

de Pedraza, Pablo; Vollbracht, Ian

Working Paper

The Semicircular Flow of the Data Economy and the Data Sharing Laffer curve

GLO Discussion Paper, No. 515

Provided in Cooperation with:

Global Labor Organization (GLO)

Suggested Citation: de Pedraza, Pablo; Vollbracht, Ian (2020) : The Semicircular Flow of the Data Economy and the Data Sharing Laffer curve, GLO Discussion Paper, No. 515, Global Labor Organization (GLO), Essen

This Version is available at:

<https://hdl.handle.net/10419/215779>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

The Semicircular Flow of the Data Economy and the Data Sharing Laffer curve

Pablo de Pedraza and Ian Vollbracht¹

Pablo de Pedraza

Corresponding author
European Commission
DG Joint Research Centre
Unit I.1 - Monitoring, Indicators and Impact
Evaluation.
TP 361 – Via E. Fermi 2749 – I-21027 –
Ispra (Va) - ITALY
Tel: +39 0332 78 3805
pablodepedraza@gmail.com
pablo.depandraza@ec.europa.eu

Ian Vollbracht

European Commission
DG Joint Research Centre
Directorate I - Competences
Unit I.1 - Monitoring, Indicators and Impact
Evaluation.
TP 361 – Via E. Fermi 2749 – I-21027 –
Ispra (Va) - ITALY
Tel: +39 0332 78 3805
ian.vollbracht@ec.europa.eu

Abstract

This paper presents a theoretical conceptualization of the data economy that motivates more access to data for scientific research. It defines the semicircular flow of the data economy as analogous to the traditional circular flow of the economy. Knowledge extraction from large, inter-connected data sets displays natural monopoly characteristics, which favours the emergence of oligopolistic data holders that generate and disclose the amount of knowledge that maximizes their profit. If monopoly theory holds, this level of knowledge is below the socially desirable amount because data holders have incentives to maintain their market power. The analogy is further developed to include data leakages, data sharing policies, merit and demerit knowledge, and knowledge injections. It draws a data sharing Laffer curve that defines optimal data sharing as the point where the production of merit knowledge is maximized. The theoretical framework seems to describe many features of the data-intensive economy of today, in which large-scale data holders specialize in extraction of knowledge from the data they hold. Conclusions support the use of policies to enhance data sharing and, or, enhanced user-centric data property rights to facilitate data flows in a manner that would increase merit knowledge generation up to the socially desirable amount.

Keywords: Big data, Artificial Intelligence, Government.

JEL: H1, O38, P48

¹ The scientific output expressed does not imply a European Commission policy position. Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use that might be made of this publication. The authors would like to acknowledge comments and suggestions from Bertin Martens, Daniel Vertesy, Michaela Saisana, Sven Langedijk, and Ignacio Sanchez and previous work and discussions conducted at Webdatanet Task Force 25 during the cost action webdatanet (IS-1004, <https://webdatanet.usal.es/>) and at a series of seminars held at the European Commission Joint Research Centre in Ispra, Italy. The authors have no affiliations with or involvement in any organization or entity with any financial interest in the subject matter or materials discussed in this manuscript.

1.- Introduction

On the one hand, knowledge, as a particular form of information plays a fundamental role in the market economy and in defining the appropriate role for governments (Stiglitz 2001). Knowledge is a key component in productivity and growth (Romer 1986), the one ring of globalization that rules trade, capital flows and immigration (Freeman 2013). In addition, *inclusive* institutions spread economic benefits more widely and foster innovation, technology and long-run economic growth (Acemoglu and Robinson 2012, Galenson 2017). On the other hand, the traditional price-quantity approach has limitations in capturing the data intensive economy picture (Khan 2017) where access to Big data (BD) and Artificial Intelligence (AI) determine knowledge creation, welfare (Duch-Brown, 2017a, 2017b, and 2017c), innovation, wealth and power distribution (OECD 2019, ITU 2018).

Against that background this paper presents the semicircular flow of the economy, a theoretical framework that augments the traditional view of the economy for goods and services by adding data and knowledge flows to the traditional circular flow model (Samuelsson 1948, Samuelsson and Nordhaus 2010). This paper uses the semicircular flow of the economy to derive implications for digital governance and data policies (European Commission, 2018a) for social good (European Commission 2020).

Our theoretical approximation of a data intensive economy has the following characteristics.

Households' and firms' daily activity generate BD that data holders collect to produce knowledge using AI and disclose it in the form of services. AI is a scaled-up automated application of existing statistical techniques that enables recognition of patterns, regularities, and structures in data without an a priori theoretical framework (Boisot and Canals, 2004; Duch-Brown et al., 2017; Vigo, 2013). We take AI to mean a very broad definition of machine learning and related methods that can be used to analyze BD in order to generate knowledge. BD are data characterised by their volume, velocity and variety (Laney, 2001, 2012). Massive data points can be collected, organized, combined, searched and used for a wide variety of analysis purposes. AI models can be tested and continuously improved with new BD. Algorithms trained on one data set can be transposed to other complementary data sets and adjacent data (Duch-Brown et al., 2017) to obtain more and better predictions.

Knowledge production displays natural monopoly characteristics: economies of scale and scope; high fixed costs; and low variable costs. This tends to lead to the emergence of oligopolistic “giant” data holders that innovate and generate knowledge and services. Their innovation capacity and market power are likely to attract further investment, which fosters their data collection capacity and their further expansion. As in other sectors where market

power is concentrated among a small number of firms, there are incentives for oligopolistic data holders to collude and set barriers to entry in access to data, knowledge production and knowledge disclosure in order to protect and foster their market position.

In this market structure for BD acquisition and AI innovation, the amount of knowledge produced and disclosed in the economy is the amount that maximizes oligopolistic data holders' profit ($Q=Q_m$). By analogy with monopoly/ oligopoly theory, this amount is below the socially desirable amount, Q^* ($Q^*>Q_m$). Monetary taxes and fines do not necessarily solve anti-trust concerns because they do not address the fundamental causes of the market structure for BD acquisition and AI innovation and its consequences for knowledge underproduction, asymmetries and inequalities.

The semicircular flow of the economy is an analogy of the circular model of the economy for goods and services. In this paper, the same intervention rationale for government in that context is applied to data and knowledge flows. The government's goal in the data economy is to increase the amount of knowledge produced and disclosed in the economy up to the socially optimal amount. Data transfers (in what follows we use the terms "data transfers" and "data leakages" interchangeably, with both being analogous to the use of the term "leakage" in the traditional circular economy framework) towards merit users of BD enable further knowledge generation towards this socially desirable level. *Data-sharing policies* (analogous to a certain extent to *taxation* in the traditional model) are the data policy that generates data flows to promote merit uses of data while discouraging demerit ones. The *Data-sharing Laffer curve* is the theoretical relationship between the data sharing rate and the amount of knowledge generated in the economy. It is also, to a certain extent, a merit knowledge production possibilities frontier. Just as monetary taxation is hypothesized to have a level of maximum tax revenue, analogously, data-sharing policies have a point where the economy generates the maximum amount of merit knowledge possible.

The semicircular flow theoretical framework seems to describe a number of elements of today's BD and AI sector and may therefore provide a useful framework to simplify the complexity of the digital economy for public policy analysis. The semicircular flow approach includes, among other aspects, data as a means of payment, consumers' utility maximization, data holders' behavior and concentration process, and economies of scale and scope in knowledge production. The approach also suggests that the current production of knowledge is $Q=Q_m$. We explore channels by which data sharing policies can generate data flows to increase knowledge generation towards the socially desirable amount Q^* .

We therefore present a simple theoretical framework that is consistent with the existing literature conceptualizing digital economy governance. It is also in line with existing models showing that the amount of knowledge disclosed in online markets tends towards monopoly levels (Board and Lu 2018). Conclusions for policy action support several existing proposals: The World Economic Forum multiple stakeholders approach, data taxation (Askita 2015), the establishment of a data authority (Scott Morton 2019, Martens 2017), and further implementation and development of data portability rights (De Hert 2018, European Union 2016) by means of Personal Data Stores (PDSs) (Bolychevsky and Worthington 2018). In the light of our model, monetary taxation approaches (Pratley, 2018; Sandle, 2018; D'Andria 2019) are considered necessary but insufficient as they do not take into account the data and knowledge dimensions of the economy. Other approaches, such as Macron's agreement with Facebook (Scott and Young, 2018, NSF 2017) do tackle data and knowledge issues raised in this paper, but are not, in isolation, of sufficient scale and scope to have a major impact on overall merit knowledge generation. Conclusions also support United Nations' call for a global partnership to improve quality statistics available to citizens and governments to reduce gaps between the private and public sectors (UN, 2013; UN, 2014).

The remainder of the paper is organized as follows. Section 2 conceptualizes a data intensive economy. Section 3 explores how the world fits the semicircular model conceptualization. Section 4 identifies economic principles behind socially desirable knowledge production and optimum data sharing defining a data sharing Laffer curve. Section 5 identifies the implications of the amount of knowledge generated and the existing policy reactions aiming to move the economy along the data sharing Laffer curve. Section 6 concludes.

2.- A Data intensive Economy

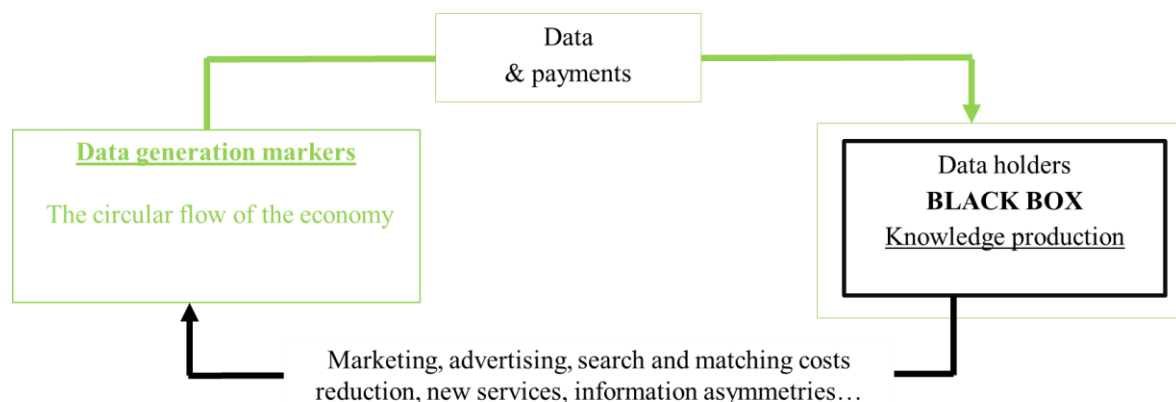
The semicircular flow of the economy is a theoretical model that simplifies the reality of a data intensive economy (Pedraza and Vollbracht 2019). Figure 1 represents an economy with the following characteristics. On the left-hand side, households and firms operate according to the circular flow of the economy model, exchanging goods and services for money and labour for wages (Samuelsson 1948, Samuelsson and Nordhaus 2010). Their activity generates a flow of data and monetary payments to data holders. On the right-hand side, data holders use Big Data (BD) and Artificial Intelligence (AI) techniques to extract knowledge and information from data. Knowledge production generates new and innovative services that influences the left hand side markets through matching efficiency, marketing, advertising, and reduction of search and transaction costs. Data flows are semi-circular: from households and firms to data holders but not in the other direction. Household and firms receive data-driven services created by

data holders based in part on their own data, but do not receive unprocessed data. This is the fundamental distinction with the traditional circular flows model in which workers supply labour to firms for an explicit money wage which they then spend on goods and services for explicit prices in the wider economy. The additions to the traditional circular flow of the economy are therefore the prominent role of data flows, data holders, knowledge production and knowledge flows.

