LUISS Guido Carli

LIBERA UNIVERSITÀ INTERNAZIONALE DEGLI STUDI SOCIALI

Master in Management with major in Innovation and Entrepreneurship

Subject: Markets, Regulations and Law

Data Economy: Moving from Digital Feudalism toward a Digital Capitalistic Model

SUPERVISOR

Professor Luca Arnaudo

STUDENT Matr. 683461

Pierfrancesco Marano

CO-SUPERVISOR

Professor Massimiliano Granieri

ACADEMIC YEAR 2018/19

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Data economy and data-driven innovation (DDI) are key pillars in the economy of the 21st century. The increasing digitalization of every kind of socio-economic activities and the decline in the cost of data collection, storage and analysis, has led to a new kind of economic paradigm, centered on the use of huge volumes of data, the so-called Big Data. Data are considered as the new world's most valuable resource, the oil of the new century and the engine of the digital economy. The innovative characteristics of data are conventionally summarized in four Vs: (1) volume, the quantity of data created every year is impressive and it is exponentially growing year over year; (2) velocity, data are collected and analyzed quickly, even in real-time; (3) variety, as we will see, there are many different types of data, and (4) value, data are becoming a valuable asset for firms and a new factor of production. We can talk of a data revolution, or a new Schumpeterian wave, that is fostering new industries, processes and products, while replacing the old ones. However, as always happens when there are such radical changes in the economy, there are also challenges to be addressed. The capability of private firms to collect and analyze huge amount of personal data, indeed, has created a wide array of policy issues, ranging from privacy and consumer protection to competition concerns. Particularly, governments will need to anticipate and address the disruptive effects of Big Data on the economy and overall well-being. For the moment, how we will see, they have not been capable of keeping pace with the technological challenges of this new economy.

In this work, we are going to analyze the data economy from different point of views, with a particular focus on its critical aspects, and eventually, on how they may be addressed. In the first chapter (Introduction to data economy), we will analyze more in details the intrinsic characteristics of data with some numbers, trends and statistic of the global data economy. Afterwards, we will see the legal implications and the social impact of data. In the second chapter (European Union Data Economy), we will narrow the focus on the data economy in Europe, and on the projects of the European Union to exploit this new technological revolution through the analysis of the Digital Single Market Strategy and the European Data Market Monitoring Tool. Eventually, we will further narrow the focus analyzing the numbers and the trend of the Italian data economy. The third chapter (The value of data), focuses on how data's economic value can be computed, through the analysis of the Data Value Chain. Then, we will

define data-driven innovation and eventually, we will try to understand where the actual value of data lies. In the fourth (Privacy) and the fifth (Competition) chapters the focus switch on the critical aspects of the data-driven innovation disruption, and the policy issues it has brought. Specifically, in the fourth we will analyze the privacy implications of the mass surveillance business model that dominates the new economic landscape. We will analyze the current privacy law in Europe, and the new Directive, the General Data Protection Regulation (GDPR) issued in May 2018. The fifth chapter focuses on the competition harms of data economy, and the challenges competition authorities must overcome to regulate the new industries arisen from the data revolution. In the final paragraphs, three case studies are analyzed in order to better understand how competition authorities are trying to figure out how to avoid big tech giants to act anti-competitively. Eventually, the sixth and last chapter (Digital Capitalism) explains how the data economy can become more efficient and socially fair, with a revolutionary approach supported by always more scholars and experts of the data economy.

1. INTRODUCTION TO THE DATA ECONOMY

1.1 Data as oil

"*The world's most valuable resource is no longer oil, but data"*
(The Economist)

Data in the twenty-first century are the equivalent of what was oil in the twentieth century. As an oil refinery, also data centers fulfill the same role: produce raw materials that fuel the entire economy. Indeed, the refining of data powers all kind of online services. As oil, data are extracted, stored (in this case in *data warehouse*), refined, valued, bought and sold in different ways. Particular importance has the refining process, i.e. the so called *data mining* process, in which happen the discovery of patterns[1] and relations inside the data sets. It is a crucial process, since the quality of data is much more important for its deployment than the quantity of them. The data economy is a recent phenomenon, but it is disrupting every sectors and industry, that indeed we can talk of a data revolution.
Nevertheless, the parallelism between data and oil is not as accurate as it may seem, because there are also several and important dissimilarities:

- there is a fixed amount of oil in the world, and the value of oil derives from its scarcity and from the difficulty of extracting it. Data, on the other hand, can be created infinitely and easily.
- oil is a single use commodity, data instead can be used and reused for new purposes and insights.
- oil increase its value merely accumulating it, and the more oil you can accumulate the more is its value. Conversely, the more data collected the greater the complexity and the costs to maintain proper protection and to extract value from them. Indeed, the value extracted from data comes from the insights generated through analytics and combinations of different data sets, and data needs to be collected and maintained in a usable and accessible format in order to achieve economic benefits.

[1] A pattern is a structure, a model, or in general, a concise representation of a given data set.

- due to its scarcity oil is stored and locked behind borders; on the contrary, data transfers are an essential component of creating worldwide benefits. Restricting data behind borders would mean the reduction of both economic and social benefits. To exemplify this concept quantitatively, the movement of data across borders generates yearly global economic gains equivalent to the GDP of France.

Thus, it is clear that when comparing these two commodities it is also important to have in mind the above mentioned differences. However, the important concept here is not how much oil and data are equivalent to each other, "data is the new oil" mantra wants to emphasize the broader economical shift of the twenty-first century that had caused the rise of data as the most important resource to the detriment of oil, so that data can be considered undoubtedly the oil of the digital economy.

1.2 Data economic revolution

The digital universe is expanding at a tremendous pace, an expansion that include not only people and enterprises online, but also all the smart devices connected to the internet. This ever-expanding universe is doubling in size every two years[2], and by 2020 the amount of data circulating in this universe will reach 44 zettabytes (or 44 trillion gigabytes), the world produces 2.5 quintillion bytes a day, and 90% of all data has been produced in just the last two years[3].

Unlike the physical universe, the digital one is created and defined by a human-created machine: software. Software development affected all sectors of the economy, since they allow the arrival of new actors able to reduce fixed costs while providing cheaper and better service: Amazon in the retail market, Google in marketing and advertising, music with Spotify, etc. All these innovations need an ever-increasing quantity of data, and these new platforms gather and reprocess personal data from users in order to provide a more efficient service. As everything

[2] The digital universe of opportunities: rich data and the increasing value of the IoT; IDC (International Data Corporation). Access at: https://www.emc.com/leadership/digital-universe/2014iview/index.htm (2014).

[3] T. Vasudha, G. Arvind, *The value of data*, World Economic Forum (2017)

is going online, there are always new ways of exploiting data, the so-called *data-driven innovation*, defined by the OECD as "the use of data and analytics to improve or foster new products, processes, organizational methods and markets". This kind of innovation has allowed the transformation of data into assets, allowing enterprises to learn deeply the customer's behavior, to speed the business cycle, to flatten the organizational structure and broadly to make better predictions and smarter decisions. However, the positive impacts go beyond these business oriented goals, since it can also have a social impact contributing to inclusiveness and the well-being of citizens. In fact, the OECD identified education, healthcare services and public administration as the sectors where data-driven innovation can lead to the highest impact. As a matter of fact, data do already have a huge impact on the output of the world economy, the Mckinsey Global Institute estimates that data flows have boosted the world GDP by more than 10%[4]. By the end of 2019, Digital Transformation spending will reach 1.7 trillion dollars worldwide, a 42% increase from 2017. Despite this already astonishing results, there is also a new revolution coming from the data driven innovation, the deployment of data extracted from embedded systems, i.e. the advent of the Internet of Things (IoT). This kind of data accounted in 2013 only for the 2% of the total data in the digital universe, and they will account for 10% in 2020. The era of the IoT involves the computerization of things, adding software and intelligence to them, and it requires Artificial Intelligence (AI) in order to work properly.

Internet companies' control of data has given them an enormous amount of power. However, the abnormal bulk of data circulating in our economy is exploitable only for the corporations capable of managing this large amount of data, and the best in exploiting them are now the companies that are dominating the world economy. Therefore, it is also useful to analyze firstly the percentage of the total amount of data circulating in the digital universe that can be used -i.e. tagged and analyzed-, and secondly how much of these useful data are actually analyzed. According to the 2014 IDC's study, in 2013 the amount of useful data was only 22% of the total, and of these only 5% were valuable (also called "target-rich data"). However, as firms take advantage and get expertise on how to exploit Big data and analytics technologies, these percentages are growing rapidly, and the 22% of useful data in 2013 will become a 35% in 2020, while the *target-rich data* will reach a 10% of the total amount. Here we can see again a

[4]G. Cattaneo, G. Micheletti, D. Osimo, K. Jakimowicz, *First report on policy conclusions*, IDC Italia (2018).

clear parallelism between data and oil, since also world's oil reserves are exploited in a smaller percentage of the total amount, according to the difficulty and economic viability of extracting it.

The corporations that had already exploited the potential of Big data at its best in the last decade, are nowadays the most valuable firms in the world. They are platforms-companies and data-aggregators: Alphabet, Facebook, Amazon and Microsoft, each one of them capitalize on individual data. To reconsider the parallelism between data and oil, if we take the same list of just 12 years ago, in 2006, it was dominated (except for Microsoft) by oil and energy companies: Exxon Mobil, General Electric and Shell Oil. These platforms-giants are now running to become the first trillion dollars company (as market cap) in the history, Apple with its market capitalization of more than 900 billion dollar (as of May 2018) is leading this competition. Moreover, they had revolutionized the classic economic accretion, because while the old monopolies create and shares economic benefits for many, the new tech giants are generating significant value for billions of people, that can benefit of free services, but very few reap the economic benefits. If, indeed, we compare the market cap:workforce ratio (market cap per employee) of General Motors with the one of Facebook, this situation emerges clearly: in 2016 GM created economic value of approximately 231 thousands dollar per employee, while the Zuckeberg's platform generated 20.5 million dollar per employee, i.e. almost a hundred times the value of GM[5]. If instead we consider Amazon, the Earth's biggest store, it is worth more than its biggest "competitors" (Walmart, Target, Macy's, Kroger, Nordstrom, Tiffany & Co., Coach, Williams-Sonona, Tesco, Ikea, Carrefour and The Gap) all combined, with a market capitalization of more than 700 billion dollars (as of May 2018).

This is the disruptive power of Big data, and particularly the power of the *data-network effect:* using data to attract more users, who then generate more data, which help to improve services, which in conclusion attract more users.

Another important issue to be considered is the information security and privacy, since many of the generated data needs some level of protection, from privacy protection to full-encryption lockdown. According to the IDC's study, in 2014 the portion of data not needing protection, like for example camera phone photos, public website content or open source data, accounted

[5] S. Galloway, *The Four - The hidden DNA of Amazon, Apple, Google and Faceboo*k, Bantam Press (2017).

for a 57% of the total. The portion sensible to different level of protection, like corporate financial data or personally identifiable information or medical records, accounted for the remaining 43%; of these only 48% were actually protected. Thus, it is clear that information security will need to be improved and tightened. In this direction the European Union has already made the first move with the new General Data Protection Regulation (GDPR), that impose a stricter regulation and a higher transparency for the protection of the privacy for those companies who handle them. However, the burden and complexity of managing data at an enterprise level will become heavier and heavier with the years, considering that the number of IT professional figure within the companies cannot keep up with the evolution of the digital universe. According to the IDC's study, the number of gigabytes per IT professional will grow by a factor of 8 until 2020: in 2014 the worldwide workforce of IT pros was 28 million, each of them handling on average 230 gigabytes of data, in 2020 the workforce will be of 36 million with a burden on each one of them of 1231 gigabytes.
The digital universe will continue to expand and flood already saturated infrastructures and the people that manage them; in addition, the IoT will add new levels of complexity, but also of opportunity. The enterprise that will succeed in the future should solve foremost organizational challenges, more than technical ones. The IDC's study has identified 3 organizational steps enterprises should follow in order to succeed:

1. Create a C-level position in charge of developing new business opportunities
2. Develop an executive team focused on understanding the new digital landscape: its risks, opportunities and the short and long-term steps to take in order to ensure an efficient digital transformation
3. Re-allocate resources across the business based on digital transformation priorities, identifying the gaps in talent and skills required.

Therefore, the big data challenge is firstly a management revolution, since as the tools and philosophies of big data emerge, there will be changes about the value of experience in the process of decision making. Many business executives are still figuring out the differences between big data and analytics. Basically, they are related, since they both seeks to deduce insights from data and translate that into business advantage; but big data differentiates from the analytics of the past for 3 key characteristics: volume, velocity and variety. The most critical

consequence of this managerial revolution is its impact on the decision-making culture of the firms: a shift from the classic way of relying on intuition and experience of the senior executives (particularly on "HiPPO" - the highest-paid person's opinion), to a data-driven approach that firstly ask "What do the data say?" when faced with an important decision. Nevertheless, big data's power does not eliminate completely the need for vision or human insight, but it will change the role of leadership: instead of finding the right answers it will be crucial for senior executives to ask the right questions. Enterprises will still need business leader that set clear goals, spot future opportunities, understand how the market in which he operates is developing, think creatively, articulate a compelling vision and persuade people to embrace it and work hard to achieve it. All of this cannot be handled alone by big data, no matter how many and how well structured they are, data will be an important tool that will always need a human-mind's control.

The evidence that data-driven companies perform better, has been shown by a study of Harvard, together with the MIT Center for Digital Business and in partnership with McKinsey's business technology office in 2012[6]. This study conducted interviews with executives of 330 public North American companies about their organizational and technology management practices, and gathered performance data from their annual reports. The result was clear: the more the companies characterized themselves as data-driven, the better they performed on objective measures of financial and operational results; in particular companies in the top third of their industry in the use of data-driven decisions making were, on average, 5% more productive and 6% more profitable than their competitors, moreover, this difference was reflected in measurable increase in stock market valuations. Thus, the evidence of this study is that data-driven decisions tend to be better decisions, and leaders should embrace this new revolution in order to survive in the digital universe.

1.3 Characteristics and classification of data

Nowadays, each of us can be defined as a *walking data generator*, thanks to all sorts of devices that are becoming sources of data, from cars and airplane to toys and toilet seats, and obviously all of our mobile and wearable digital devices. Practically, the things around us will become

[6] E. Brynjolfsson, A. McAfee, *Big Data: The Management Revolution*, Harvard Business Review (2012).

the eyes and ears of the Internet, and the privacy implications of all this connectivity are profound. The majority of this data is a type of data called *metadata*, they are data about data, information a computer system uses to operate or data that are a by-product of that operation. For example, in a text message system, the messages themselves are data, but the accounts that sent and received the message, the date and time of the message, are all metadata. Data is content and metadata is context, and the latter can be much more revealing than data, especially when collected in aggregate.

Analyzing the economic characteristics, data can be classified as non-rivalrous, excludable and intangible good. The non-rivalry nature implies that data should be used as much as possible, since the marginal costs of an additional use of them is zero. The second characteristic, the excludability, is an important precondition for ensuring the possibility of revenues for the data holders, since it allows them to exclude others from using their data. Data is a capital with increasing returns, since it can be re-used as input for further production. Eventually data is a general purpose input with no intrinsic value, indeed, they can be an input for multiple purposes and its value depends on complementary factors related to the capacity to extract information.

Another important distinction is between primary data and generated data:

- Primary data are produced by single individuals, and they include identity data (last name, first name, home etc.), and their sensitive data, such as sexual orientation or religious affiliation. They are any stored information, recorded on a digital device, connected to a person and allowing this person to be identifies. These data are the heath of the data economy.
- Generated data, instead, are those data collected by entities for a lucrative purposes or not, through cookies or trackers. They include consumption data and financial data.

Analyzing more deeply the intrinsic characteristics of data, a fundamental classification of data is about personal and non-personal data:

- Personal data: any information relating to an identified or identifiable living individual (data subject). Personal data that has been rendered anonymous in such a way that the

individual is not or no longer identifiable, is no longer considered personal data, but the anonymization must be irreversible. Examples are: name and surname, home address, email address, Internet Protocol (IP) address, a cookie ID or an identification card number.

- Non-personal data: conversely they are all the others, in particular they refer to machine-generated data, especially raw data created through sensors.

1.3.1 Legal implications

The distinction between personal and non-personal data has important implication from a legal perspective. In the European Union there are two separate and different legislations respectively. The protection of personal data has been elevated to a fundamental right of every European citizen, since it has been disconnected from the right to privacy. In fact, it is ratified in the article 8, Title III (Freedoms) of the *Charter of Fundamental Rights of the European Union* (the Charter), while the right to privacy is enunciated in the Article 7, and also in the Article 16 of the *Treaty of the Functioning of the European Union* (TFEU).
Article 8 of the Charter states:

1. Everyone has the right to the protection of personal data concerning him or her.
2.Such data must be processed fairly for specified purposes and on the basis of the consent of the person concerned [...] Everyone has the right of access to data which has been collected concerning him or her, and the right to have it rectified.
3.Compliance with these rules shall be subject to control by an independent authority

According to the third point, EU countries have set up national bodies responsible for protecting personal data. For example, in Italy there is the *Garante per la protezione dei dati personali.* At a European level, instead, the GDPR established the new *European Data Protection Supervisor* (EDPS), an independent EU body responsible for monitoring the application of data protection rules within European institutions, for investigating complaints and with extensive powers to determine disputes between national supervisory authorities. Furthermore, on 24 May 2016 the Regulation (EU) 2016/679 came into force, and it becomes effective on the 25 of May 2018. It is a new regulation for all the Member States, better known with the name of

General Data Protection Regulation (GDPR). It is a regulation on the protection of natural persons with regard to the processing of personal data and on the free movement of such data. The main purpose is the strengthening of the fundamental rights of the European citizens in the digital age, but also to simplify rules for business in the digital single market, since it uniforms all the different legislations in a single regulatory body.

As concerns the non-personal data, there is less brightness in term of legislations, since the European Commission are still working on finding the optimal solution for them. Until now, the Commission has proposed a new Regulation aimed at removing obstacles to the free movement of non-personal data in an official EU Communication[7], that together with the GDPR will ensure a comprehensive and coherent approach to the free movement of data in the EU. The main criticality to solve about the non-personal machine-generated data is that many of them do not have property rights as a protection, thus, most holders do not make their data accessible but utilize them only in-house. Currently, there is an ongoing academic discussion about proposals for the creation of a new sui-generis intellectual property rights (IPRs) for non-personal data. However, the discussions is complex, since to justify such a measure it is firstly necessary to prove the existence of market failures. Principally, it should be assessed whether the data economy suffers from too low incentives for producing data and whether therefore IPRs are needed to solve this incentive problem. The final economic goals of the EU Communication are mainly the facilitation of access to and sharing of machine-generated data, the assurance of a fair sharing of benefits between data holders, processors and application providers and the protection of investments in data and innovation.
This considerable concern of the European legislators about the data economy, testify the importance that this new resource has in the creation of new economic benefits for the European economy. Data are the driver of growth and change, shaping the new digital data-driven economy, that has already caused a generation of new infrastructures, new businesses, new monopolies and also new politics. All of the new technology waves have a common denominator: extracting value from data.

[7] Communication from the Commission to the European Parliament, *Building a European Data Economy* (2017). Access at: https://ec.europa.eu/digital-single-market/en/policies/building-european-data-economy.

1.4 Social role of data

Data, as we already said, play also an important role in the society that goes beyond productivity growth, as they can also directly contribute to inclusiveness, development and the well-being of citizens. The key issue is that a society with a high level of data innovation may still be unequal, with a few actors concentrating most of the power and capturing most the benefits. This issue clarifies the crucial role of policies in pursuing a balanced and participative data-driven society, where individuals maintain control of their own data (one of the principle of the GDPR).

Data have also an impact in the area of science. In fact, the ability to access and combine data coming from different sources, along with the increasing power of analytics, have allowed researchers to conduct complex experiments and potentially achieve scientific discoveries and better predictions leveraging massive data sets and algorithms, especially in the fields of medical, social and environmental sciences.

From a healthcare perspective, the collection and analysis of data can have an extremely disruptive impact. Patient data tracking could provide useful information concerning patient characteristics, illnesses, medications and therapies, which can be exploited to identify which paths to follow to achieve cost-effective outcomes, finally improving healthcare quality and performance. Data tracking can also improve preventive and emergency care, which are also benefiting from an enormous amount of health and lifestyle data collected by sensors and connected devices within modern care institutions, smart homes and smart living environments. These technologies have proven to be particularly helpful to elderly people, who can be alerted when it is the time to take a medicine or receive suggestions to maintain a healthy lifestyle.

Moreover, the full use of data analytics in the public sector could lead to improved operational efficiency and tax collection, and to lower corruption cases, thanks to a greater transparency. In addition, open access to public sector data could enhance government accountability, thus, fostering democracy and citizens' engagement, and at the same time big-data could improve government decision-making in different domains.

Smart cities are another context in which data-driven innovation is crucial, for example data collected from connected cars can be leveraged to provide real time information on traffic and lead to traffic management optimization. The Intelligent Traffic Management System of

London, for instance, is using smart technology to tackle with congestion and is becoming increasingly able to predict traffic and guide flows.

Finally, the data-driven innovation can create new job opportunities for citizens, but at the same time it is destroying old ones, and transforming others. While technologies such as robotics and machine learning may cause massive unemployment levels, other digital technologies may help create new opportunities and more flexible jobs, allowing citizens to become independent workers. Even though, this new kind of flexible/independent work life lead to other questions, such as dissatisfaction with income variability and lack of the traditional benefits associated with a stable job position.

Thus, not also it is necessary to bridge the digital gap for workers and teaching them the necessary skills to survive in the digital revolution, but it is needed a welfare plan to avoid dangerous social disruption. Furthermore, a deep reform of the school systems is required, focusing in particular on Science, Technology, Engineering and Math (STEM) skills, to prepare also the next generation.

We are in the first phase of this revolution, in which we are only starting to understand the challenge and threats presented by the data economy. There are multiple ways in which society could evolve in the next years, depending on how it faces the opportunities and threats. To better understand the potential implications of some of the policy and economic decisions that could be taken in the next years, we can hypothesize some future scenarios, one optimistic and another pessimistic:

- Power to citizens: individuals have full control of their personal data, GDPR successfully increases confidence in data, and this eventually lead businesses to get more value from their data products and services enabling a virtuous growth circle.

- Data Elysium and Big Brother 2.0: as in the movie (Elysium), economic and social inequality will spread and they will lead to a divided scenario where there is an elite with all the privilege and enjoy all the benefits of technology innovation, while the majority of the population lives in misery, working low level jobs with low salaries. In this dark scenario, the power and control are concentrated in the hands of a few dominant global companies manipulating the population through their data.

2. EUROPEAN UNION DATA ECONOMY

Narrowing the focus of the data economy only on the European Union data market, we will see the opportunities and threat for the European Member States. Among the seven pillars of the Europe 2020 Strategy, which sets objectives for the growth of the European Union by 2020, there is the European Commission's Digital Agenda. The main goal of this Agenda is to develop a digital single market, in order to generate smart, sustainable and inclusive growth in Europe. The Digital Agenda is made up of six pillars:

1. Achieving the Digital Single Market (DSM)
2. Enhancing the interoperability of devices, applications, services and networks
3. Strengthening online trust and security, in order to combat cybercrime, child pornography and privacy violation
4. Promoting fast and ultra-fast Internet access for all
5. Investing in ICT research and innovation, in order to boost growth and jobs via innovative public-private partnerships
6. Promoting digital literacy and skills, promoting employment in the ITC field[8]

In the next paragraphs we will focus mainly on the first pillar, the Digital Single Market Strategy.

2.1 Digital Single Market (DSM) Strategy

The digital revolution offers the opportunity to strengthen Europe's economy and society which have suffered, and it is still suffering, from a decade of low growth. However, this revolution needs also to be incentivized by the right legislations and policy. The McKinsey Global institute estimates that Europe has captured so far only 12% of its potential from digital technologies and it is lagging behind the United States, even though the European market has the potential to be the largest digital market in the world in size and value, only if appropriate policy and investment decisions will be made. For this reason, the European policymakers are working in

[8] European Commission Communication to the European Parliament, *A New Skills Agenda for Europe - Working together to Strengthen Human Capital, Employability and Competitiveness* (2016).

order to establish the best condition for the prospering of the digital opportunities in the European market. Indeed, with the aim of maximize the growth of Europe data economy, the European Commission has launched a strategy that goes under the name of Digital Single Market (DSM) Strategy. This strategy denotes the willingness of the European Commission to maximize the positive impacts of the digital revolution on people's lives and European business activities, ensuring access to online activities for individuals and businesses under conditions of fair competition, consumer and data protection, removing geo-blocking and copyright issues. The completion of this strategy, firstly formulated on the 6 May 2015, has been identified by the Commission as one of its 10 political priorities. Indeed, a digital single market creates opportunities for new startups and allows existing companies to reach a market of over 500 million people, and it can contribute 415 billion euro per year to Europe's economy. Consequently, this process will lead to the creation of jobs, and more broadly to the improvement of the wellbeing of the European citizens. Analyzing more in details the DSM Strategy, it has identified three main emerging challenges: (1) to ensure online platforms can continue to bring benefits to our economy, (2) to develop the European Data Economy to its full potential and (3) the protection of Europe's asset through cybersecurity. Moreover, it is built on three main pillars, and each of them has its own policy proposals:

Pillar I (Access): Better access for consumers and business to digital goods and services across Europe. In order to achieve this the Commission aims to propose:

- rules to make cross-border e-commerce easier, harmonizing EU rules on contracts and consumer protection
- the end of unjustified geo-blocking. It is the discriminatory practice used for commercial reasons, when online sellers deny consumers access to a website based on their location
- an antitrust competition inquiry into the e-commerce sector
- reduce the administrative burden businesses face from different VAT regimes, with a single electronic registration and payment, and a common VAT threshold to help start-ups selling more

Pillar II (Environment): creating the right conditions and a level playing field for digital networks and innovative services across Europe. For this goal the Commission has already taken several initiatives:

- the cross-border portability of online contents services (films, sporting broadcasts, music, e-books or games) to which they have subscribed in their country
- the abolition of mobile roaming charges in June 2017
- the new single set of EU rules on data protection and privacy (GDPR) and the first ever common cybersecurity law to help keep network and information systems safe in all Member States, both in May of this year

Furthermore, the Commission is working on a comprehensive analysis of the online platforms (search engines, social media, app stores etc.) about how they use the information they acquire and the promotion of their own services to the disadvantage of competitors.

Pillar III (Economy & Society): maximizing the growth potential of the digital economy. The Commission aim to:

- create a European free flow of data to promote the free movement of data, since new services are obstructed by restrictions on where data is located, that have nothing to do with protecting personal data
- define priorities for standards and interoperability in area critical to the DSM, such as e-health, transport planning and energy
- create an e-government action plan to connect business registers across Europe, in order to ensure that businesses and citizens only have to communicate their data once to public administrations

Overall, since May 2015, the European Commission has delivered 35 legislative proposals and policy initiatives to pursue the Digital Single Market Strategy.

2.2 European Data Market (EDM) Study Monitoring Tool

In order to pursue the DSM strategy efficiently, market intelligence is necessary to guide strategic decisions. For this reason, there is the European Data Market (EDM) Study Monitoring

Tool: an instrument designed by IDC in collaboration with the Lisbon Council to provide the European Commission with a comprehensive view of the data-driven economy through annual reports. According to the EDM study[9], the data workers in 2016 were 6.16 million and they will reach 10.43 million by 2020, the data companies were 255 thousand in 2016 and they will be 359 thousand by 2020 and the data economy value as a whole was almost 300 billion euro in 2016, and it will be 739 billion euro in 2020.

More than these simple numbers, the EDM Monitoring Tool provides a perspective of the data ecosystem in Europe, analyzing the role of policies in shaping the future scenarios of the data market. In order to do so, it utilizes six interrelated indicators:

1. Data professionals: workers who collect, store, manage, analyze, interpret and visualize data as their primary or as a relevant part of their activity

2. Data companies: they can be *data suppliers*, when their main activity is the production and delivery of digital data-related products, services and technologies; or *data user*s, which are organizations that generate, exploit, collect and analyze digital data intensively and use what they learn to improve their business

3. Data's supplier revenue: the aggregated value of all the data-related products and services generated by EU data supplier companies

4. Value of the EU data market: the marketplace where digital data is exchanged as products or services as a result of the elaboration of raw data

5. Value of the EU data economy: the overall impacts of the data market on the economy as a whole

6. Data skills gap: the potential gap between demand and supply of data skills in Europe

[9] *How the power of data will drive EU economy;* The European Data Market (EDM) Monitoring Tool Report (2018). Access at: http://datalandscape.eu/webinars/how-power-data-can-drive-european-union-economy-european-data-market-study-and-monitoring

These indicators can be grouped in four different blocks:

1. The *Data Professionals Block*: composed of data professionals and data skills gap

2. The *Data Market Block*: composed of the value of data market, the number of data companies and the value of data revenue

3. The *Data Companies Block*: composed of the number of data companies and of data users

4. The *Data Economy Block*: composed of the value of data revenues and of the data economy

The Report of 2017 highlights that the pace of growth of the European data economy accelerated compared to the previous years, thanks also to the more favorable economic climate. The number of data suppliers and data users increased, their demand of data services and products drove up the value of the data market, that has reached the valuation of 335.5 billion euro with an increase of 11.8% compared to 2016, corresponding to 2.4% of the EU28 GDP. Also the number of data professionals has increased of 8% from 2016, reaching 6.7 million professionals employed, with a share on total EU employment of 3.2%. A key benefit to take into consideration is the jobs creation of the innovative markets; enterprises are employing a wide range of skills combining mathematics and analytics, business intelligence and ICT. However, the increasing demand for innovative data skills is not fully satisfied by the market of workers, and this is exactly what the indicator 6 of the EDM Monitoring Tool control. The gap increased from 428 thousand excess demand to 449 thousand in 2017, corresponding to a 6% of the total skills demand in EU28. At this level, the skills gap is not worrying, since a vacancies ratio of this extent is still manageable; what is worrying is the persistence of the gap and the increasing trend for the future years.

Finally, the EDM Monitoring Tool measures a more limited set of indicators for three other international economies: the United States, Brazil and Japan. In particular, it looks more closely at the U.S., examining the competitiveness implications. The latest estimates confirm an overall

development of the six indicators, and that the European data economy is the second largest after the US and well ahead of Brazil and Japan. Moreover, Europe is catching up on the gap with the US year over year, in fact, it has a yearly growth 2016-2017 of 9.2%, more than the double of the US and three times stronger than in Japan. However, with respect to the US, Europe has a higher level of fragmentation when it comes to the usage of digital technologies across companies, because of the higher presence of SMEs, and to cultural and educational factors. It is also for these reasons, that Europe lacks of successful companies as Amazon, Google, Facebook and Apple. European companies are less able to capitalize and exploit the enormous amount of data on which they sit on, with few notable exceptions, such as Spotify in Sweden, Shazam in the UK or Yoox in Italy.

The final picture emerging from the Report is mixed: Europe still lags behind the United States in terms of size, but it is growing at a faster pace. Europe's position will not improve until the Digital Single Market Strategy will become effective throughout all the Member State, removing the barriers that still impede the free flow of data within the EU both at cross-border and at intra-company level. European companies should be incentivized to open up and share their data with adequate policies. In addition, Europe market need big investments in the improvement of digital infrastructure and digital skills

2.3 Italian data economy's legislation and numbers

The Italian legislations about the protection of personal data is based on the Code for the protection of personal data (*Codice in materia di protezione dei dati personali)*, known as the Code of Privacy (*Codice della privacy*), released with the Legislative Decree n.196, the 30 June 2003 and entered in force from 2004.
This decree was the Italian response to the European Union Directive 95/46/CE, the guideline directive for all the member States' legislation about the treatment of personal data. However, the European Commission acknowledging the technological and social mutation occurred in the past twenty years, released the new EU Regulation n.679/2016, the already mentioned General Data Protection Regulation (GDPR), "on the protection of natural persons with regard to the processing of personal data and on the free movement of such data", with a new philosophy based on the principle of accountability. Therefore, the new Regulation has

substituted the old directive of the 1995, and being a EU Regulation it is binding in its entirety and directly applicable in all the Member States. Nevertheless, many Member States, including Italy, need a legislative intervention, in order to harmonize the Regulation with the national law. For this purpose, it was nominated a Commission in January of this year, with the aim of ratify a Legislative Decree that harmonize the Italian law with the new European GDPR. The harmonization is necessary for two reasons:

1. the abrogation of old Italian norms in contrast with the EU Regulation

2. the definition of new norms in topics that the Regulation allows Member States to define autonomously, even if always in respect of the principles and norms of the Regulation itself

As regards the number of the Italian data economy, the already mentioned EDM Monitoring Tool, gives us some statistics that can give an idea of the value of this economy in Italy. In 2016, the Italian Data Market Value was of 4.6 billion euro, expected to reach in 2020 a value of 5 to 9 billion euro (from Baseline scenario to High Growth scenario), while the value of the Italian Data economy was 28 billion euro in 2016. The Italian market share compared with the overall European economy was about 8% in 2016, and it is expected to grow to a 9% in the best-case scenario. However, the impacts of the data economy on the overall economy is only less than 2%, highlighting that much has yet to be done in Italy to exploit the full potential of this new economic revolution.

3. THE VALUE OF DATA

As already mentioned in the first chapter, the three V that characterized big data are velocity, volume and variety. In addition, the further development of the use of big data has led to a fourth V: value. However, despite the increasing value of data for the competitive success of Internet firms, data are difficult to price, because they are not fungible. Therefore, their economic value is extremely difficult to estimate.

3.1 Data Value Chain

The economic and social value of data depends on many different factors: the veracity of data (i.e. the authenticity and trustworthiness), the type of information they bring, the structure of data (structured or unstructured) and also the way the data are collected. To all effects, data have become a new factor of production[10], in the same way as hard assets and human capital. Thus, having the right technological basis and organization structure to exploit data has become essential for public and private corporations. The value of data can be analyzed through a value chain, more precisely a Data Value Chain, that describe the flow of information through a series of steps, necessary for the economic valorization of the data sets. Already in 2013, the European Commission saw the data value chain as the "center of the future knowledge economy, bringing the opportunities of the digital developments to the more traditional sectors". The firsts to apply the value chain metaphor to information systems were Jeffrey F. Rayport and John Sviokla in their work *Exploiting the Virtual Value Chain,* on the Harvard Business Review in 1995. They analyzed how business competes in two different worlds: a physical world made of resources, and a virtual one made of information, distinguishing the former as the *marketplace*, and the latter as the *marketspace*. The processes to create value in the two different worlds are not the same, and they noticed that those who understand how to master both the processes can create and extract value in the most efficient and effective manner. In that period, instead, the value chain was conceived as a management decision support tool, developed by Micheal Porter, used to describe a series of value-adding activities connecting a company's supply side with its demand side, and the information were treated only as a supporting element of the value-adding process, while rarely were used or considered able to generate new value for the customer.

[10] OECD, *Human Capital – The Value of People,* OECD Publishing (2005).

Nowadays, the ability to effectively manage information and extract knowledge from them, is seen as a key competitive advantage. The data value chain is an important instrument for coordinate the different activities regarding an information system[11]. A value chain is made up of a series of subsystems each with inputs, transformation processes, and outputs, and it can be applied to information flows to understand the value creation of data technology. In a Data Value Chain, information flow is described as a series of steps needed to generate value and useful insights from data. The following are the key high-level activities:

- Data Acquisition: the process of data gathering, filtering and selection, before they are put in a data warehouse (or any other storage solution) on which data analysis can be carried out. This phase is one of the major big data challenges in terms of infrastructure requirements.

- Data Analysis: the phase that transform the raw data into something meaningful for the purpose of the decision process. The aim of the analysis of data is to highlight relevant data, synthetizing and extrapolate hidden and useful information with high potential from a business decision perspective. From this phase derive activities like *data mining, business intelligence* and *machine learning*.

- Data Curation: the active management of data during their life cycle. This process aims to validate data, to ensure the necessary data quality requirements for their effective use. In this phase there can be made different activities by data curators (also known as scientific curators, or data annotators), such as content creation, transformation, validation and preservation of data.

- Data Storage: the persistence and management of data in a scalable way that satisfies the needs of applications that require fast access to the data. For nearly 40 years, the unique solution for data storage was the Relational Database Management Systems (RDBMS), but this system has become unsuitable for big data scenarios. Therefore,

[11] *L'Economia dei Dati - Tendenze di Mercato e Prospettive di Policy*, ITMedia Consulting (2018).

NoSQL[12] technologies have been developed with the scalability goal in mind and they are prevailing in the data storage businesses.

- Data Usage: the final step that covers all the data-driven business activities that needs access to data. It is the output of the process, the reason why of all the previous phases. Data usage in business decision-making (in Chapter 1 we talked about the data management revolution) can enhance competitiveness through reduction of costs and increased added value.

Therefore, data represent a fundamental input, and in order to generate a successful informative output, it is necessary that the raw information goes through all the above mentioned distinct phases. In the first phase de-structured data are gathered, selected, pre-processed and transformed in order to reduce the complexity and improve the precision of the results of the analysis. In the second phase the selected data are analyzed to get the relevant information for the decision process. Eventually, the information obtained is evaluated and interpreted until it forms an informative output coherent with the objectives. The value generated by data and from their analysis, has revolutionized the traditional relationship between customers and producers. In the past, corporations sold their products to the clients in exchange of money and maybe some data of negligible value. Nowadays, on the contrary, transactions, and every interaction with clients, creates precious information. Therefore, corporations are willing to offer free service in order to obtain those valuable information, the transaction is no longer money in exchange of the products, but data in exchange of the free use of the product or the service.

In the next paragraphs we will analyze more in detail each phase of the Data Value Chain.

3.1.1 Data acquisition

The majority of data acquisition scenarios is characterized by high-volume, high-velocity, high-variety, but low-value data. Therefore, it is important for data controller and data processor to

[12] NoSQL databases use a variety of data models for accessing and managing data, such as document, graph, key-value, in-memory, and search. They are optimized specifically for applications that require large data volume, low latency, and flexible data models.

have adaptable and time-efficient gathering, filtering, and cleaning algorithms that ensure that only the high-value fragments of the data are actually processed by the data warehouse analysis. The core of data acquisition lies in the phase of gathering data from different information sources, ranging from data gathered by sensors to data depicting online transactions, and also an ever-increasing part produced in social media and via mobile devices, or unstructured (text, video, pictures and media files) and at a very high pace (tens of thousands events per second). Then, once gathered, the aim is to store those data in scalable, big-data capable data storage. In order to achieve this goal, there are three required components (the above mentioned infrastructure requirements):

1. Protocols: they allow the gathering of information for distributed data sources of any type (unstructured, semi-structured, structured)

2. Frameworks: they allow the collection of data from the distributed sources by using different protocols

3. Technologies: they allow the persistent storage of the data collected through frameworks

These data acquisition tools have to deal with high-velocity, variety, and real time data acquisition, and they have to ensure a very high throughput. The main challenge in this phase is to provide frameworks and tools that ensure the required throughput for the problem at hand without losing any data in the process.

3.1.2 Data Analysis

The data acquisition process enables the subsequent tools of the Data Value Chain, the data analytical tools, to do their work properly. In fact, the data acquisition phase usually ends by storing the raw data in an appropriate master dataset, and connecting it with the analytical pipeline. As already said, data come in many forms, but there is one dimension to consider in the analytical phase: the amount of *structure* contained in the data. The more structure a dataset has, the more useful it will be to machine processing, indeed, structured data provides greater

analytical capabilities by defining a structured representation associated with the data. Therefore, big data analysis can be defined as the sub-area of big data concerned with adding structure to data in order to support decision-making. Without big data analysis, most of the acquired data would be useless, and the rest of the value chain would not function, since it transforms raw structured or unstructured data, composed of many different formats, in data ready for curation, storage and usage.
Today, industry is applying large-scale machine learning and other algorithms for the analysis of huge datasets. Machine learning algorithms use data to automatically learn how to perform tasks such as prediction, classification and anomaly detection.

3.1.3 Data curation

Data analytics in order to work properly need a good quality of the information analyzed, since the quality of data can have a significant impact on business operations. For this reason, the third phase of the value chain is data curation, that address the data quality issue, with the goal of maximizing the usability of the data through methodological and technological data management support. It is a crucial process in particular where there is an increase in the number of data sources and platforms for data generation. Data curation can be categorized into different activities such as: content creation, selection, classification, transformation, validation and preservation. However, being an emergent activity, there is still vagueness and poor understanding of the role of data curation inside the big data lifecycle. Indeed, the costs of data curation are usually not estimated or underestimated. In the future years, with the growth in the number of data sources and of decentralized content generation, ensuring data quality will be always more important for data management environments. Therefore, investments in the evolution of data curation methods and tools is a cornerstone element for ensuring data quality at the scale of big data.

3.1.4 Data storage

Big data storage is concerned with storing and managing data in a scalable way. The ideal goal to reach, for any data storage system, is to storage a virtually unlimited amount of data, dealing efficiently and effectively with a range of different data models, supporting both structured and

unstructured data, and working only on encrypted data for privacy reason. Obviously, all these needs cannot be fully satisfied at the same time, but in recent years, technological advancements have emerged that partly address these challenges. Corporations can now store and analyze more data at a lower cost and at the same time improving their analytical capabilities. These technologies are allowing traditional pre-internet sectors to become more data driven (such as the health sector). One of these technologies is the cloud based storage, that makes possible to save data on a remote database (the cloud) rather than keeping it on proprietary hardware. According to IDC study, by 2020, 40% of the digital universe will be touched by cloud computing, and as much as 15% will be maintained in a cloud. Cloud storage can be used not only by enterprises, but also by end users. For them, storing their data in the cloud enables access from everywhere and from every device in a reliable way. For enterprises, cloud storage provides flexible access from multiple locations and quick and easy scale capacity, as well as cheaper storage prices. However, privacy infringements are one of the main concerns related to the clouds, as we already saw with the famous scandal of the photo leak in Hollywood in 2014. In conclusion, we can say that there is a strong need to increase the maturity of storage technologies so that they fulfil future requirements and lead to a wider adoption also in non-IT-based sectors.

3.1.5 Data usage

Advanced data usage is the final phase of the Data Value Chain, it is the wide field of activities that addresses the goal of supporting the process of business decision-making. Indeed, data usage integrates the data analyses into the business decision-making, that includes the reporting, the exploration of data and the exploratory search. Big data usage scenarios can have the greatest impacts in the discovery of new relations and dependencies in the data that lead to economic opportunities and more efficiency, and also provide a better understanding of these dependencies, making the system more transparent and supporting economic and social decision-making processes. In order to reach those goals, besides the technology infrastructure needed for the interaction and collaboration of services from multiple sources, there are three fundamental requirements to be met: (1) regulations and agreements on data access, ownership, protection and privacy, (2) demands on data quality, (3) access to the raw data and to

appropriate tools for big data usage. Respecting these requirements lead to a more transparent, objective and reproducible decision process.

Sectors on which data usage has an huge impact are the manufacturing, energy, transportation and logistics sector. These sectors are undergoing a transformational change as part of an industry-wide trend, called *Industry 4.0*: a process of digitization and interlinking of products, production facilities, and transportation infrastructure as part of the Internet of Things development, which brings an evolution of old manufacturing processes. In this process, there is a need of aligning hardware technology (machine and sensors) with software technology (i.e. the data representation, communication, storage, analysis and control of the machinery).

Broadly, data usage as a decision support system relies on different techniques, from the least complex to the more complex they can be categorized as:

- Lookup: at the lowest level of complexity, it regards fact retrieval and search for known items, both processes in which data are only retrieved for different purposes without deeper analyses.
- Learning: in this case, there are more complex processes, such as comparison, aggregation and integration of data. These functionalities can support knowledge acquisition and interpretation of data, enabling comprehension.
- Investigation: the highest level of complexity in decision support systems, in which data are analyzed, accreted and synthesized. At this level, true discoveries are supported and they can influence planning and forecasting of the future business activities.

There is also an even higher level of complexity, where the above mentioned functionalities are automated to provide predictive analyses. An example is one of the application of data usage in the Industry 4.0 trend, i.e. the predictive maintenance of machinery based on big data, that allows factories to lower the cost of maintenance. When the effective application of big data is successful in bringing measurable benefits, measurable data quality and security, the big data usage become a *smart data* usage scenario, in which new business model are made possible or existing one are made more efficient and profitable.

3.2 Two approaches for the extrapolation of value

Corporations can extrapolate value from data through two distinct approaches: (1) the direct commercialization of data and (2) the approach based on data driven innovation.

3.2.1 Big Data commercialization

In the first case, with the commercialization of data, corporations aim to monetize the value of data as quickly as possible, with their diffusion and circulation. In this process corporations can have different functions and act as:

- Data suppliers: they extract value from data generated by themselves and organized in data sets, simply selling those data through different price schemes, such as pay-per-use, subscription, ad insertion or freemium. The kind of data they exchange are data that are reusable, but mainly rough and little differentiated. Data suppliers are usually those firms that acts in point of intense traffic of data, like the telecommunication, media and entertainment industries. For example, a telecommunication company can sell data generated by its users to a second firms that need those data for information on road traffic.

- Data managers: they are represented by those firms that classify, clean and analyze the information behind the rough data. In other words, they add value to rough data, allowing the usability, efficiency and functionality of those data. Therefore, data managers ease the use of big data, but they are not the direct users or re-users of data. Their price models are similar to that of the data suppliers.

- Data custodians: they act as trust framework providers, allowing the re-use and re-sell of big data supplying a solid and secure infrastructure. They usually works only with subscription price schemes. Substantially, they guarantee the integrity and quality of data from the sourcing until their effective use.

- Application developers: they are all those firms that project, build and sell applications that permit the commercialization of big data. They build software that allow the easy interpretation of big data for the human mind. Application developers usually work in partnership with analytical firms, data suppliers and industrial partners in order to develop solutions to innovative analysis to the final clients.

- Service providers: they develop services based on big data, offering services tailored for specific purposes directly to the users of data. They utilize data science, data mining, predictive model and results analysis to extrapolate information from rough data for specific contexts. The most advanced of this service is the *what if* analysis, in which big data become more valuable telling a specific information, such as the more profitable target of clients, or their commercial habits.

- Data aggregators: they are those firms that search, cross and contextualize data in such a way to find correlations, find efficiencies or visualize possible existing interrelations. The results of these analysis are then sold as service with higher value to firms of final customers, or government bodies. The classic example of data aggregator is represented by those comparing price services, such as Skyscanner for flights, or Trivago for hotels.

All the business model above analyzed are not mutually exclusive, since corporations can also create value through different approaches simultaneously. In this scenario, we can see how the Big Data commercialization is not circumscribed only to the trade of data, but instead is characterized by a full interaction with the data of the final users in order to reach specific typologies of transactions and services.

3.2.2 Data Driven Innovation

The second way corporations can extrapolate value from data is the process of Data Driven Innovation (DDI). DDI is an iterative innovation process consisting in the exploitation of any kind of data inside the company (*in house*), in order to create new value. Data driven innovation is leading to the development or improvement of new products and services, the improvement of the marketing strategies or more in general of the business decision process. Indeed,

according to an OECD study in 2015[13], there is a productivity improvement of corporations related to Data Driven Innovation of 5-10% and a reduction in administrative costs, for public entities, of 15-20%, thanks to a better efficiency, a greater tax revenue and a lesser risk of frauds and human mistakes. At an organizational level, there are two type of strategic actions resulting from the Data Driven Innovation processes and the underlying Data Value Chain:

1. the first actions concern the internal organization of the firms, with the final aim of making information inside the company more usable, in order to have a greater efficiency and efficacy of the business processes. In other words, this actions concern the improvement of the corporations' Information Management Systems (IMS)[14]. In order to achieve a better exploitation of data within the corporations, it is crucial to find a balance between three dimensions: a clear business strategy, a supporting IT infrastructure technology and an analytical widespread culture inside the company (as illustrated in the image below). The lack of balance between these three dimensions explains the reason why, commonly, there is a scarce analytical culture inside corporations, while it is only applied to a limited set of problems or decisional processes. To tackle this problem and enlarge the analytical approach to the entire organization, the Analytical Competency Centre (ACC) are cross-functional organizational teams designed specifically for this purpose, supporting and promoting the effective use of Business Intelligence (BI) across an organization. Moreover, another goal of these strategic actions, is to reduce as much as possible the time between the discovery phase of the data and the commercial exploitation of them.

2. The second type of actions concern the exploitation of data from client, with the aim of improving the customer relationship or to exploit the benefits of targeted advertising. A clear example is the Netflix algorithm used to analyze in details the preference of their users in order to give them a better service, tailored to their tastes. In particular, the Netflix algorithm gathers different cluster of users not based on socio-demographic

[13] OECD, *Data Driven Innovation Big Data for Growth and Well-Being.* Access at: http://oe.cd/bigdata (2015).

[14] Information Management System (IMS) is a general term for software designed to facilitate the storage, organization and retrieval of information.

aspects, but exclusively on their streaming preferences. Therefore, it is not important if a user is from Italy or from Japan, the only thing that matters is the common preferences they have in terms of movies and TV series. This is an important upgrade, made possible by the Big Data revolution. Indeed, in this new era, the categorization of users in cluster is not anymore based on macro categories (such as nationality, gender, age etc.), but instead, it requires the use of complex algorithm based on more sophisticated criteria.

Nevertheless, it is important to notice that the Data Driven Innovation paradigm does not guarantee automatically the creation of added value or a successful business. On the contrary, it is necessary that corporations are endowed of a long term business strategy and an appropriate level of technological investment in IT infrastructure. At the end, the positive effects of data are both for customers, with more innovative and personalized products and services, and for corporations, which can exploit new business strategies and become more competitive.

3.2.2.1 Data Driven Innovation in different sectors

The public sector is nowadays more conscious than the past of the potential of the value that the Data Driven Innovation paradigm can bring to them. Public administrations, indeed, gather on a daily basis a huge amount of data. However, there are some obstacles that need to be overcome in order to fully exploit the economic advantage that data can bring. They are in particular:

- the interoperability of data, i.e. the fragmentation of the property of data across different public administrations
- the legislative process, slower and longer compared to the velocity of technological processes
- the privacy and security of data problematics
- the issue of competency among the human capital employed in the public sector

The finance-insurance sector, instead, is the clearest example of a data driven industry. As a matter of fact, financial and insurance institutions worked with enormous volume of data regularly also before the advent of the Big Data technology; thus, they already had developed

Business Intelligence (BI) techniques. What is now changing for those institutions with the new technologies are the rapidity and accuracy of the data analysis, with lesser investments required. Data about the clients are of crucial importance for credit and insurance institutions, and the correct use of their data has the potential to transform their business, realize new revenue opportunities and handle efficiently the risks, simultaneously improving the relationship with the clients. A key concept is the 'customer intimacy', i.e. the process of following the requests of clients in real time, understanding their needs instantaneously in order to help them reaching their financial objectives, keep them informed constantly and identify the necessities of unsatisfied clients; substantially improving the overall quality of the service. Recent studies have shown as the 71% of enterprises in the financial sector at a global level are exploring data driven technologies and predictive analysis, and 70% of them declared the importance of data in their business. Moreover, 50% of firms in this sector has hired a Chief Data Officer (CFO)[15]. In 2015 a IDC survey affirmed that 35,5% of financial-insurance institutions consider data as their priority for the next years[16].

Eventually, data driven innovation has revolutionized the way of gathering, processing and exploiting data in all the economic sectors. In particular, the disposal of a greater volumes of data is extremely favorable for those services in which the data is the product itself, or in the case it is strictly related to the offered product. However, in order to exploit at the maximum the potential of this innovation process is necessary to develop coherent strategies for the use of data, i.e. investing in human capital and in IT infrastructure, and at a legislative level, create a clear legislation in terms of privacy and competition.

3.3 How to maximize the value of data

There are infinite ways to exploit the economic potential of data. For this reason, it is firstly necessary to understand to which specific business necessity data should respond. To each business necessity corresponds a different *big data type*. For example: in the utility sector, there

[15] Accenture, *Exploring Next Generation Financial Services: The Big Data Revolution* (2016). Access at: https://www.accenture.com/t20170314T051509__w__/nl-en/_acnmedia/PDF-20/Accenture-Next-Generation-Financial.pdf

[16] *Key success factors for digital transformation in the banking industry*, IDC Report (2015).

can be business problems about the prediction of power consumption, the big data type that better fits with this issues are the machine-generated data; or in the marketing sector, in which marketers conduct sentiment analysis, the data type will be web and social data. After this first phase of identification of the data type, there is a second phase in which the big data type is analyzed in order to classify the *big data scenario*. The elements to be defined in this phase are:

- Analysis type - it depends on whether the data is analyzed in real time or batched for later analysis. This choice affects several other decisions about products, tools, hardware, data sources and data frequency. Also a mix of both types may be required
- Processing methodology - the type of technique to be applied for processing data, it can be a predictive, analytical or 'query and response analysis'. It depends on the business requirements, and usually they may need a combination of different techniques.
- Data frequency - it is defined by how much data is expected and at what frequency. Knowing these aspects helps determine the storage mechanism, and the storage format. It depends on the type of data sources: on demand (as with social media), continuous feed or real time feed.
- Data type - it is the type of data to be processed: transactional, historical, etc. This distinction helps segregate the data in storage.
- Content format - the content of data may be structured, unstructured or semi-structured. Knowing this information help to determine how data need to be processed.
- Data source - where data is generated: web and social media, machine-generated, human generated etc. Identifying all the data sources helps determine the scope from a business perspective.
- Data consumers - it is a list of all the possible consumers of the processed data (business users, business processes, individual people in various business roles etc.)
- Hardware - it simply depends on the type of hardware on which the big data solution will be implemented.

Therefore, a data set has no intrinsic value, unless it is processed following the right criteria for the business necessities at stake. The main challenge in this process is the difficulty in analyzing completely different data. The level of difficulty is determined also by the volume of data collected across the entire organization and the many different ways different types of data can

be combined, contrasted and analyzed to find patterns and other useful information. The first challenge is in breaking down silos to access all data an organization stores in different places and different systems. With this massive heterogeneous volume of data is no longer possible to utilize the old *data warehouse*. Nowadays, indeed, the level of complexity is higher, and behavioral data can be analyzed together with transactional data and all other types of data. The concrete ways data can be used to generate value are:

(i) Optimization of costs: data can be employed for predictive analysis in order to have better information on maintenance works, supply chain, logistic planning etc. With a greater efficiency in those activities there is an inevitable reduction of administrative costs.

(ii) Optimization of revenues: insights deriving from the analysis of data can lead to the exploitation of new opportunities, entering in new markets, improving the products and services and improving the relationship with the final clients.

(iii) Marketing and targeted advertising: more data are collected by firms about their clients, easier is for them to convert this information into targeted advertising. Through personalized ad, the efficacy of ad campaign double on average, and according to a recent study of Forbes[17], 89% of marketers use predictive analysis in order to improve the ROI of the marketing investments.

(iv) Market intelligence: all those activities relevant to a company's markets, gathered and analyzed in order to improve the decision-making process.

(v) Market-making: they play a crucial role in making possible the meeting of the needs of buyer and sellers in one place. Clear example of market making firms are ride-sharing or online dating apps.

[17] L. Columbus, *89% of B2B Marketers Have Predictive Analytics on their Roadmaps for 2016,* Forbes (2016). Access at: https://www.forbes.com/sites/louiscolumbus/2016/01/24/89-of-b2b-marketers-have-predictive-analytics-on-their-roadmaps-for-2016/#5b9d55451822

(vi) Training data for artificial intelligence: machine learning and deep learning need huge amount of training data.

Eventually, the value of data lies in their uniqueness and modality of use. Trying to understand the value of each bit of information that should be gather, combined and analyzed is difficult mainly because corporations themselves cannot fix the value of their data until they are able to specify and understand how to use them. Potentially, data within an organization can represent only a small percentage of the total revenue, but, at the same time, they can also become a key success factor for the future growth of the business.

3.3.1 Data lake

The first challenge in the process of valorization and analysis of data within an organization is to grant access to the data all over the company departments. Indeed, no matter the sector in which a company is, the majority of them has inside their structure a department dedicated to the creation of silos of data. These silos make difficult the free circulation and sharing of data within the different departments. This organizational barrier limits the exploitation of data, and make it also more expensive. In order to overcome this organizational barrier, there is an optimal solution called *data lake.* Data lake is an instrument able to simplify the access to data for all the departments inside a company, through the integration of all the disposable data in a single *repository*, that is easily and flexibly accessible. The entire organization can access to all the information contained inside the data lake with a simple query. Moreover, the data lakes can memorize in real time huge volume of data, also of different formats, and they are the ideal space of work for data scientists (in it they can build data-driven application and conduct analytical projects). All the most advanced digital corporations, such as Amazon and Facebook, utilize this system in order to handle their huge amount of data in an effective and efficient manner. Therefore, data lake are platform useful for handling data in a flexible way, for aggregating data of different type in one *lake,* and to allow the exploration of all data inside a company for every department. The most common data lake platform is Hadoop, claimed to be used by more than half of the companies in the Fortune 500. Hadoop's strength comes from its flexible nature, since companies can expand and adjust their data analysis operations as their business expands. In principle, every business sector would benefit from the capacity of

integrate data of any format and quality, in particular the public sector is the one that can achieve the greatest benefits. Data lake integration platform, indeed, contribute to a holistic approach in the analysis of data, combining socio-demographic and behavioral data, giving back a more defined profile of the final customer.

3.3.2 Data mining

Data analysis is a process of inspection, cleaning, transformation and modelling of data through algorithms, with the aim of find and highlight relevant information that support the decision-making process, and can also lead to new business opportunities. The standard algorithms for the analysis of data are based on deductive and inductive principles, i.e. there is a distributive model associated to the data set, on which mathematical deduction are made. Nevertheless, even when the results of the analysis are accurate, there is a certain degree of uncertainty, due to the interpretation of the results that should be made.
There are many different approaches to analyze data, one of them is the process of data mining. Data mining is the discovery process, based on predictive model, of relations, patterns[18] and information inside data sets, previously unknown and potentially useful for the business organization. Therefore, it is an analytical process finalized to select, explore and model huge volume of data in order to find relations and information concretely transformable in commercial actions, with the final aim of getting an economic advantage. Data mining instruments do not substitute the traditional type of knowledge, instead, they are an integrative tool for the decision process. Therefore, results obtained through data mining should be presented, communicated and shared with the company's department that will need this information the most. Moreover, data mining techniques can be divided in two groups:

1. Supervised data mining (also known as predictive or directed): it is a top down approach used when there is a specific target value to predict. In this model the goal is to acquire long term knowledge, i.e. applicable also in the future.

2. Unsupervised data mining (also known as descriptive or undirected): it is a bottom up approach, in which there is not a target value, but where the goal is to find hidden

[18] Patterns represent a structure, a model, or more in general, a concise representation of data.

structure and relation among data. This approach is commonly used in the first exploratory phases.

4. PRIVACY

"Once we have surrendered our senses and nervous systems to the private manipulation of those who would try to benefit from taking a lease on our eyes and ears and nerves, we don't really have any rights left. Leasing our eyes and ears and nerves to commercial interests is like handing over the common speech to a private corporation, or like giving the earth's atmosphere to a company as a monopoly." (Marshall McLuhan, Understanding Media - 1964)

4.1 Privacy rights and data protection rights

In general terms, privacy may be defined as the concept that one's personal information is protected from public scrutiny. Nonetheless, privacy has multiple dimensions[19], they are:

- Physical privacy: it concerns the protection against physical intrusions and right to bodily integrity.
- Informational privacy: it is the dimensions more directly related to data-driven economy, it concerns limiting access to personal data and rights against unauthorized publication of personal information.
- Decisional privacy: it is the freedom of taking personal decision without the interference of the State, for example abortion rights, divorce rights etc.
- Proprietary privacy: it is the right to control personal identities attributes.
- Associational privacy: it concerns the ability of meeting with selected individuals of one's own choice.
- Intellectual privacy: it refers to the so called freedom of thought.

The right to privacy can be considered as an inherent human right, since it derives from our physiological needs. Indeed, the innate desire for privacy is common not only among humans but in all mammals, because we consider surveillance as a physical threat (animals are surveilled by predators in the natural world), and it makes us feel like prey. Moreover,

[19] M.E. Stucke, A.P. Grumes, *Big Data and Competition Policy*, Oxford University Press (2016).

psychologists, sociologists, philosophers and technologists have all written about the effects of constant surveillance on human mind, even only the perception of it, and they all agree that it has negative consequences for our health, and it threatens ourselves as individuals.

Around the world the right of privacy is mentioned in more than 150 national constitutions[20]. In the Italian Constitution it is enunciated in the Article 15: "Freedom and confidentiality of correspondence and of every other form of communication is inviolable. Limitations may only be imposed by judicial decision stating the reasons and in accordance with the guarantees provided by the law".

In the rhetoric of the European Union law, the right to data protection was initially strongly connected to the right to privacy and early data protection instruments were explicitly linked to the right to privacy. Thus, the right to data protection was seen as sub-set of privacy interests, and not as fundamental right itself. The first frameworks for data protection on a European level were issued by the Council of Europe in 1973 and 1974, and were titled Resolution 'on the protection of privacy of individuals vis-à-vis electronic data banks in the private and public sector'. Here, the right to data protection were still seen as a part of the right to privacy. Later on, in 1981 the *Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data* by the Council of Europe specified in its object and purpose: 'The purpose of this Convention is to secure in the territory of each Party for every individual, whatever nationality or residence, respect for his rights and fundamental freedoms, and in particular his right to privacy, with regard to automatic processing of personal data relating to him ("data protection"). In 1995, in the Data Protection Directive there are 13 references to the right to privacy[21] and in Article 1, concerning the objective of the Directive, it stated: "[..] Member state shall protect the fundamental rights and freedoms of natural persons, and in particular their right to privacy with respect to the processing of personal data". Gradually, however, when the European Union, instead of the Council of Europe, started to be more interested in the field of data protection, the perspective of the issue changed. More precisely, where the Council of Europe's main focus is the protection of human rights, the one of EU has

[20] Source: Wikipedia. Access at: https://en.wikipedia.org/wiki/Right_to_privacy

[21] B. Van Der Sloot, *Legal Fundamentalism: is data protection really a fundamental right?*, Chapter from book *Data Protection and Privacy: Invisibilities and Infrastructure* (2017).

always been the defense of the internal economic market. Therefore, there was a gradual shift that has led to consider data protection as an economic matter. All this bring us to the new General Data Protection Regulation, entered in force this year, in which the term privacy has been removed entirely. For instance, terms like 'privacy by design' have been renamed 'data protection by design' and 'privacy impact assessments' have become 'data protection by assessments'. Even the Article 1, about the objective of the Regulation, does not refer to the right of privacy, it instead states: "[..] This Regulation protects fundamental rights and freedoms of natural persons and in particular their right to the protection of personal data." Therefore, there is an ongoing fully disconnection between right to privacy and data protection. This is also reflected on higher regulatory level, in Chapter second of the Charter of Fundamental Rights of European Union, entered in force in 2009 after the Treaty of Lisbon, the right of privacy is granted in the Article 7, just before the Article about the protection of personal data (Art. 8), as the *Respect for private and family life:* "Everyone has the right to respect for his or her private and family life, home and communications".

The fundamental reason at the basis of this separation is the inadequacy of the right to privacy for the modern challenges of large automated data processing, and besides this disconnection between the two different but intertwined rights, data protection has also been regulated on an ever higher regulatory level and with ever more detailed legal regimes, as the challenges and the complexity increase hand in hand with the technological improvements.

4.2 The golden age of surveillance

Privacy debates have always been at the center of a global controversial debate, in particular for what concerns the government surveillance on our life, and the question of how much we can sacrifice of our privacy for public safety concerns. In the last decade the amount of surveillance from governments and corporations has increased enormously, and nowadays we are all open books to both governments and corporations. They both gather, store and analyze an enormous amount of data that we create and trade for nearly nothing, as we move through our digitized lives. The "privacy paradox" consists in the desire of the majority of people to protect their personal information, while, at the same time, giving away this data explicitly for free. Corporate surveillance and government surveillance are not separate, they are intertwined, with different purposes but with the same goal: knowing everything about everyone. Therefore,

governments and corporations support each other in a public-private surveillance partnership at a global level, without a formal agreement but with an alliance of interests. As an evidence of this partnership, a 2013 Washington Post story reported that 70% of the US intelligence budget goes to private firms[22]. The main criticality is the opaqueness of such alliances, since much government control of corporate communications infrastructure occurs in secret, and we only hear about it occasionally, as it happened in 2013 with the Snowden documents leaks. Those documents made it clear how much the US National Security Agency (NSA) relies on US corporations to spy on people. Through programs like PRISM, for example, the NSA legally oblige companies like Microsoft, Apple and Google to provide data on several thousand individual of interest. Sometimes corporations work with the NSA accordingly; but it is not always the case, in fact, sometimes corporations are forced by the courts, largely in secret, and other times NSA hacks into corporations' infrastructure without their permission. An historic example of how government force corporations to spy for them, is the Communications Assistance for Law Enforcement (CALEA), a US law passed in 1994 after a great pressure on the Congress from the FBI, a law that require telecommunication services to re-engineer their digital switches to have eavesdropping capabilities built in it, enabling and facilitating FBI's telephone surveillance. More recently, in 2008 the US government secretly threatened Yahoo with a 250 thousand dollar per-day fine, if it did not join the NSA's PRISM program. Therefore, it is clear that if FBI or the NSA wants to turn a business into a mass surveillance tool, they are entitled to do so, and they can do all of this in secret and then force corporations to keep that secret too.

Unfortunately, this is not an US phenomenon, but it is happening all over the world, with many countries that are using corporate surveillance capabilities to monitor their own citizens. Examples are the UK Government Communication Headquarters (GCHQ) that pay telco like Vodafone to give it access to bulk communications all over the world, or the French government that do the same with Orange; or the Italian company Hacking Team that sell its products of phone intrusion and monitoring to various governments, such as the one of Colombia, Egypt, Morocco, Nigeria, Saudi Arabia, Susan, Italy and many others. Therefore, much of the infrastructure that governments around the world use for mass surveillance is firstly built for corporate use. The technology, indeed, is value neutral. There is no technical difference

[22] R. O'Harrow Jr., *The outsourcing of US intelligence raises risks among the benefits*, The Washington Post (2013).

between corporate and government uses, so, for instance, a legitimate corporate tools for blocking employees from e-mailing confidential data can be also used by repressive governments for surveillance and censorship.

What has changed over the last years that has led to this mass surveillance era is that our interactions and conversations, and all our actions either online and offline, are no longer ephemeral as once were through most of our history. This means that everything is recorded and permanently available, as a consequence we will not forget anything because we will always be able to retrieve data from the past from some computer memory. The reason behind this phenomenon is the technological innovation, that had caused the drop of the cost of computing technology. As technology improved and prices dropped, corporate surveillance has grown from collecting as little data as necessary to collecting as much as possible and governments broadened their surveillance as well. As the cost of data storage became cheaper, computers were able to save more data and for a longer time, and as big data analytical tools became more powerful, it became profitable to save more information, eventually leading to the birth of the surveillance-based business model. The surveillance-based business model has become the dominant model for the Internet because people like free services. Nevertheless, the word "free" is misleading, because we do not pay with money but with our personal data in exchange of those free services. We let corporations spy on us and eventually we end up giving away those data for less than their value. It is important to remember in this case, that if something is free, it is because we are not the customer (or at least not only), but we are also the product.

The common denominator and primary goal of all this corporate internet surveillance is advertising. It all started when commercial services first enter into the Internet, when people were not willing to pay even a small amount for access, and there was a debate on how to charge people for them. Thus, the only possible revenue model that made sense was the advertising one, and surveillance has made advertising even more profitable. Before the advent of internet and targeted ad, advertising was very inefficient. Indeed, without the sophisticated means of internet and all the information that we create using it, much of the money spent for advertisement went wasted because it addressed un-targeted audience. For instance, a beer ad is wasted on someone who do not drink a beer, or a car ad is wasted unless the audience has the money and the willingness to buy a new car. Now, with surveillance-based marketing,

corporations know exactly who wants to buy a beer, or a car, and even the exact moment when someone is close to a place where he/she can buy one.

The other reason, beyond freedom, people willingly give personal data to corporate interests, is convenience. As a matter of fact, the surveillance-based services are not only free, but also valuable and useful. For example, Amazon recommendations are highly appreciated by users, since it recommends things to buy based on previous purchases, or services like Google maps, Uber, Waze, the more information they have about their users the more they work faster and better. Even personal and targeted advertising is useful, since we appreciate ads about things we are interested in. Nevertheless, there is a subtle threshold that ads must not overcome, otherwise advertising that is too targeted and personalized feels creepy. This is related to another interesting question, that is: how much more data helps? Where is the limit of information about customers that should not be exceeded? To make it clear with an example, it is very useful for a car company to know that someone is interested in a convertible instead of a SUV, but it is only marginally valuable to know that the same customer prefers the grey model or the black one. In economic terms, there are diminishing returns for the amount of information companies store.

4.3 General conflicting principles

The negative repercussions of this ubiquitous mass surveillance era on society are numerous, and the costs for society as a whole and for individuals undoubtedly outweigh the benefits. What needs to change firstly, is the cultural bias according to which there is a trade-off between security and privacy, while, as a matter of fact, it is a false belief. Actually, privacy and security are two concepts aligned with each other: when we have no privacy the sense of insecurity increase, and when personal spaces and records are not secure, we have less privacy. Moreover, the costs of privacy loss are hardly quantifiable and only become tangible with their aftereffects, and this is another reasons why people easily agree to lose some degree of privacy in order to feel more secure, and also the reason why people undervalue privacy when they have it, while only recognize its true value when they lose it. Moreover, the problem of policies that sacrificed privacy for more security is that the entire weight of insecurity is compared with the incremental invasion of privacy. To make it clearer, if for example there is a terrorist attack threat, the likelihood of that attack to happen with or without the hypothetical program of privacy

reduction only slightly change. It is not sure that the attack happens without the surveillance, but is also not sure that the attacks does not happen with the surveillance. Thus, the final goal of policy makers is not to find an acceptable trade-off between security and privacy, because both can and should coexist.

Another important conflict of concepts is the one between security and surveillance. They are conflicting design requirements, since a system built for security is harder to surveil, while, on the contrary, a system built for easy surveillance is harder to secure. However, for societies, security is more important than surveillance, then, it is more efficient a secure information infrastructure that inhibits surveillance, instead of an insecure infrastructure that allows for easy surveillance. Indeed, infrastructures, as technologies, are value neutral, and it depends on the use that people made of them, that can be for good or bad purposes. The general principle is that systems should be designed with the minimum surveillance necessary for them to function, and where surveillance is required they should gather the minimum necessary amount of information, retaining them for as little time as possible.

As concern transparency, it is essential to any open and free society and it is a necessary condition for another important principle: accountability. Transparency is the foundational basis of the fundamental social contract of a democracy, in which people cede power to institutions but they must know how this power is used, so that if there is an abuse or misuse of that power, there is also a corresponding coercion. Transparency basically allows citizens to know how governments act and oversee its activities, indeed, it is the only way that citizens have for protecting themselves after ceding power to politicians and institutions. The same concept is relevant also for corporations, corporate disclosure laws are meant to increase transparency between corporations and their stakeholders. Nevertheless, some degree of secrecy is also needed for both governments and corporations in order to work properly. More secrecy, in fact, corresponds to more power: institutional secrecy increases institutional power, while individual privacy increases individual power. Thus, the more transparency the less the power in the hands of corporations and institutions. Applied to the data economy, the principle of transparency requires that people should be entitled to know what data about them is being collected, what data is being achieved, for how long and how those data are used, and also by whom. In the new GDPR the role of transparency is emphasized more than the previous

Directive, in the Premise 39 of the Regulation it is stated that the collection, use, or consultation of personal data concerning natural persons should be transparent to them, and also that: "The principle of transparency requires that any information and communication relating to the processing of those personal data be easily accessible and easy to understand, and that clear and plain language be used." Nevertheless, the GDPR does not introduce the concept of transparency *ex novo,* instead, this was only the result of a long normative evolution started with the principle of accuracy in the Convention 108[23] and the principle of loyalty in the 1995 Directive[24].

4.4 Convention 108

The first binding international instrument in the data protection field towards a more transparent and accountable use of data was taken in 1981, in technological terms a different geological era, with the *Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data,* known as the Convention 108. The principal aim was to protect the individual against abuses which may accompany the automatic collection and processing of personal information relating to them. The Council of Europe, already then, notice how 'information power' brought with it a social responsibility of the data users both in the private and in the public sector. In the explanatory report they wrote: "It is essential that those responsible for these files should make sure that the undeniable advantages they can obtain from automatic data processing do not at the same time lead to a weakening of the position of the persons on whom data are stored. For this reason, they should maintain the good quality of the information in their care, refrain from storing information which is not necessary for the given purpose, guard against unauthorized disclosure or misuse of the information, and protect the data, hardware and software against physical hazards". Despite the fact that those words were written almost 40 years ago, they underline the same problematics that we are still facing today. Moreover, to reconnect to the previous paragraph, the explanatory reports also highlighted how

[23] Article 5: "Personal data undergoing automatic processing shall be obtained and processed fairly"

[24] In the Premise 38 it was stated: "If the processing of data is to be fair, the data subject must be in a position to learn of the existence of a processing operation and, where data are collected from him, must be given accurate and full information, bearing in mind the circumstances of the collection".

the established legal systems of the Member States were devoid of rules that could have helped accomplish these aims, since they only had laws on privacy.

Under this Convention, the parties (also non-Member states) were required to take the necessary steps in their domestic legislation to apply the principles it enunciates. The resolutions, indeed, listed a number of ground rules to be observed when personal information are stored in electronic data banks, but it was left to the discretion of the Member States how they would give effect to these rules. The two general principle that all national laws must had to recognize were: (1) the principle of publicity, i.e. the existence of automated data files should be publicly known; (2) the principle of control, i.e. that public supervisory authorities can require that the rights and interests of those individuals are respected by the data users. In addition, the Convention also addressed for the first time the issue of the transfer of flows of personal data, a problematic still discussed and unsolved today. The question is to establish to what extent national data protection laws afford adequate protection when data concerning individuals flow across borders. Indeed, protection of persons become weaker when the geographic area is widened. Already in 1981, the explanatory report to the Convention 108 stated: "Concern has been expressed that data users might seek to avoid data protection controls by moving their operations, in whole or in part, to *data havens*, i.e. countries which have less strict data protection laws, or none at all." A first step to solve this issue was made with the Directive 95/46/CE, now abrogated by the GDPR, with the introduction of the principle of adequacy, further improved by the new Regulation.

Such problem can be satisfactory solved only in one way, through international cooperation. Some steps forward have been made in Europe, as we will see more in details with the GDPR, but Europe is still only a part of the globe, thus, until an utopian global agreement is reached, the problem of the transfer of personal data across borders will persist.

4.5 General Data Protection Regulation

The Regulation 2016/679 of the European Parliament, ratified the 27 April 2016, better known as the General Data Protection Regulation (GDPR), entered into force the 25 May 2018, replaces the already-mentioned Data Protection Directive 95/46/EC, adopted the 24 October 1995. GDPR aim is the harmonization of data privacy laws across all European Member states, in order to protect and empower all European Union citizens' data privacy and to reshape the

way corporations across the continent approach data privacy. It is the last step, in chronological order, taken by European Union toward a more transparent and accountable use of personal data. It is an intervention of the EU legislators with the final aim of a formal and substantial harmony of the safeguard system in Europe, and it is the result of a four years troubled negotiations. Since it is a European Regulation, it is self-executing, with immediate and directly applicable norms in each Member State. For this reason, the legislative framework of a Regulation is characterized by more complete and clear regulatory contents. This effort of the European legislators is justified by the divergences created during the twenty years between the Member States of the old Data Protection Directive, starting from 1996, both at a legislative and jurisdictional level. This Regulation aims to reduce those divergences, towards the advancement of the Digital Single Market; however, the other side of the coin is that it can create an excessive regulatory rigidity.

4.5.1 The guiding principles

The three pillars of this new regulation are transparency, responsibility and accountability, for the data controllers and processors[25]. As the Data Protection Directive, also the GDPR is based on the fundamental principles defined by: the *Fair Information Practice Principles (FIPPs)*[26], the OECD Guiding Lines of 1980 and the Convention 108. Those principles are stated in the Chapter II, on Principles indeed, and more precisely in the fifth and sixth articles of the Regulation. They are:

- *Lawfulness and accuracy*: the processing of data is lawful only at the conditions written in Article 6, i.e. when:
  (a) the data subject has given consent to the processing of his or her personal data for one or more specific purposes.

[25] In Article 4, about the definitions, 'data controller' is defined as the natural or legal persons, public authority, agency or other body which determines the purposes and means of the processing of personal data; while 'data processor' means a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller. This distinction is one of the novelty of the GDPR.

[26] FIPPs form the basis of the United States Federal Trade Commission's privacy compliance policies and procedures governing the use of *Personally Identifiable Information* (*PII*)

(b) the processing is necessary for the performance of a contract to which the data subject is party.
(c) processing is necessary for compliance with a legal obligation to which the controller is subject.
(d) processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller.
(f) processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child.

- *Transparency*: besides being a duty for the data controller and processor and a right for the data subject, transparency must be an essential characteristics of the processing of data itself. It is first enunciated in Article 5, point (a), together with the principles of lawfulness and fairness. Then, there is Article 12 in Section 1 of Chapter III (Rights of the data subject), on 'Transparency and Modalities', and it states that the controller shall take appropriate measures to provide any information relating to processing to the data subject in a concise, transparent, intelligible and easily accessible form. For this reason it is requested that data should be collected in a transparent manner, with clear and comprehensible information policy.

- *Purpose limitation* (art. 5, point b): according to this principle personal data should be collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes. There is, however, an exception, when further processing are in the public interest, or for scientific or historical research purposes.

- *Data minimization* (art. 5, point c): it means that personal data collection and processing should be "adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed".

- *Accuracy* (art. 5, point d): in this case, the principle of accuracy simply requests that personal data should be always accurate and kept up data, and consequently, that all the inaccurate data should be erased or rectified.

- *Storage limitation* (art.5, point e): it requires that personal data should be stored for no longer than is necessary for the purposes for which the personal data are processed. Also in this case, there is the exception in the case of public interest, historical or scientific purposes, in which personal data may be stored for longer periods.

The Regulation, then, in point *f* of article 5, states the new principle of '*Integrity and Confidentiality*', requiring that "personal data shall be processed in a manner that ensures appropriate security of the personal data, including protection against unauthorized or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organizational measures". Another important novelty of the GDPR, is the conditions for consent[27] (Article 7), which should be always unequivocal and preventive. In addition, the Regulation forbids the tacit consents in any case and the data subject shall have the right to withdraw the consent at any time in an easy way.

4.5.2 The main innovation of the GDPR

The GDPR emphasize the rights of the individual, the duties of the corporations and public administrations, and the relations between them, in order to tackle all the most critical aspects that concern the data economy. It creates new figures and roles, and for the first time it defines, in Article 4, a distinction between the 'Data Controller' (Art.4, n.7), i.e. "the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data"; and the 'Data Processor' (Art.4, n.8), i.e. "a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller." There is also a third figure, the 'Data Recipient', that

[27] In Article 4, about the definitions, 'consent' is defined as "any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her.

is as in the other cases a natural or legal person, a public authority, agency or another body, to which, in this case, the personal data are disclosed.

4.5.2.1 New obligations for the processor (or controller) of personal data

The actual reform of the GDPR, that strengthen the protection of personal data, lies in the new set of responsibilities of the data controller and data processor. Beyond the already mentioned new transparency obligations, the European legislators has introduced new duties for who control and process personal data. Specifically, the most important are enunciated in Article 30, 33 and 34:

- Article 30 - *Records of processing activities*: in the case of enterprise or organization employing more than 250 persons, data controller shall maintain a record of processing activities under its responsibility. The record should contain a list of minimum information, between which: the name and contact details of the controller, the purpose of the processing, the categories of recipients to whom the personal data have been or will be disclosed, transfers of personal data to a third country or an international organization and a general description of the technical and organizational security measures. Those records should be in writing, including electronic form.

- Article 33 - *Notification of a personal data breach to the supervisory authority*: as the title says, this article concerns the duty of data controller and data processor to notify the personal data breach[28] to the competent supervisory authority. Moreover, the notification should happen without delay, i.e. not later than 72 hours after the data controller has become aware of the breach. In case the notification does not happen within the 72 hours, it shall be accompanied by the reasons of the delay. In addition, the notification should include a minimum amount of information, such as the nature of the

[28] A data breach is an unauthorized access and retrieval of sensitive information by an individual, group or software system.

personal data breach, the likely consequences of the data breach and the measures taken or proposed by the controller to address the issue.

- Article 34 - *Notification of a personal data breach to the data subject*: it is the same obligation of article 33, but in this case the notification concern the individual persons damaged by the data breach. The notification is required when the leak of data is likely to result in a high risk to the rights and freedoms of the natural persons involved. On the contrary, the communication is not needed when the controller has implemented appropriate technical and organizational protection measures in order to avoid the breach (such as encryption[29]), or when the controller has taken subsequent measures which ensure that the risk is no longer likely to materialize, and also when the notification would involve disproportionate effort (in this case it will be sufficient a public communication or similar measure to inform the data subjects in an equally effective manner).

4.5.2.2 Territorial scope

In Article 3, concerning the 'Territorial scope', the Regulation addresses the issue of the transferability of data across borders and introduce considerable progress in the European legislation. Indeed, the GDPR abrogates the *country of origin principle*[30] of the old Data Protection Directive, and expands its field of application, ratifying the applicability of the Regulation "to the processing of personal data in the context of the activities of an establishment of a controller or a processor in the Union, regardless of whether the processing takes place in the Union or not." Moreover, the Regulation applies also to the processing of personal data of data subjects who are in the Union, while the controller or the processor is not, when the offering of goods or services to such data subjects are in the Union[31], or when the monitoring

[29] Encryption is a particular action that rends the personal data unintelligible to any person who is not authorized to access it.

[30] Article 4 of Data Protection Directive (95/46).

[31] Article 3, n.2, point (a)

of their behavior as far as their behavior takes place within the Union[32]. In other terms, the GDPR has an extraterritorial effectiveness, since it is applicable either if the data subject is actually or virtually in the EU territory, or if the processor or controller are established in the Union area and also when the processing take place outside the Union area. This innovation is born of the digital evolution and globalization process, and it aims to respond to the challenges set by the technological advancements and from new business model, at the same, taking care of personal data safety needs detected by the European citizens[33].

In addition, the GDPR regulates also the transfers of data across borders, a common practice that is growing year after year. The Regulation puts some restrictions on those transfers outside the Union, to third countries or international organizations[34].
In particular, in Article 45, it enunciates stricter criteria for the evaluation of the adequacy (*adequacy test*) of the third country, to give the permission of the transfer. Indeed, in the assessment of the third country process, the Commission must take into account: the rule of law, the respect for human rights and fundamental freedoms, the relevant legislation, the access of public authorities to personal data, the existence and effective functioning of one or more independent supervisory authorities and the international commitments the third country in question has entered into. In the case of the absence of a decision according to the criteria established in Article 45, in point 1 of Article 46, the Regulation allows the controller or processor the transfer of personal data to a third country, only in the case they provide appropriate safeguards, and on condition that enforceable data subject rights and effective legal remedies for data subjects are available. In point 2, the article lists which are those appropriate safeguards to be provided, like, for example: a contractual clause between the controller or processor and the controller, processor or the recipient of the personal data in the third country or international organization, or a provision to be inserted into administrative arrangements between public authorities which include enforceable data subject rights.

[32] Article 3, n.2, point b)

[33] Garante per la protezione dei dati personali, *Guida al nuovo regolamento europeo in materia di protezione dei dati personali* (2016).

[34] The GDPR for the first time include also international organization in the adequacy test, while the old Directive considered only third countries.

4.5.2.3 Supervisory authorities

As a consequence of the self-executing nature of the GDPR, there is a stronger focus on the functions and roles of the national supervisory authorities, since as a Regulation it has to establish in details the functioning, prerogatives and also the coordination between the different national authorities. At this purpose, the GDPR has introduced two new chapters:

- Chapter 6 about the independent supervisory authority[35].
- Chapter 7 about the cooperation between the lead supervisory authority and the other supervisory authorities concerned.

Particularly important, in Chapter 7, is the introduction of a new mechanism, called *one-stop-shop* (or consistency mechanism), in order to tackle the common circumstance in which the same processing of data is operated in more than one country in the Union, thus, involving more national supervisory authorities. According to the *one-stop-shop* principle, when the processing of personal data takes place in more than one Member State, one single supervisory authority (the lead authority) should be competent for monitoring the activities of the controller or processor throughout the Union and taking the related decisions. This lead authority should be the supervisory authority of the Member State in which the controller or processor has its main establishment. In this way, the GDPR tries to ensure a fast, consistent supervisory decision, providing legal certainty and reducing administrative burden. The decisions of the lead authority will be binding also for the other processing of data of the same data controller in the other Member States, in order to avoid that the same violations of the right to data protection can be treated differently in different States.
Moreover, the GDPR replaced the *Article 29 Working Party*[36] with the European Data Protection Board[37] (EDBP). The EDBP Board is established as a body of the Union and it has

[35] Article 51, point 1, defines Supervisory Authority as independent public authorities responsible for the application of the GDPR, in order to protect the fundamental rights and freedoms of natural persons in relation to the processing and to facilitate the free flow of personal data within the Union.

[36] Article 29 of the Data Protection Directive (95/46).

[37] Article 68 GDPR.

legal personality, it is composed of the head of one supervisory authority of each Member State and of the European Data Protection Supervisor. Also the European Commission has the right to participate in the activities and meetings of the Board designating a representative, without voting rights. The principal task of the Board (Article 70) is to ensure the consistent application of the GDPR, acting independently when exercise its power. To that end it can advise the Commission on any issue related to the protection of personal data in the Union, and issue guidelines, recommendations and best practices.

4.5.2.4 Remedies and penalties

A further framework that distinguish the GDPR from the old Directive is the one concerning remedies and sanctions. In Chapter 8, about remedies liability and penalties, the Regulation, as in the chapters on the supervisory authorities, is more detailed compared to the old Directive. In Article 82, it is enunciated the right of data subject to compensation when there is a material or non-material damage as a result of an infringement of the Regulation. Data subjects have the right to receive compensation from the controller or processor for the damage suffered, and they can also choose where to bring the proceedings: either in the courts of the Member State where the controller or processor has an establishment, or before the courts of the Member State where the data subject has his or her habitual residence.

As concerns the administrative fines, the GDPR has heavier fines for corporations that do not respect the Regulation. In the determination of the administrative fines, in Article 83, it is specified which variable of the infringement the authorities should take in consideration, like, for example: the nature, gravity and duration, the intentional or negligent character, the action taken to mitigate the damage, the degree of responsibility and the degree of cooperation with the supervisory authority. The maximum amount of the fines has been increased also, with a maximum fine of 20 million euro, or up to the 4% of the total worldwide annual turnover of the preceding financial year of the accused corporation. For what concerns penalties for infringements which are not subject to administrative fines pursuant Article 83, the Regulation gives free space to the national legislations, that should take all the necessary measures to ensure that those are implemented, and that such penalties are effective, proportionate and dissuasive.

5. COMPETITION

The competition in the digital economy has its own characteristics, such as "winner takes all" situation, network effects, two-sided or multi-sided platform, fast paced innovation and huge investments[38]. Data, generating revenue from user-data-based profiling and advertising, are a source of innovation that can affect the traditional levers of competition, thus they are a subject of interest also for competition authorities. However, the existing antitrust regulations were conceived for this pre-data economy. Therefore, anticompetitive behaviors of big tech companies remain mainly unpunished. In this chapter, we are going to analyze briefly the European antitrust law, and then, how antitrust authorities should analyze the relationship between anticompetitive conducts and Big Data.

5.1 Antitrust law

Antitrust laws aim to protect competition in the market, since competition encourages companies to offer consumers goods and services at the most favorable terms, and eventually, it encourages innovation and the reduction of prices. In order to be effective, competition requires companies to act independently of each other, but, at the same time, subject to the competitive pressure exerted by the others incumbents. The enormous growth of user data collection in recent years, and all the consequences that this Big Data revolution has having on the everyday life of individuals, and on the business activities in every sectors, has led to an ongoing debate on whether Big Data presents an antitrust issue or not. In any case, since Big Data are a source of innovation that affects also the traditional competitive levers, they are a subject of interest for antitrust legislators around the world.

At an European level, European antitrust policy is developed around two central pillars: Article 101 and 102 of the Treaty on the Functioning of European Union (TFEU):

- Article 101: it prohibits agreements between two or more independent market operators which restrict competition. It covers both horizontal agreements (at the same level of

[38] OECD, *OECD Digital Economy Outlook* (2017), OECD Publishing.

the supply chain), and vertical agreements. The clearest example of infringement of Article 101 is the creation of a cartel between competitors, which may involve price-fixing and/or market sharing

- Article 102: it prohibits firms that hold a dominant position on a given market to abuse that position. Abuse of dominant position can be the charging of unfair prices, the limitation of production in order to increase the price, or the refusal to innovate to the detriment of consumers.

The European Commission is empowered by the Treaty to apply these rules and has a number of investigative powers to that end. Moreover, the Commission has also the power of impose fines on undertakings which violate the EU antitrust rules. At a national level, there are the National Competition Authorities (NCAs), empowered to apply the above mentioned articles of the Treaty.

5.2 Competition in the Big Data realm

As concerns data economy, there are the myriad of pro-competitive benefits for the economy created by the exploitation of Big Data by private firms[39]. As a matter of fact, unprecedented consumer benefits have already been realized through the use of big data. The first and most obvious benefit to consumers has been the ability of firms to offer heavily subsidized, usually free, services, as consumers give those firms permission to monetize consumer data on the other side of their business (two-sided platform). Moreover, online data are used by firms to improve and refine products and services in a number of ways, and to develop brand new innovative product offerings. The complications for antitrust authorities derive from the difficulty of prove how the exploitation of Big Data create a barrier to entry, while, data-driven markets are actually characterized by low entry barriers, where little user data is required as a starting point for most online services. Data driven markets, indeed, are in theory characterized by low entry barriers, where little user data is required as a starting point for most online services. Firms can enter with innovative new products that address new customer needs, and thanks to its

[39] G. Sivinski, A. Okuliar, L.Kjolbye, *Is Big Data a Big Deal? A Competition Law Approach to Big Data*, European Competition Journal (2017).

innovative solution, they can quickly collect data from users, which they can use to further improve the product and finally compete with the incumbents on equal terms. The history of the digital economy offers many examples of that kind, where a simple insight into customer needs enabled new entrants a rapid success, despite the established network effects. This was the case, for example, of the dating online application Tinder, when it was launched on the market in 2012, it had no access to user data, but still, it became the market leader within a couple of years, thanks to the strength of its underlying solution (a simple user interface). Therefore, according to this line of thought, new entrants are unlikely to be at a significant competitive disadvantage relative to incumbents, since they may only lack of asset equivalence (since the incumbents have larger data store), but this has never been a sufficient basis to define a barrier to entry up until now. Further elements in support of this thesis, are the intrinsic characteristics of data (already analyzed in the first chapter), mainly their non-exclusive and non-rival nature. Therefore, big data cannot be compared to other key inputs, since if one provider has a piece of data, another provider is not prevented from collecting the same piece of data. In other words, the collection of a piece of data by one firm does not occur at the expense of another firm. Once collected, stored and analyzed, data has a limited lifespan, since the value of data lessens over time, therefore, any competitive advantage data provide is momentary, and entrants are unlikely to be significantly disadvantaged relative to incumbents. In conclusion, according to this theory, potential competitors do not need to create a data store equivalent to the size of the incumbent, while they need to build a sustainable competitive advantage focusing on the managerial toolkit and organizational competences, that allows to turn big data into value to final consumers, carving out a niche of their own. Nevertheless, there is the other side of the coin, where the growing importance of Big Data as an input, and the consistent increase in the four Vs of data, is seen as a way for the corporations to gain and sustain an unfair competitive advantage. