential input (at least in quantities that can be really functional to a firm in order to compete in the data market) nor one replicable by competitors, a preliminary conclusion could be drawn, on the background of the experience of the liberalization of the European telecommunications

¹⁰ Definition of relevant market: The relevant market combines the product market and the geographic market, defined as follows:

- a relevant product market comprises all those products and/or services which are regarded as interchangeable or substitutable by the consumer by reason of the products' characteristics, their prices and their intended use;
- a relevant geographic market comprises the area in which the firms concerned are involved in the supply of products or services and in which the conditions of competition are sufficiently homogeneous.

markets, and data could be considered to somehow have risen to the role occupied in the '80 and '90 by not replicable physical networks. In addition, turning to consider also intangible assets, it should be recalled that the European Court of Justice has ruled to open up to competitors any proprietary database, as long as it is deemed to be essential with a view to the competitive development of innovative services in the related downstream markets.¹¹ Something similar has been proposed by Luigi Zingales and Guy Rolnik on the pages of The New York Times. The two Authors suggest the reallocation of property rights via legislation in order to provide more incentives to compete. And, as claimed by them, the idea is not new:

«Patent law, for example, attributes the right to an invention to the company a scientist works for, to motivate companies to invest in research and development. Similarly, in the mobile industry, most countries have established that a cellphone number belongs to a customer, not the mobile phone provider. This redefinition of property rights (in jargon called “number portability”) makes it easier to switch carriers, fostering competition by other carriers and reducing prices for consumers.

The same is possible in the social network space. It is sufficient to reassign to each customer the ownership of all the digital connections that she creates—what is known as a “social graph.” If we owned our own social graph, we could sign into a Facebook competitor—call it MyBook—and, through that network, instantly reroute all our Facebook friends’ messages to MyBook, as we reroute a phone call.

*If I can reach my Facebook friends through a different social network and vice versa, I am more likely to try new social networks. Knowing they can attract existing Facebook customers, new social networks will emerge, restoring the benefit of competition».*¹²

A similar approach would allow to re-establish information as a public good. Within data market, information has been indeed transformed into a private asset able to generate

¹¹ For more information, see: <https://www.twobirds.com/en/news/articles/2004/ecj-ruling-ims-health>

¹² Luigi Zingale and Guy Rolnik, “A Way to Own Your Social-Media Data,” *The New York Times*, 2017, <https://www.nytimes.com/2017/06/30/opinion/social-data-google-facebook-europe.html>.

strong dependent relationships between platforms and users, capturing these latter in an informational aftermarket¹³.

¹³ Antonio Nicita, "I Big Data, La Privacy e La Prova Concorrenza," *Il Sole 24 Ore*, n.d., <http://www.ilsole24ore.com/art/commenti-e-idee/2018-05-14/i-big-data-privacy-e-prova-concorrenza-224307.shtml?uuid=AEaCFIoE>.