The semicircular flow of the economy does not include explicitly data generated by firms and citizens in their interaction with the public sector. This is the case of social security records, administrative data, or medical records of public health services. Data holders have a competitive advantage in using AI technologies to obtain knowledge extraction technologies. They very often successfully compete in public competitions and have access to public sector data (Lomas 2019).

Data are a means of payment from firms and families to data holders. While in the circular flow of the economy explicit prices are a fundamental variable, data, as a means of payment, are ambiguous. While money is easy to use and understand, data are not. Data are not easily priced, their value is not clear, especially at the individual level. Data flows do not generate clearly comparable signals like prices. No authority is in charge of setting data value such as central banks that set interest rates in order to regulate the supply of money to the wider economy. We need to think beyond the traditional prices and quantities space. A space where prices are paid with data and quantities refer to knowledge in the form of digital services cannot be drawn in such a simple manner, but economic principles nonetheless apply to it.

Figure 1.- Semicircular flow of the data economy



It is assumed that in the semicircular data economy consumers maximize their utility considering only the monetary part of the prices that prevail, while assuming the value of their personal data equals zero. In this way, if the price of a digital service only includes payment with personal data, consumers consider it to be a “free” service. We refer to this as the “individuals’ data zero value assumption”. It can be motivated from two perspectives. First, the economic value of individual data, before merging with other data from other individuals, is indeed typically close to zero. Although individuals are owners of their personal data, they only capture its value after it has been merged and processed, i.e. through the digital services provided by data holders. Even if there is a clear legal corpus assigning individuals the property of their personal data, they have no capacity to process them via BD and AI methods. Second, in the early stages of digitalization consumers very often do not realise that they are generating data. From a microeconomic point of view consumers’ utility maximization is a function of monetary prices (PM), individual data value (IDP), and quantity (Q). $IDP=0$ if data holders’ activities within a black box do not undermine rights such, as privacy, nor have any negative impact in the functioning of competitive markets and the rule of law. Individuals’ data zero value assumption fails however, and therefore $IDP>0$, if individuals are aware that payment using data bears a cost in terms of individual rights, such as privacy, or via foregone income as a result of providing data for free which in fact has a positive market value.

Data holders are profit maximization agents. They obtain “de facto” ownership of data, build huge valuable BD sets and draw value from them by knowledge extraction using AI. In knowledge production, efficiencies arise from volume (scale) and variety (scope) and involve lower average cost the bigger the data set. This makes massive data sets very valuable even if $IDP=0$. In statistical terms, scale refers to the number of observations (N) and scope to the number of explanatory variables (X). Volume facilitates the determination of the specification of models because the larger the number of individuals observed (N), the greater the degrees of freedom to include more variables (X). Scale and scope are a direct consequence of two Vs of the BD definition (Laney, 2001, 2012): volume and variety. Knowledge extraction from BD using AI also has high fixed costs and almost negligible variable costs. Concentration arises from the efficiencies that derive from lowering the average cost of producing knowledge/information from bigger and more detailed data sets. Concentration leads to the emergences of a small number of digital-giant-data-holders that expand to cover as many human activities as possible.

Oligopolistic data holders compete but have incentives to collude. They compete in a data collection race towards $N=\text{all}$ and $X=\text{everything}$ but knowledge production is a natural monopoly where the presence of economies

of scale and scope, high fixed costs, small marginal costs and other barriers to entry operate together. Diminishing returns to scale in knowledge extraction from BD never arrive. To take the extreme hypothetical case, where AI technology is common across all firms, then its efficiency is enhanced by access to the largest BD “lake”. This creates a “winner takes all” dynamic loop in which the holder of largest BD lake generates the largest knowledge rents and innovation, which can then, in part, be used to further augment the size of the BD “lake” to which the Data holder has access. Continuing the natural monopoly logic, a small number of operators that have incentives to collude influence the market as a whole and may tend towards one single operator over time in order to avoid wasteful duplication of resources. The amount of knowledge produced, its disclosure, its prices and its quantities then follows monopoly theory (Shumpeter 1942):

First, data holders’ expansion across sectors and activities generates a process of creative destruction that replaces less efficient and effective traditional operators that lag behind in their ability to collect BD and generate knowledge. Traditional operators may not be not less efficient in terms of monetary profit but their data generation capacity. For example, Uber does not yet make monetary profit but is more efficient than traditional taxis in generating data and attracting investment.

Second, market structure and lack of competition attracts investment for R&D and innovation.

Third, data holders are able to set prices and quantities. Regarding prices paid with data, the market works on a “take it or leave” basis: digital services are often only available to consumers that are ready to provide data as an (implicit) part of the bargain. Regarding quantities, data holders set the amount of knowledge production to the amount that maximize their profit. Such a quantity is expected to be below the socially desirable amount. In practice, controlling knowledge production and disclosure implies information asymmetries between data holders and the rest of the agents in the economy such as consumers, central banks, antitrust authorities, and the scientific community. Data holders’ “de facto ownership” keeps citizens and the public sector outside the black box and without access to the data lake. Underproduction of knowledge implies an opportunity cost for the whole society. Like in a monopoly, part of the knowledge is not produced and therefore the society loses it. Part is produced but not disclosed, and it is therefore subtracted from consumers’ surplus in a hypothetical BD-knowledge space.

In the early stages of digitalization, Governments intervene without considering the data and knowledge dimension of the economy. Equation 1 captures the traditional country-level macroeconomic equilibrium, without the data dimension. The state of (macro) economic equilibrium occurs when total leakages (savings (S) + taxes

(T) + imports (M)) are equal to the total injections (investment (I) + government spending (G) + exports (X)) that occur in the economy. This can be represented by:

$$S + T + M = I + G + X \quad (1)$$

Disequilibrium occurs when leakages are not equal to the total injections. In such a situation, changes in expenditure and output will lead the economy back to the state of equilibrium. Such changes will depend on the type of inequality ($S + T + M > I + G + X$ or $S + T + M < I + G + X$)

As data holders are supra-state agents that operate globally but typically concentrate in low-tax jurisdictions, the ability of governments to collect taxes decreases ($T \downarrow$) and reduces Government's financial capacity, spending ($G \downarrow$) and ability to respond to market failures and promote efficiency, equity and stability. Data holders' ability to collect valuable data increases their financial power and capacity to attract investment ($I \uparrow$). Data and the capacity to process them are a critical ingredient of innovation, which makes data holders an attractive store of value for investors. The resulting financial strength allows data holders to predate markets. Creative destruction applies also to sectors traditional provided by the State such as health, education, public transport and national defense.

$$S + T \downarrow + M = I \uparrow + G \downarrow + X$$

There are no data nor knowledge dimensions in Equation 1. Monetary taxation of digital activities contribute to balance equation (1) without the need to reduce Government expenditure ($G \downarrow$) but do not tackle the data and knowledge aspects of the semicircular flow of the economy.

3.- Does the world fit the semicircular model?

Means of payment in many digital services are personal, accompanying usage data and sometimes a monetary payment (Evans, 2013; Scott Morton et al., 2019; Tett, 2018, Brynjolfsson et al., 2018). This is the case of services offered by search engines such as Google and social networks like Facebook or LinkedIn, other services like Dropbox, Spotify (Kramer and Kalka, 2016), and platforms like Airbnb, Couchsurfing, Zipcar, Uber, Lyft, BlaBlaCar, TaskRabbit, myTaskAngel, Freelancers, etc.... For example, Facebook offers a 'free to use' digital service that allows people to stay in touch. Users generate data when they create a user profile indicating their name, occupation, schools attended. and when adding other users as 'friends', exchanging messages, statuses, pictures, videos, links, 'likes', and

other Facebook reactions together with the exhausted data (paradata, environmental data, or footprints) related to their activity. Similar examples are Instagram, a photograph- and video-sharing service, or WhatsApp, a messaging service, which are free to use and do not generate direct revenue but do generate data.

More and more devices contain sensors and more activities generate data. There is an increasing capacity to pump zettabytes of unstructured data towards data holders (The Economist, 2017). Daily activity is a data factory that produces feedstocks: data about personal relationships, health, mood, locations, movements, diverse amount economic activities, C2C, P2P, B2B, B2C ...

Regarding consumers' utility maximization, the "individuals' data price equals zero" is, in the initial stages of digitalization, a realistic assumption. Increasing digital literacy, however, makes individuals more aware about several issues: the ability of data holders' to protect their privacy, the limitations of traditional policies to guarantee the rule of law and competitive markets in a data intensive economy, and the dangers of excessive state intervention (Freedom House 2019). At the moment, individuals have very few alternatives to exercise their personal data property rights but to accept unclear terms and conditions from data holders (Cakebread, 2017, Obar and Oeldorf-Hirsch 2018). In addition, they do not have the ability to merge nor process data with other individuals to obtain valuable data sets bypassing data holders. Citizens' control over their own data is very limited.

Data holders are profit maximization companies. Sometimes they do not obtain a direct monetary compensation from the digital services but the "de facto ownership" of data. They use data to produce knowledge about patterns, regularities and structures of human behaviour and activities (Dosis and Sand-Zantman, 2018; Jones and Tonetti, 2018, Scott Morton et al., 2019, Boisot and Canals, 2004; Duch-Brown et al., 2017; Vigo, 2013). Their activity benefits society via innovation and expansion of AI (towards $X = \text{everything}$ and $N = \text{All}$) that generate a process of creative destruction: uber vs taxis, rbnb vs hotels... that also affects sectors traditionally provided by the public sector such as public transport (Evgeny 2015), health care (Carrie Wong 2019), banking (Mercola 2020) and national defence (Brustein and Bergen 2019).

Regarding, markets' valuation of individual data vs market valuation of data factories, according to the *Financial Times* (Steel 2013, Steel et al., 2013), data brokers pay between EUR 0.0005 and EUR 0.66 (calculations made in October 2018) for the data of individuals, depending on personal characteristics and the amount of detail. Huge amounts were paid for (apparently) non-profitable companies that have developed services with network effects and the ability to collect data (Bond and Bullock 2019; Kaminska, 2016; McArdle, 2019). Instagram and WhatsApp's

acquisitions by Facebook in 2012 and 2014 respectively and Google's acquisition of YouTube in 2006 are good examples. Individual data is almost valueless (Steel et al., 2013), only having hundreds of millions adds value to data (Worstell, 2017), and such a value is only realised after knowledge extraction. Acquisition of data, the capacity to generate more data, and competition reduction justifies the valuation of these data factories.

Knowledge production using BD and AI resembles a natural monopoly. It has very high fixed cost and negligible variable costs (Duch-Brown, 2017a). Fixed costs refer to connectivity infrastructure such as broadband (Unctad, 2017), research and development, data centres, cloud computing arms, and data refineries to handle data generation, collection, and processing (The Economist, 2017). The more data feed self-optimising AI algorithms (Silver et al., 2017) the more AI improves. Despite the clear positive impact of merging data in data value, identifying where economies of scale stop and give way to diminishing returns is an empirical question on which there is little evidence (Codagnone and Martens 2016).

Costs of diversification and innovation oppose scale, scope, and concentration in services, products, and data production markets but not in knowledge production. For example, Facebook, WhatsApp and Instagram may compete as social networks with different specializations in the digital service markets but extraction of knowledge is more efficient if data obtained from them are analyzed using the same tools and methods. Platform economy has its limits (Azzellini et al. 2019) but data holder expand into physical production and sectors where platforms are not yet taking over. Amazon acquisition of Whole foods, that increases data collection to offline activities, illustrates how knowledge extraction is a specialization itself (Hirsch, 2018; Krugman 2014). Sofa Sounds partnership with Uber and AirBnB also illustrate data driven expansions without high diversification cost because it does not imply a new specialisation. Knowledge extraction is not specifically included in the NACE classification system as a sector per se which would pave the way to better data economy concentration indexes and measures.

Data-driven acquisitions, interconnections and partnerships between companies resembles and spaghetti bowl. For example, MasterCard Advisors are IBM Watson partners. PayPal is, in principle, a Mastercard competitor, but Mastercard owns a percentage of PayPal and PayPal is a Facebook partner. Facebook has received investment from PayPal. In China, social networks and the payment industry are already integrated into the same company through the Chinese 'WeChat', which, in a single application, offers services such as Instagram, Facebook, and WhatsApp together with payment services. Google's acquisition of DeepMind, the world AI leader, in 2014 also illustrates the reinforcing nature of BD and AI. DeepMind also has access to public records through its agreement with the United

Kingdom's National Health Service (Lomas 2019). IBM's acquisition of the Weather Company in 2015 illustrates that concentration goes beyond personal data to information on variables that determine consumer behaviour.

There are incentives to centralise knowledge production in what could be collusion practices in knowledge extraction if partnerships imply data merging. For example, during the process of obtaining European Commission approval to merge Facebook and WhatsApp (European Commission, 2017), Facebook pledged that it would not merge user -bases but, as far as we know, there has been no authority supervising that it does not do so.

BD and AI reinforce each other and the concentration process. Data holders expand investing in companies able to generate data but also in AI companies. For example, Facebook's investment in DeepText, an AI natural language processor able to learn the intentions and context of users in 20 languages, and face recognition technologies show that concentration follows a BD and AI reinforcing loop (Pedraza and Vollbracht 2019). In general, data is a critical ingredient to feed AI models and innovation (OECD, 2019). Expansion also affect mobile devices and gadgets, such as smart watches, that generate more data. There are evidences supporting market concentration in the global economy (Mckinsey 2019, 2018, Scott Morton 2019) and the digital sector (OECD 2019, Unctad 2017). A shrinking number of companies dominates increasing number of industries which is accompanied by declining in start-up growth and less financial resources for them, fewer high-growth young firms and growing inequality (Khan 2017, Porter 2016, Jarsulic et al. 2016, Decker et al 2018).

All the above explains the Khan's anti-trust paradox (Khan 2017): the limitation to cognize harms to competition from short-term prices and outputs when the long-run competitive advantages from knowledge generation and innovation is an important driving force behind concentration. Data holders and investors maximize data collection and expand their data collection infrastructure because access to data and knowledge shapes globalization, innovation, and wealth and power distribution (OECD 2019, ITU 2018, Freeman 2013). The data economy is not a small add-on the circular flow but the key to long run growth market power and dominance (Arthur 2011). In fact, although dominance has grown also thanks to merges and proprietary market places allowing data holders to crush competitors and favor its rankings and sell their own brands, digital giants often have meagre profits but set their priorities on intensive data-hungry growth. In the meanwhile, advertising is often their main revenue (Facebook, 2014; Statista, 2015, Khan 2017).

Data are not only the "oil of the twenty-first century", the key input to knowledge and innovation, AI development and knowledge generation market power (OECD 2019, Liem and Petropoulos 2016, UTI, 2018, OECD

2019). They are also the source of additional market failures. Access to data generate information asymmetries opening opportunities for price discrimination, steered consumption, and unfair competition in sectors different to knowledge generation (White House, 2015, Ursu, 2015, Mikians et al. 2012, Shiller 2014, Chen et al. 2015, Möhlmann and Zalmanson 2017, Uber 2018, Ezrachi and Stuke 2016). Discrimination can go beyond prices and lead to unfair treatment and discrimination in general (Isaac, 2017; Wong, 2017). Asymmetric information may also support predatory pricing and monopsony behaviours (Bensinger, 2012; Bond and Bullock, 2019; Kaminska, 2016; McArdle, 2019, Codagnone and Martens, 2016). Regarding rule of law, services emerging in the data economy, especially in the sharing economy, challenge aspects like consumer protection, professional licenses and regulations vs informal supply of services, working conditions (Hall and Krueger 2015, Cook et al. 2018), quality standards (Codagnone and Martens, 2016, Vaughan and Hawksworth, 2014; Malhotra and Van Alstyne, 2014), and tax avoidance ($T\downarrow$) (D’Andria, 2019). In addition, some hedge funds operating in markets around the world employ AI models, and treasure BD lakes and human intelligence to obtain inside information about the economy. Their expertise has recently been used in electoral campaigns in a decisive way (Grassegger and Krogerus 2017; Kosinski et al, 2013). The same methods, agents and algorithms to model both electoral campaigns and trade on the financial markets generate a situation that goes beyond market concentration and resembles the separation of powers problem (Kee 2018; Cadwalladr, 2017).

Such a background makes data holders very attractive to investors ($I\uparrow$), and expands the creative destruction to the State itself ($G\downarrow$). $T\downarrow$ and $I\uparrow$ generate a disequilibrium like $S + T + M < I + G + X$ where $G\downarrow$, changes in public expenditure, lead the economy back to an equilibrium where the role of State diminishes. Similarly as the antitrust paradox (Khan, 2017), such an equilibrium misses the data and knowledge dimension of the economy. Knowledge generation lacks transparency, occurs within a black box. Data holders “De facto ownership” operates as a ‘breastplate’, a shell that avoids additional merit knowledge. The United Nations (UN, 2014, UTI, 2018) has reported growing inequalities in access to data, information, and the ability to use them. Distribution of information generates asymmetries and foster inequalities (Duch-Brown, 2017; Stiglitz, 2001). No authority is in charge of efficiency, equity, stability and redistribution from a data and knowledge points of view.

4.- Socially desirable Knowledge production: Optimum data sharing

There are goods that if provided by the free market can be under-consumed (merit goods like education) or over-consumed (demerit goods like illegal recreational drugs) (Musgrave 1959). For example, without education and maturity an individual cannot make a well-informed choice about the amount of education he should consume. His decision affects the whole society because individuals' education displays positive externalities on societal well-being, citizen's security and economic growth (Lucas 1988, Munich et al 2018). Similarly, a drug addict cannot decide for himself and illegal drug consumption generates negative externalities on the rest of the society through health and security expenditures. The idea behind merit and demerit goods is that a well-informed society is in a better position to identify the amount needed of certain goods. As a result, governments impose community standards and support consumption of merit goods and ban, or discourage, demerit ones. In the traditional circular flow of the economy, a fiscal authority follows policy decisions to collect taxes (so-called leakages) from agents and deliver merit goods (so-called injections) to the whole society (Samuelsson 1948, Samuelsson and Nordhaus 2010). Scientific knowledge about individual and community consequences of merit and demerit goods is the basis of the government's intervention. It aims to improve consumers' and citizens' capacity to take informed decisions by tackling information failures.

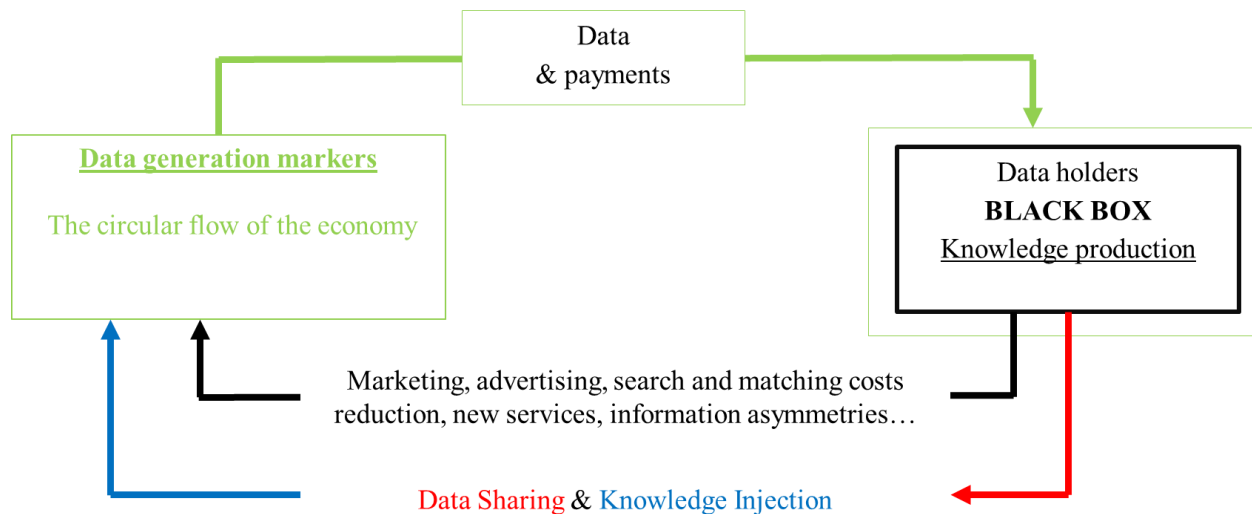
Depending on the usage and type of knowledge generated, data can be a merit or a demerit good. Data are a merit good when used to innovate and reduce market frictions, information costs and asymmetries. In those cases they generate better matches between supply and demand and, using the sharing economy as an example, facilitate the full utilization of private assets that otherwise would be idle. Data are also a merit good if used to conduct scientific research and obtain empirical evidence to support policies. For example, to study market structures and anti-trust concerns, adapt the existing legal corpus to the new digital reality, find ways to foster competition, and promote transparency, rule of law and enforcement, or to forecast the economic cycles and deliver nimbler and faster anti-cyclical policies. Data are a demerit good when used to violate privacy, to generate market power, to set barriers to entry, to generate information

asymmetries or unacceptable distribution of wealth, to control market places and damage competition, to charge unfair fees or prices, to monitor and control citizens' lives, manipulate political campaigns or to impose excessive regulations limiting innovation.

Following the conceptualization of merit and demerit uses of data, we can define the *under-production of knowledge* as a situation characterised by the existence of barriers to entry to the production of merit knowledge from data. We can also define *Data-sharing policy* as analogous to *taxation* in the traditional model, as a data policy that generates a data flow (leakage) to promote (merit) knowledge production (injection) from data generated from economic activity. A merit data (taxation) sharing policy would only promote merit uses of data. A Pareto efficient data sharing policy improves the situation of agents that are the beneficiaries of the intervention, mainly households and firms, displaying positive externalities for society at large without generating negative consequences on efficient allocation of resources, nor discouraging investment and R&D activities. A Pareto efficient intervention does not rival data holders' activities. It generates a higher level of data flows available to the agents in the economy that are able to produce merit knowledge. Generating a direct flow of data back to firms and families would not solve information asymmetries because they, in general, do not have the ability to extract knowledge from BD. Following the leakages and injections analogy, figure 2 represents data flows (leakage) towards agents able to extract merit knowledge from data generation and additional flows of knowledge back to society aiming to increase efficiency, equity and stability (knowledge injection).

There are many public authorities (potentially) able to produce merit knowledge as a result of non-rival and Pareto efficient *data sharing policies* that would increase the supply of data available to them. For example, central banks, antitrust authorities, the scientific community, and other agents that are not direct competitors to data holders could contribute to a better-informed society. Central banks could improve their understanding of the economic cycle. Antitrust authorities could enhance research on sources of unfair competition, deliver antitrust policies and balance information asymmetries specific to a data intensive economy. The scientific community has shown that, if not limited by data access, it could enhance knowledge about many research topics and phenomena (Schroeder and Cowsls 2014).

Figure 2.- Semi-circular flow of the economy, leakages and injections.



There are three possible channels to enhancing data-sharing to increase the flow of data towards merit users that State intervention could promote. First, data holders, as de facto owners of data, can share their data sets with non-rival and merit users as part of their data philanthropic marketing. Second, as with other merit goods, the State can promote enhanced data flows to merit users via incentives and disincentives linked to a data tax system and similar to traditional taxation. Third, as owners of their personal data, citizens could be encouraged to provide their personal data for research purposes and in order to promote social welfare.

State intervention may of course solve certain market failures but generate new ones. Figure 8 represents the *Data sharing Laffer curve*: the theoretical relationship between the *data sharing in the economy* and the amount of merit knowledge generated. The main idea is that, as monetary taxation has a point of maximum tax revenue, *data sharing*, has a point where the economy generates the maximum amount of knowledge possible given the state of technology.

The horizontal axis represents the level of data sharing. We use a number from 0 to 100 purely for convenience and ease of explanation. Accordingly "0" represents an economy free of any data-sharing

responsibilities and “100” represents a “Big Brother” Orwellian world of total data sharing obligations by all actors in the economy. The vertical axis represents the amount of merit knowledge generated in the economy.

The relationship between knowledge generation and the level of data sharing resembles that of the traditional Laffer curve. At $DS=0$ only data holders draw value from data generated in the economy to maximize their profit and their market power. $DS=100$ represents the total negation of data holders’ de facto ownership and individuals’ property and privacy rights. It implies total “Orwellian” surveillance of all activities by the state. At $DS=100$ neither data holders nor citizens have incentives to participate in a data generation economy and no knowledge is generated from BD. Between 0 and 100 there is an infinite number of data policy options. Up to the optimum rate of data-sharing the relationship is positive: increases in data sharing generate higher levels of merit knowledge. Beyond that point, the higher levels of data sharing lead to a declining amount of merit knowledge production.

If the *Data sharing Laffer curve* is also a merit knowledge production possibilities frontier, bigger data lakes, developing scale and scope, innovation, increasing data collection activities, and transparency and the rule of law enforcing trust in the data economy move the curve upwards. Barriers to entry, lack of trust and competitive markets, and disincentives to invest in innovation move the curve downwards.

From $DT=0$, data holders can voluntarily start opening BD making them available to other agents by means of APIs or ad hoc non-disclosure agreements. Data holders “data philanthropy”, marketing and willingness to activate research community around their interests may increase data sharing up to $DS=$ “de facto ownership” generating higher amount of knowledge than at $DT=0$. Data taxation based only on data holders’ good intentions does not increase the amount of knowledge generated to its optimum level. Data holders have incentives to keep BD and knowledge generation internally and knowledge disclosure at the level where they keep market power, barriers to entry and information advantages and asymmetries as much as possible.

If DS is “de facto ownership” then $DS < DS^*$ because data sharing is still in the positively sloped part of the curve. Further increases in data sharing up until the point $DS=DS^*$ allow higher non-rival and

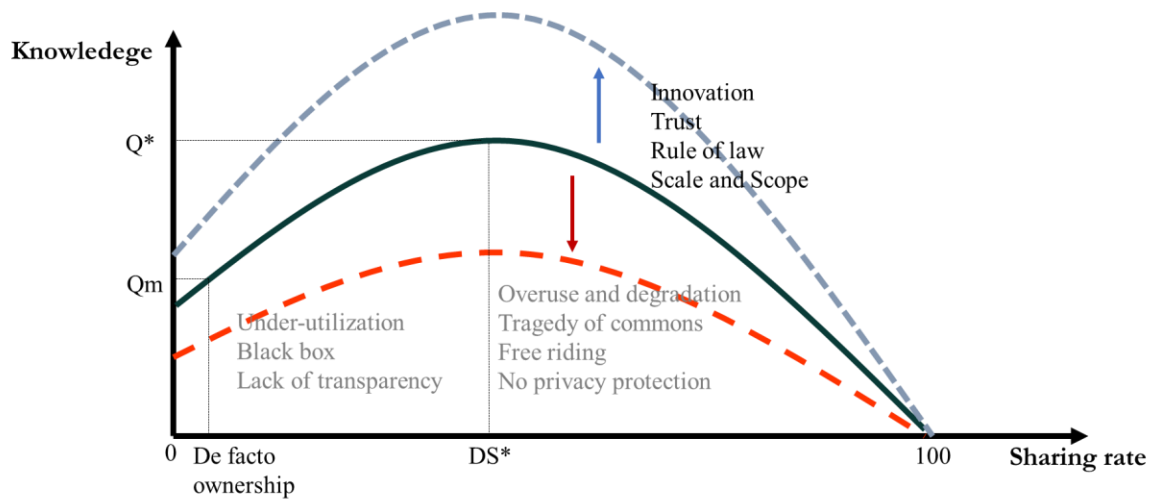
merit knowledge generation. Carefully designed data policy can therefore generate a movement along the curve. Up to the “optimum data taxation point” ($DS=DS^*$) more data sharing has a positive impact on merit knowledge production. Resulting knowledge injections into the wider economy in turn increase social wellbeing.

However, beyond the optimum level of $DS=DS^*$, where $DS>DS^*$, additional increases in data sharing generate demerit and rival uses of data, which have negative consequences upon the data generation process. Rival uses of data create inefficiencies. Free riding upon the data discourages data holders’ investment in data generation activities and innovation. At these high levels of obligatory data sharing, privacy and other citizens’ rights start to be disregarded. This situation is akin to a kind of “tragedy of commons” in the data economy. For example, data sharing is used to monitor and control citizens’ lives which erodes the legitimacy of the system itself. It also reduces consumers’ willingness to use their personal data as a means of payment, thereby reducing the amount of data generated in the economy. As a consequence at $DS>DS^*$ increases in data sharing obligations generate lower levels of merit knowledge production relative to DS^* .

At $DS=100$ economic activity generates no data. Households and firms see their privacy violated and they do not participate in activities where they (would) pay with data. Data holders do not find it profitable to invest in activities that produce data. As a result, there is no data sharing (tax) base. This is analogous to the theoretical situation whereby there would not be any monetarily-defined economic activity if traditional tax rate on economic activity would be set at 100%.

At $DS=DS^*$ society generates the maximum amount of merit knowledge by optimising the size of data lakes available to society. At DS^* , data policy preserves incentives to invest in data generation, fosters innovation, trust, transparency, the rule of law and increases confidence in data as a means of payment. This situation increases the amount of data generated in the economy, data-driven innovation and moves the Laffer curve upwards. It also facilitates governments’ role, fosters economic stability, reduces market failures, barriers to entry, balances information asymmetries, and fosters competition.

Figure 3.- Data sharing Laffer curve (theoretical presentation)



5. -Where is the economy located in terms of the Data taxation Laffer curve?

There are three AI leaders worldwide: the USA, China, and the EU (European Commission, 2018a) with the EU lagging behind the first two.

In USA and Europe, corporate data holders decide for what and to whom users' data are available. Scientists access data in several ways. First, they can explore the surface of the digital economy by web crawling (Pedraza et al 2019). Second, they can benefit from (non-disclosure) agreements, but such agreements may generate a data divide among scientists putting replicability and FAIR (Findable, Accessible, Interoperable, Reusable) principles in danger (Wilkinson et al. 2016, Taylor et al. 2014, Codagnone and Martens 2016, Malhotra and Van Alstyne, 2014, Hall and Krueger 2015, NSF 2017). Third, they can use the data crumbs that data holders make available to activate the researchers' community to obtain new perspectives of their own business. This is the case of Google trends and other "Data Philanthropy" initiatives (Pawelke and Tatevossian 2013). Internet searches contain insights into diverse human activities (Askatas and Zimmerman 2009, 2011a, 2011b, 2011c, Choi and Varian 2011) but the data released is not enough to build and test consistent and stable models (Artola et al 2015) as was shown by the Google flu predictions (Ginsber et al. 2009, Butler 2013).

Although all these data have generated thousands of academic papers, the economy seems to be at or close to the data sharing rate shown in the figure 3 as data holders' "de facto ownership" and therefore well below the optimum data sharing policy. First, because the amount of knowledge produced in the economy is the amount that

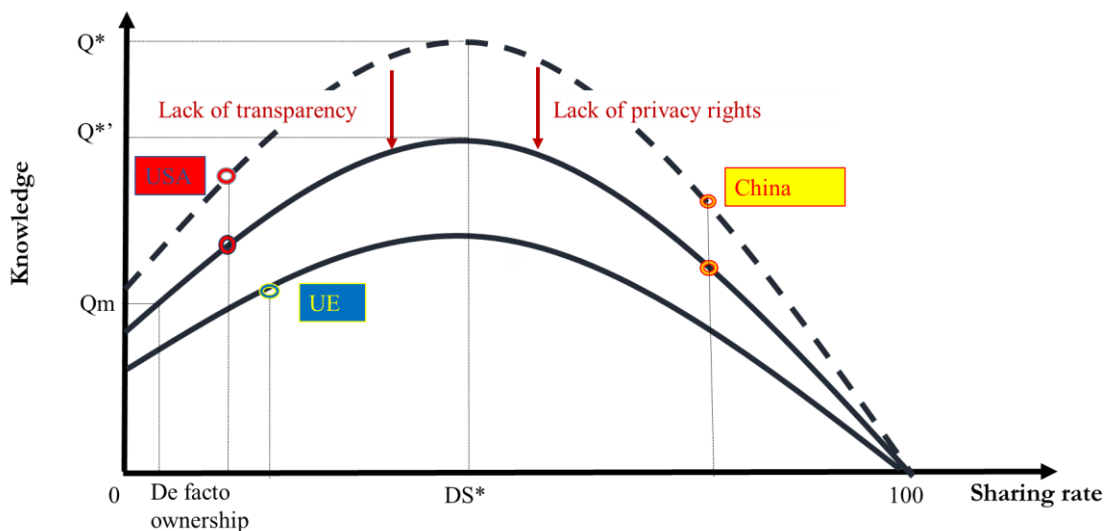
maximizes data holders' profits and their data collection, further knowledge production relies on their good intentions via data philanthropy (Taylor et al. 2014, Einav and Levin 2013) rather than state intervention or individuals' exercise of their data property rights. Second, because current data sharing towards merit users is not enough to produce the socially desirable amount of knowledge: Regulatory authorities and the scientific community remain unable to fully tap into innumerable aspects of digital policymaking (Khan 2017, Scott Morton et al., 2019; Taylor et al., 2014; Butler 2013; Artola et al 2015; Lazer et al.2014).

Figure 4 is an attempt to represent the respective AI leaders' Laffer curves assuming, for clarity of explanation, common slopes and optimum data sharing rates DS^* across countries. Both the EU and the USA have similar, data sharing levels. The sharing rate in the EU is represented slightly higher because the GDPR regulation supports data portability (according to art. 20 of the GDPR data can be transferred from one controller to another) and the European Commission's Data Strategy supports data sharing (European commission 2020). However, the US knowledge production frontier is higher because of the strength of American corporations that possess huge BD lakes and AI capacity. US is the world leader in start-ups and venture capital. The amount of knowledge produced by the EU is below US but still in a good position regarding AI publications (European Commission, 2018a). Centralization in China, without clear distinction between data holders, state, supervision, surveillance and the presence of multifaceted tools like WeChat gives the country a competitive advantages to develop huge BD lakes. China is the world leader in turning research into patents (European Commission, 2018a). We represent China's data fiscal pressure as beyond DS^* and in the downward sloping part of the curve. According to Freedom House (2019) China is the world leader in developing and exporting social media surveillance tools. The US, although considered a free internet country according to the same Freedom House assessment, has also suffered, among other things, a proliferation of false content on the internet that has done harm to the capacity for free and fair elections to be held. One case represents excessive intervention where a central power disregards individual rights, the other represents a situation in which private interests can operate with very little transparency.

According to our theoretical rationale, none of those three situations is static. As digital literacy evolves, actors in the economy become more aware of implications of *de facto* payments using data. When data are used against consumers' interests, for example, to generate demerit knowledge to support unfair competition, price discrimination, manipulation, political distortion, surveillance..., actors change their data generation behaviour (Toscano 2019). For example, although by October 2018, 74% of Facebook users were not aware that advertisers were able to make use of

their lists of interests for targeting purposes, after the Cambridge Analytical scandal, many users (54%) adjusted their privacy setting, started using it less frequently or even left the app (Gramlich 2019). This is one example of the hypothesis that either excessive intervention or complete lack of transparency reduces trust and individuals' willingness to "pay" with data. Both move the Laffer curve downwards, reducing the knowledge production possibilities and effective and potential innovation capacity (from Q^* to Q^{**}). Just as payments in non-reliable currencies are not accepted, economic agents, sooner or later, will demand legal security to use their data as a means of payment.

Figure 4.- Data sharing Laffer curve (some current examples)



Against this background there have been four types of reactions: Taxes and fines, data holders' conglomerates break ups, data taxation and algorithm transparency, and user centric data property rights.

First, following a traditional view of the economy, some countries have approach the issue from a monetary point of view with unilateral taxes and fines (Pratley, 2018; Sandle, 2018; D'Andria 2019). Other countries and international institutions have followed and are now devising how to tax digital activity (European Commission, 2018b, OCDE 2019, D'Onfro and Browne 2018, European Commission 2018b, Khan and Brunsden 2018) and collect money through fines. In Germany, for example, banned content not removed by Facebook within 24 hours faces fines of up to 50 million euros. In a more user-centric approach, Posner and Weyl (2018) propose that agents could be compensated by the data they generate just as they are compensated for their labour or in the form of a dividend (Ulloa,

2019). Such compensation still has a ‘monetary’ view of the data economy and does not take into account difficulties in pricing individual data and lack of competitiveness in generating value added from knowledge extraction. These reactions increase financial power of Governments and their ability to tackle the equilibrium described in equation 1. However, they do not contribute to produce merit knowledge. It does not build trust or transparency, nor does it shed light on demerit knowledge generation. It keeps merit users and citizens displaced and outside of the “black box”.

Second, it has been argued that breaking up companies like Amazon, Facebook and Google (Alphabet) would generate enhanced competition. It has been proposed, for example, that platforms should be broken up from any participant on that platform, and forced to meet non-discriminatory standards and forbid any transfer of data to third parties. Breaking up, however, may imply duplication of resources and lower innovation from not taking full advantage of economies of scale and scope. The division of conglomerates does not solve the issue of giving people more control over their own data, nor does it add transparency to the black box and the spaghetti bowl of data alliances.

Third, there have been claims for data and algorithm transparency and awareness (European Commission 2016, Connolly 2016) and agreements at local level, such as the city of Boston and Uber (Evgeny 2015) and at government level, such as the Macron-Stakelberg agreement (Scott and Young 2018, Barzic et al 2018). Boston accepted Uber as legal in exchange of quarterly data that can improve traffic and urban planning. The Macron-Stakelberg agreement was a posteriori reaction to the alleged political ad-targeting scandals in electoral campaigns. It allowed six French officials to work at Facebook for six months examining how to combat hate speech. Macron considered the agreement an experimental approach to a new “smart regulation”. Both are good examples of merit, non-rival Pareto efficient intervention (Gold 2019, Askitas 2015). These efforts are, however, not coordinated, and not stable or large enough to develop lasting economies of scale and scope. None of them allows a systematic exploration of data and algorithms. They can be criticized for giving data holders the right to operate as data intermediaries (Evgeny 2015) and exercise market power (Bergemann and Bonatti 2018). They are also not user-centric. The good news is that these initiatives go beyond the traditional view of the economy and tap into the data and knowledge dimension of it. Similar formulae could expand to other realms of digital life and to other data holders. As pointed out by Askitas (2015, 2018), governments may further encourage corporate data taxation opening up data for the benefit of society, while also protecting their corporate interests and user’s privacy concerns. Macron’s “smart regulation” can be considered a first experimental step of Askitas’ proposal that would generate a flow of data (leakage) to merit users. In terms of the Laffer curve, it would generate a movement along the curve. If accompanied

by knowledge based policies focusing on trust, consumer rights and so on, it would also move the Laffer curve upwards. It would generate positive externalities to society as a whole, including data holders whose businesses depend on trust in the data generation process and willingness to use data as a means of payment.

Sometimes state intervention, instead of solving existing problems, generates new ones. Data taxation can also have negative consequence on trust and move the curve downwards if it is not clearly differentiated from surveillance (Lyon 2014). According to Freedom House (2019) 40 out of 65 countries have installed advanced social media surveillance programs to monitor users. Data sharing and merit knowledge are not about authoritarian governments using advanced tools and artificial intelligence in a way that undermine civil rights and freedom. User centric, bottom-up, approaches based on individual decisions may be more efficient in maintaining trust which leads to the next initiative.

Fourth, clear and real property rights have always been a prerequisite for a well-functioning market economy and consumers' utility maximization. Data sharing towards merit users could be based on consumers' individual decisions. In practice, free movement of data and user's ownership of their personal data are very limited: users are the legal owners (European Union, 2016, Jones and Tonetti 2019) but data holders, collect, control and draw value from them. Data property rights that citizens are able to exercise have to be accompanied by tools and infrastructures that enable citizens' decisions and empower them.

From a legal point of view, the European GDPR (European Union, 2016), which came into force in 2018, is a legal global benchmark that sets the legal basis of a user-centric approach. The GDPR intends to facilitate the free flow of personal data with the goal of protecting the rights of citizens. According to De Hert et al. (2018) the right to data portability is the novel feature of the GDPR that forms the basis for additional regulation beyond data protection and towards competition law or consumer protection.

From an infrastructure point of view, Personal Data Stores (PDSs) are an emerging business model that aims to facilitate users' exercise of their personal data property right giving users more options to control their data in terms of permissions to access and generation of value (Bolychevsky and Worthington 2018). A data authority, as proposed by (Martens, 2016 and Scott Morton et al. 2019), that enforces existing data protection and other rights with specialist and data analytics staff and infrastructure could trigger the user-centric approach and empower users (GDPR art 51, European Union 2016). Users 'consent for merit access to data would contribute to benefit the whole society with the positive externalities of data aggregation and knowledge extraction. According to the GDPR (European Union, 2016),

lawful processing (Art 6) of data can be based, for example, on consent (Art 7) that has to be given for each specific and explicit purpose (Art 5.b). Data holders who have received informed consent from a data subject can only use the data for the specific and explicit purpose for which consent is given.

The OECD Cancun Ministerial declaration on the digital economy (OECD 2016) recognised the need of collective and internationally coordinated action to promote research, rule of law, trust, competition, and transparency, consumers' protection, working conditions and regulation in general. The World Economic Forum (WEF) stakeholders approach (WEF 2019) implies careful blending and balance of many kinds of organizations, from both the public and private sectors, international organizations and academic institutions. As there are International offices for specific topics, there could be an International Data Organization, in charge of coordinating and developing the infrastructure and the general ethics and principles. International monetary taxation of digital activity could provide the resources for these developments.

From a scientific point of view, a data-intensive economy needs data-intensive science to develop new forms of digital sciences (Martone et al 2016), new theories, methods (Varian, 2013, Steinmetz et al. 2014) and discoveries. A Data Authority could deal with the challenges of data not created for scientific research but with increasing scientific interest: data accessibility, removal of barriers to data use and re-use, data management, repositories, collaborative data infrastructures, preservation, citation principles, reproducibility, openness accessibility, ethical principles, data scientists' professional codes and standards (Wilkinson et al. 2016, Starr 2015, Lecarpentier et al 2013) ... Making data from economic activity available and reusable for behavioural, economic and social sciences, like other sciences are doing (Bauch 2011), would generate synergies and cross-fertilization of disciplines.

From an economic intuition point of view, consumers maximizing their utility beyond the individual's data zero-value-assumption and data holders maximizing their profit should lead to the production of the socially desirable amount of knowledge (Jones and Tonetti 2018). In short, a higher level of data sharing would generate knowledge that would fit in the definition of a public good: 'a good that all enjoy in common and each individual's consumption of it leads to no subtractions from any other individual's consumption' (Samuelson, 1954, 1955).

6.- Conclusions

Economic theory is a toolbox that helps us to understand the complexities of the world around us. This paper develops the semicircular flow of the data economy, a theoretical simplification of the data intensive economy that we use to explore challenges of digital governance. Knowledge production using BD and AI play a central role in a

data intensive economy where the traditional explicit prices and quantities view of the economy fails to provide a complete conceptual framework.

AI knowledge production exhibits many natural monopoly characteristics in a manner that triggers a process of market concentration and fosters the emergence of giant data holders specialized in collecting and processing big data (BD). In the absence of clearly exercisable property rights, data holders set knowledge production, knowledge disclosure and data sharing to the levels that maximizes their BD collection and economic profit. Incentives to collude and set barriers to entry for potential new entrants limit access to BD that could increase merit knowledge production beyond current levels. On the one hand, data holders speed up innovation via a creative destruction process. On the other hand, they cause market failures and information asymmetries. But if monopoly theory holds, the amount of (merit) knowledge produced is below the socially desirable amount.

Assignment of property rights and effective political institutions are key determinants of the long run economic success or failure. While *extractive* institutions allow elites to capture society's resources, *inclusive* institutions spread economic benefits more widely and foster innovation and technology being technological change the most important long-run source of growth (Acemoglu and Robinson 2012, Galenson 2017). From a policymaking point of view, the government's role in a data-intensive economy is to promote competition, reduce market failures, protect privacy, promote merit knowledge and reduce demerit knowledge, and to help individuals to fully exercise their data property rights. This implies the development of an infrastructure to generate the provision of data towards merit users thus promoting additional knowledge injections into the economy. This can be fostered by developing data sharing policies (Askitas, 2015) as well as being driven by individuals' decisions (De Hert et al., 2018, European Commission 2016). Data sharing and knowledge injections could be coordinated by a data authority, as proposed by Martens (2016) and Scott Morton et al. (2019), able to benefit from economies of scale and scope. It could generate positive externalities in the whole of society, including for data holders. However, for developments of this type to be in line with overall societal interests, such policies must also guard against the risks of over-acquisition of data by the state in a manner that could facilitate intrusive surveillance without fostering innovation to the benefit of the wider economy.

This theoretical analogy needs further micro and macro developments and further analytical evidences to be fully supported. It also needs to build upon existing codes, standards and very strict data experts' deontology and existing computer science developments.

References

- Acemoglu, Daron, and James Robinson. 2012. *Why Nations Fail: The Origins of Power, Prosperity, and Poverty*. New York: Crown Bus
- Artola, C., Pinto, F. and de Pedraza, P., 2015, 'Can internet searches forecast tourism inflows?', *International Journal of Manpower*, Vol. 36, No 1, pp. 103-116 (<http://www.emeraldinsight.com/doi/pdfplus/10.1108/IJM-12-2014-0259>).
- Askatas, N. and Zimmermann, K. F., 2009, *Google econometrics and unemployment forecasting*, IZA Discussion Paper No 4201, June 2009, Institute of Labor Economics.
- Askatas, N. and Zimmermann, K. F., 2011a, *Health and well-being in the crisis*, IZA Discussion Paper No 5601, March 2011, Institute of Labor Economics.
- Askatas, N. and Zimmermann, K. F., 2011b, *Detecting mortgage delinquencies*, IZA Discussion Paper No 5895, July 2011, Institute of Labor Economics.
- Askatas, N. and Zimmermann, K. F., 2011c, *Nowcasting business cycles using toll data*, IZA IZA Discussion Paper No 5522, February 2011, Institute of Labor Economics.
- Askatas, N., 2015. Google search activity data and breaking trends. IZA World of Labor 2015: 206 doi: 10.15185/izawol.206
- Askatas, N., 2018, A Data tax for Digital Economy. IZA World of Labout, October 2018 <https://wol.iza.org/opinions/a-data-tax-for-a-digital-economy>
- Azzellini, D., Greer, I., Umney, C. 2019. *Limits of the Platform Economy: Digitalization and Marketization in Live Music*. Working Paper Forschungsforderung number 154, August 2019, Hans Blockler Stiftung.
- Barzic, G., Rose, M. and Rosemain, M., 2018, 'French officials are going to work at Facebook for 6 months' World Economic Forum (<https://www.weforum.org/agenda/2018/11/france-to-embed-regulators-at-facebook-to-combat-hate-speech/>)
- Bauch, A., Adamczyk, I., Buczek, P., Elmer, F. J., Enimanev, K., Glyzewski, P., Kohler, M., Tomasz Pylak, T., Andreas Quandt, A., Ramakrishnan, C., Beisel, C., Malmström, L., Aebersold, R., Rinn, B. 2011, OpenBIS: a flexible framework for managing and analyzing complex data in biology research, *BMC Bioinformatics* volume 12, Article number: 468 (2011)

- Bensinger, G., 2012, 'In Kozmo.com's failure, lessons for same-day delivery', *Wall Street Journal*, 2 December (<https://blogs.wsj.com/digits/2012/12/03/in-kozmo-coms-failure-lessons-for-same-day-delivery>).
- Bergemann, D. and Bonatti, A., 2018, Markets for information: An introduction. Cowles Foundation discussion paper No. 2142 <http://cowles.yale.edu/>
- Board, S. and Lu, J., 2018, Competitive Information Disclosure in Search Markets, *Journal of Political Economy*, 2018, vol. 126, no. 5.
- Boisot, M. and Canals, A., 2004, 'Data, information and knowledge: have we got it right? *Journal of Evolutionary Economics*, Vol. 14, No 1, pp. 43-67 (DOI: 10.1007/s00191-003-0181-9).
- Bolychevsky, I. and Worthington, S. 2018, Are personal data Stores about to become the next big thing? <https://medium.com/@shevski/are-personal-data-stores-about-to-become-the-next-big-thing-b767295ed842>
- Bond, S. and Bullock, N., 2019, 'Uber IPO prospectus shows ride-hailing revenues stalled', *Financial Times*, 11 April (<https://www.ft.com/content/c68d3662-5c76-11e9-939a-341f5ada9d40>).
- Brustein, J. and Bergen, M., 2019, Google's defence dilemma: the company wants the military business, most of its employees don't, *Bloomberg Businessweek* November 25, 2019, Europe edition, pp. 38-43.
- Brynjolfsson, E., Eggers, F. and Gannamaneni, A., 2018, *Using massive online choice experiments to measure changes in well-being*, NBER Working Paper 24514 (<http://www.nber.org/papers/w24514>).
- Butler, D. (2013), "When Google got flu wrong", *Nature*, Vol. 494 No. 7436, p. 155, available at: https://www.nature.com/news/polopoly_fs/1.12413!/menu/main/topColumns/topLeftColumn/pdf/494155a.pdf
- Cadwalladr, C., 2017, 'Robert Mercer: the big data billionaire waging war on mainstream media', *The Guardian*, 26 February 2017 (<https://www.theguardian.com/politics/2017/feb/26/robert-mercer-breitbart-war-on-media-steve-bannon-donald-trump-nigel Farage>) accessed April 2017.
- Carrie Wong, J. 2019, Will Google get away with grabbing 50m Americans' health records? *The Guardian* Thu 14 nov 2019 <https://www.theguardian.com/technology/2019/nov/14/google-healthcare-data-ascension>
- Chen, L., Mislove, A. and Wilson, C., 2015, *Peeking beneath the hood of Uber*, (DOI: <http://dx.doi.org/10.1145/2815675.2815681>; https://www.ftc.gov/system/files/documents/public_comments/2015/09/00011-97592.pdf).

- Choi, H., and Varian, H.V., 2011, 'Predicting the present with Google Trends', *The Economic Record*, Vol. 88, Special Issue, pp. 2-9.
- Codagnone, C. and Martens, B., 2016, *Scoping the sharing economy: origins, definitions, impact, and regulatory issues*, Institute for Prospective Technological Studies, Digital Economy Working Paper 2016/0, Ispra, Italy.
- Cook, C., Diamond, R., Hall, J., List, J.A., Oyer, P., 2018, The Gender Earnings Gap in the Gig Economy: Evidence from over a Million Rideshare Drivers. June 7, 2018, Stanford Business School Working Paper No. 3637, <https://www.gsb.stanford.edu/faculty-research/working-papers/gender-earnings-gap-gig-economy-evidence-over-million-rideshare>
- Connolly, K., 2016, 'Angela Merkel: Internet search Engines are "distorting perception"', *The Guardian*, 26 October (<https://www.theguardian.com/world/2016/oct/27/angela-merkel-internet-search-engines-are-distorting-our-perception>)
- Cakebread, C., 2017, You are not alone, no one reads terms and conditions of service agreements. Business Insider, Nov. 15, 2017. <https://www.businessinsider.com/deloitte-study-91-percent-agree-terms-of-service-without-reading-2017-11>
- D'Andria, D., 2019, *The unbearable intangibility of the internet: taxing companies in the digital era*, JRC Science for Policy Brief. https://ec.europa.eu/jrc/sites/jrcsh/files/fairness_pb2019_wave03_digitaltaxation_jrc_b2.pdf
- De Hert, P. Papakonstantinou, V., Malgieri, G., Beslay, L. and Sanchez, I. 2018, 'The right to data portability in the GDPR: towards user-centric interoperability of digital services', *Computer Law and Security Review* 2018, pp. 193-203.
- Decker, R. A., Haltiwanger, J., Jarmin, R. S., Miranda, J., 2018, Where has all the skewness gone? The decline in high-growth (young) firms in the USA, NBER Working Paper 21776 <http://www.nber.org/papers/w21776>
- Dosis, A. and Sand-Zantman, W., 2018, *The ownership of data* (https://editorialexpress.com/cgi-bin/conference/download.cgi?db_name=IIOC2019&paper_id=433)
- D'Onfro, J. and Browne, R., 2018 'EU fines Google \$5 billion over Android antitrust abuse' CNBC (<https://www.cnbc.com/2018/07/10/eu-hits-alphabet-google-with-android-antitrust-fine.html>)
- Duch-Brown, N., 2017a, *The competitive landscape of online platforms*, JRC Digital Economy Working Paper 2017-04.

- Duch-Brown, N., 2017b, *Quality discrimination in online multi-sided markets*, JRC Digital Economy Working Paper 2017-06.
- Duch-Brown, N., 2017c, *Platforms to business relations in online platform ecosystems*, JRC Digital Economy Working Paper 2017-07.
- Duch-Brown, N., Martens, B. and Mueller-Langer, F., 2017, *The economics of ownership, access and trade in digital data*, JRC Digital Economy Working Paper 2017-01.
- European Commission, 2016, Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions — Online Platforms and the digital Single Market Opportunities and Challenges for Europe (COM(2016), 288 final, Brussels, 25.5.2016).
- European Commission, 2017, ‘Mergers: Commission fines Facebook €110 million for providing misleading information about WhatsApp takeover’, European Commission press release (http://europa.eu/rapid/press-release_IP-17-1369_en.htm).
- European Commission, 2018a, *Artificial intelligence: A European perspective* (<https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/artificial-intelligence-european-perspective>).
- European Commission, 2018b, Proposal for a Council Directive laying down rules relating to the corporate taxation of a significant digital presence (COM (2018) 147 final, Brussels, 21.3.2018).
- European commission (2020) Communication from the Commission to the European parliament, the Council, The European Economic and Social Committee and the Committee of the Regions. A European strategy for data. Brussels, 19.2.2020 COM(2020) 66 final https://ec.europa.eu/info/sites/info/files/communication-european-strategy-data-19feb2020_en.pdf
- European Union, 2016, Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (OJ L 119, 4.5.2016, p. 1-88)
- Einav, L. and Levin, J. D., 2013, *The data revolution and economic analysis*, NBER Working Paper No 19035, May 2013 (<https://www.nber.org/papers/w19035.pdf>).

- Evans, D. S., 2013, *Economics of vertical restraints for multi-sided platforms*, University of Chicago Institute for Law & Economics Olin Research Paper No 626 (<http://ssrn.com/abstract=2195778>).
- Evgeny, M. 2015, Our cities shouldn't rely on Uber to devise new transport choices. *The Guardian Sun 1 Feb 2015* <https://www.theguardian.com/commentisfree/2015/feb/01/cities-need-to-fight-uber-transport-choice-evgeny-morozov>
- Ezrachi, A. and Stucke, M. E., 2016, The rise of behavioural discrimination, Oxford Legal Studies Research Paper No 54/2016; University of Tennessee Legal Studies Research Paper (<http://dx.doi.org/10.2139/ssrn.2830206>).
- Freedom House, 2019. Freedom on the Net 2019: The crisis in Social Media. https://www.freedomofthenet.org/sites/default/files/2019-11/11042019_Report_FH_FOTN_2019_final_Public_Download.pdf
- Freeman, R. B. 2013, One ring to rule the all? Globalization of knowledge and knowledge creation. *NBER Working Papers* No. 19301 <https://www.nber.org/papers/w19301>
- Galeson, D. W. 2017. Economic History. *Journal of Political Economy*, 2017, vol 125, no. 6
- Ginsberg, G., Mohebbi, M., Patel, R., Brammer, L., Smolinski, M. and Brilliant, L. (2009), "Detecting influenza epidemics using search engine query data", *Nature*, Vol. 457, pp. 1012-1014, doi:10.1038/nature07634, available at: <http://dx.doi.org/10.1038/nature07634>.
- Gold, M. 2019. Taxing Times. *The Economist* Oct 28, 2019. <https://eiperspectives.economist.com/technology-innovation/taxing-times>
- Gramlich, J., 2019, 10 facts about Americans and Facebook, Pew Research Center, Factank News in numbers, May 16 2019 <https://www.pewresearch.org/fact-tank/2019/05/16/facts-about-americans-and-facebook/>
- Grassegger, H. and Krogerus, M. 2017. The Data That Turned the World Upside Down: How Cambridge Analytica used your Facebook data to help the Donald Trump campaign in the 2016 election. https://www.vice.com/en_us/article/mg9vvv/how-our-likes-helped-trump-win
- Hall, J., and Krueger, A., 2015, *An analysis of the labor market for Uber's driver-partners in the United States*, Princeton University Working Paper 587 (<http://dataspace.princeton.edu/jspui/bitstream/88435/dsp010z708z67d/5/587.pdf>).

- Hirsch, L., 2018, 'A year after Amazon announced its acquisition of Whole Foods, here's where we stand', CNBC, 15 June (<https://www.cnbc.com/2018/06/15/a-year-after-amazon-announced-whole-foods-deal-heres-where-we-stand.html>)
- Isaac, M., 2017, 'How Uber deceives the authorities worldwide', *New York Times*, 3 March (<https://www.nytimes.com/2017/03/03/technology/uber-greyball-program-evade-authorities.html>).
- Jarsulic, M., Gurwitz, E., Banh, K., Green, A., 2016, Reviving Antitrust, CTR. FOR AM. PROGRESS (June 29, 2016), <http://www.americanprogress.org/issues/economy/report/2016/06/29/140613/reviving-antitrust>
- Jones, C. I. and Tonetti, C., 2018, *Nonrivalry and the economics of data*, Version 0.6, 31 July (<https://www.gsb.stanford.edu/faculty-research/working-papers/nonrivalry-economics-data>).
- Kaminska, I., 2019, 'The taxi unicorn's new clothes', *Financial Times*, 1 December (<https://ftalphaville.ft.com/2016/12/01/2180647/the-taxi-unicorns-new-clothes>).
- Kee, T. H., 2018, 'Trump has trained stock market investors', *Market Watch* (<https://www.marketwatch.com/story/trump-has-trained-stock-market-investors-2018-07-20>) accessed October 2018.
- Khan, M., and Brunsdon, J., 2018, 'France and Germany abandon plans for EU digital tax', *Financial Times*, 4 December (<https://www.ft.com/content/fc7330d4-f730-11e8-af46-2022a0b02a6c>).
- Khan, L. M. 2017. Amazon's Antitrust Paradox. *The Yale Law Journal*, vol. 126, number 3, pp. 710-805
- Kramer, A. and Kalka, R., 2016, 'How digital disruption changes pricing strategies and price models', in Khare, A., Stewart, B. and Schatz, R. (eds), *Phantom ex machina: digital disruption's role in business model transformation*, Springer, Dordrecht (<https://doi.org/10.1007/978-3-319-44468-0>).
- Kosinski, M., Stillwell, D. and Graepel, T., 2013, 'Private traits and attributes are predictable from digital records of human behaviour', *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 110, No 15, pp. 5802-5805 (<http://www.pnas.org/content/early/2013/03/06/1218772110.full.pdf+html>).
- Krugman, P., 2014, 'Amazon's monopsony is not O.K.', *The New York Times*, 19 October (<https://www.nytimes.com/2014/10/20/opinion/paul-krugman-amazons-monopsony-is-not-ok.html>).

- Lecarpentier, D., Wittenburg, P., Elbers, W., Michelini, A., Kanso, R., Coveney, P., Baxter, R., 2013, EUDAT: A New Cross-Disciplinary Data Infrastructure for Science International journal of digital curation Vol 8 No 1 (2013) DOI: <https://doi.org/10.2218/ijdc.v8i1.260>
- Lomas, N., 2019, Google completes controversial takeover of Deepmind Health 2019, *TechCrunch Sep 19 2019* <https://techcrunch.com/2019/09/19/google-completes-controversial-takeover-of-deepmind-health/>
- Liem, C., and Petropoulos, G., 2016, 'The economic value of personal data for online platforms, firms and consumers', Bruegel blogspot, 14 January (<http://bruegel.org/2016/01/the-economic-value-of-personal-data-for-online-platforms-firms-and-consumers/>).
- Lucas, R.E. 1988, "On the Mechanics of Economic Development", *Journal of Monetary Economics*, Vol. 22, pp. 3-42
- Lyon, D., 2014, Surveillance, Snowden, and Big Data: Capacities, consequences, critique. *Big data and Society*, July-December 2014: 1-13.
- Malhotra, A. and Van Alstyne, M., 2014, 'The dark side of the sharing economy ... and how to lighten it', *Communications of the ACM*, Vol. 57, pp. 24-27.
- Martens, B., 2016, *An economic policy perspective on online platforms*, Institute for Prospective Technical Studies Digital Economy Working Paper 2016/05.
- Martone, M. E., 2011, FORCE11: Building the Future for Research Communications and e-Scholarship *BioScience*, Volume 65, Issue 7, 01 July 2015, Page 635, <https://doi.org/10.1093/biosci/biv095>
- McArdle, M., 2019, 'Uber and Lyft are losing money. At some point, we'll pay for it', *Washington Post*, 5 March (https://www.washingtonpost.com/opinions/uber-and-lyft-are-losing-money-at-some-point-we'll-pay-for-it/2019/03/05/addd607c-3f95-11e9-a0d3-1210e58a94cf_story.html
- Mckinsey, 2018, *Superstars. The dynamics of firms, sectors, and cities leading the global economy*, Mckinsey Global Institute Discussion Paper, October 2018. https://www.mckinsey.com/~media/mckinsey/featured%20insights/innovation/superstars%20the%20dynamics%20of%20firms%20sectors%20and%20cities%20leading%20the%20global%20economy/mgi_superstars_discussion%20paper_oct%202018-v2.ashx
- Mckinsey, 2019, *Twenty-five years of digitization: ten insights into how to play it right*, Briefing note prepared for the Digital Enterprise Show, Madrid, 21-23 May, Mckinsey Global Institute

- (<https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/twenty-five-years-of-digitization-ten-insights-into-how-to-play-it-right>).
- Mercola, J. 2020, Google and Big Tech bough Congress, *Mercola*, Jan 02 2020
<https://articles.mercola.com/sites/articles/archive/2020/01/07/google-big-tech-bought-congress.aspx>
- Mikians, J., Gyarmati, L., Erramilli, V. and Nikolaos Laoutaris, N., 2012, *Detecting price and search discrimination on the Internet* (http://conferences.sigcomm.org/hotnets/2012/papers/hotnets12-final94.pdf?utm_source=Bruegel+Updates&utm_campaign=656e7da39b-Blogs+review+11%2F02%2F2017&utm_medium=email&utm_term=0_eb026b984a-656e7da39b-278510293
- Möhlmann, M. and Zalmanson, L., 2017, 'Hands on the wheel: navigating algorithmic management and Uber drivers' autonomy', in Proceedings of the International Conference on Information Systems (ICIS 2017), 10-13 December, Seoul, South Korea.
- Munich, D., Psacharopoulos, G. (2018) Education Externalities: What are they and what we know. EENEE Analytical report No. 34.
[file:///C:/Users/pedrpab/AppData/Local/Packages/Microsoft.MicrosoftEdge_8wekyb3d8bbwe/TempState/Downloads/EENEE_AR34%20\(1\).pdf](file:///C:/Users/pedrpab/AppData/Local/Packages/Microsoft.MicrosoftEdge_8wekyb3d8bbwe/TempState/Downloads/EENEE_AR34%20(1).pdf)
- NSF, 2017, Leading cloud providers join with NSF to support data science frontiers, National Science Foundation news release 18-009, https://www.nsf.gov/news/news_summ.jsp?cntn_id=244450
- Obar, J. A. and Oeldorf-Hirsch, A. 2018, The biggest lie on the Internet: ignoring the privacy policies and terms of service policies of social networking services. *Information, Communication and Society* pp. 128-147 vol. 23, issue <https://www.tandfonline.com/doi/full/10.1080/1369118X.2018.1486870>
- OECD, 2019, *Digital innovation: seizing policy opportunities*, Organisation for Economic Co-operation and Development <http://www.oecd.org/publications/digital-innovation-a298dc87-en.htm>
- Pawelke, A. and Tatevossian, A. R., 2013, 'Data philanthropy: where are we now?', United Nations Global Pulse, 8 May (<https://www.unglobalpulse.org/data-philanthropy-where-are-we-now>).
- Pedraza, P. de, Visintin, S., Tijdens, K. Kismihok, G. (2019). Survey vs scraped data: comparing time series properties of web and survey vacancy data. *IZA Journal of Labour Economics*. Vol. 8: issue1. <https://doi.org/10.2478/izajole-2019-0004>.

- Porter, E., 2016, With Competition in Tatters, the Rip of Inequality Widens, N.Y. TIMES (July 12, 2016), <http://www.nytimes.com/2016/07/13/business/economy/antitrust-competition-inequality.html>
- Pratley, N., 2018, 'UK finally takes on arrogant tech giants with digital services tax. Budget levy on giants such as Facebook, Google and Amazon could go further — but it's a start', *The Guardian*, 29 October (<https://www.theguardian.com/uk-news/2018/oct/29/uk-digital-services-tax-budget-facebook-google-amazon>).
- Romer, P. M., 1986, Increasing Returns and Long-Run Growth, *Journal of Political Economy*, Vol. 94, No. 5 (Oct., 1986), pp. 1002-1037.
- Starr J, Castro E, Crosas M, Dumontier M, Downs RR, Duerr R, Haak LL, Haendel M, Herman I, Hodson S, Hourclé J, Kratz JE, Lin J, Nielsen LH, Nurnberger A, Proell S, Rauber A, Sacchi S, Smith A, Taylor M, Clark T., 2015, Achieving human and machine accessibility of cited data in scholarly publications, *PeerJ Computer Science* 1:e1 <https://doi.org/10.7717/peerj-cs.1>
- Sandle, P., 2018, Britain to target online giants with new 'Digital Services Tax', Reuters, 29 October. (<https://uk.reuters.com/article/us-britain-budget-digital-tax/britain-to-target-online-giants-with-new-digital-services-tax-idUKKCN1N3265>).
- Schumpeter, J. A., 1942, *Capitalism, Socialism and Democracy*. New York: George Allen & Unwin
- Scott, M. and Young, Z., 2018, 'France and Facebook announce partnership against online hate speech. Emmanuel Macron has teamed up with Mark Zuckerberg to review the country's regulatory response to the issue', Politico, 11 December (<https://www.politico.eu/article/emmanuel-macron-mark-zuckberg-paris-hate-speech-igf/>).
- Scott Morton, F., Bouvier, P., Ezracchi, A., Jullien, B., Kazt, R. Kimmelman, G., Melamed, A. D. and Morgenstern, J., 2019, *Report for the study of digital platforms market structure and anti-trust subcommittee*, 15 May. <https://www.judiciary.senate.gov/imo/media/doc/market-structure-report%20-15-may-2019.pdf>
- Shiller, B. R., 2014, *First-degree price discrimination using big data* (http://benjaminshiller.com/images/First_Degree_PD_Using_Big_Data_Jan_27,_2014.pdf?utm_source=Bruegel+Updates&utm_campaign=656e7da39b-Blogs+review+11%2F02%2F2017&utm_medium=email&utm_term=0_eb026b984a-656e7da39b-278510293

- Steel, E., 2013, 'Companies scramble for consumer data', *Financial Times*, 12 June (<https://www.ft.com/content/f0b6edc0-d342-11e2-b3ff-00144feab7de>) accessed 28 November 2018.
- Steel, E., Locke, C., Cadman, E. and Freese, B., 2013, 'How much is your personal data worth? Use our calculator to check how much multibillion-dollar data broker industry might pay for your personal data', *Financial Times*, 12 June (<https://ig.ft.com/how-much-is-your-personal-data-worth/>) accessed 28 November 2018.
- Steinmetz, S., Slavec, A., Tijdens, K., Reips, U. D., Pedraza, P. de ..., Yavor Markov (2014). WEBDATANET: Innovation and Quality in Web-Based Data Collection. [Special Supplement] *International Journal of Internet Science*, 9, 64-71. https://pure.uva.nl/ws/files/2171902/152617_1406_ijis9_1_supplement.pdf
- Stiglitz, J. P., 2001, 'Information and change in the paradigm in economics', Prize lecture, 8 December (<https://www.nobelprize.org/prizes/economic-sciences/2001/stiglitz/lecture/>).
- Taylor, L., Schroeder R. and Meyer, E., 2014, 'Emerging practices and perspectives on big data analysis in economics: bigger and better or more of the same?', *Big Data & Society* July-December, pp. 1-10.
- Tett, G., 2018, 'Recalculating GDP for the Facebook age. The true impact of social media? Economists are approaching the question from a different angle', *Financial Times* (<https://www.ft.com/content/93ffec82-ed2a-11e8-8180-9cf212677a57>).
- The Economist, 2018, American tech giants are making life tough for startups, June 2nd 2018 edition, <https://www.economist.com/business/2018/06/02/american-tech-giants-are-making-life-tough-for-startups>
- The Economist, 2017, 'Fuel of the future: data is giving rise to a new economy', *The Economist*, 6 May (<https://www.economist.com/briefing/2017/05/06/data-is-giving-rise-to-a-new-economy>).
- Toscano, J. 2019, China will outpace AI capabilities but will it win the race? Not if we care about freedom. Forbes, Dec 3 2019. <https://www.forbes.com/sites/joetoscano1/2019/12/03/china-will-outpace-us-artificial-intelligence-capabilities-but-will-it-win-the-race-not-if-we-care-about-freedom/#32e29da66ed3>
- Uber, 2018, 'Fraudulent trips: how to recognise fraud' (<https://www.uber.com/en-ZA/drive/resources/recognising-fraud/>) accessed January 2019.
- Unctad, 2017, *World investment report 2017, investment and the digital economy*, UN Publications, Geneva.

- UN, 2013, *A new global partnership: eradicate poverty and transform economies through sustainable development*, United Nations.
<https://sustainabledevelopment.un.org/index.php?page=view&type=400&nr=893&menu=1561>.
- UN, 2014, *A world that counts: mobilising the data revolution for sustainable development*, Independent Experts Advisory Group on Data Revolution for Sustainable Development, November 2014, United Nations
<http://www.undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf>.
- UTI, 2018, *Assessing the impact of artificial intelligence*, International Telecommunications Union Issue Paper No 1, September 2018 (<https://www.itu.int/pub/S-GEN-ISSUEPAPER-2018-1>).
- Varian, H. R., 2013, 'Big data: new tricks for econometrics', *Journal of Economic Perspectives*, Vol. 28, No 2, pp. 3-28.
- Vaughan, R. and Hawksorth, J., 2014, *The sharing economy: how will it disrupt your business? Megatrends: the collisions*, PriceWaterhouseCoopers, London.
- Vigo, R., 2013, 'Complexity over uncertainty in generalized representational information theory (GRIT): a structure-sensitive general theory of information', *Information*, Vol. 4, pp. 1-30
[http://cogprints.org/8784/1/Vigo%20\(2013\).pdf](http://cogprints.org/8784/1/Vigo%20(2013).pdf).
- White House, 2015, Big data and differential pricing, White House Council of Economic Advisers
https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/docs/Big_Data_Report_Nonembargo_v2.pdf?utm_source=Bruegel+Updates&utm_campaign=656e7da39b-Blogs+review+11%2F02%2F2017&utm_medium=email&utm_term=0_eb026b984a-656e7da39b-278510293
- Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* , 2016, The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>
- Worstell, T., 2017, 'Understanding the economic value of your personal data'. www.computerweekly.com 26 May 2016
<https://www.computerweekly.com/opinion/Understanding-the-economic-value-of-your-personal-data>
- OECD, 2019 Secretariat Proposal for a "Unified Approach" under Pillar One. <https://www.oecd.org/tax/beps/public-consultation-document-secretariat-proposal-unified-approach-pillar-one.pdf>
- OECD, 2016, The Cancun Ministerial declaration on Digital Economy: Innovation, Growth, and social prosperity ("The Cancun Declaration"). <https://www.oecd.org/internet/Digital-Economy-Ministerial-Declaration-2016.pdf>

Ursu, R., 2015, *The power of rankings: quantifying the effect of rankings on online consumer search and purchase decisions*. Accessed 26 September 2018.

<https://pdfs.semanticscholar.org/b833/9f0aa1d919fdd7b58a533174eaf463645403.pdf>

WEF, 2019, Global Technology Governance A Multistakeholder Approach. World Economic Forum White Paper.

<https://www.weforum.org/whitepapers/global-technology-governance-a-multistakeholder-approach>

Wilkinson et al. 2016 Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>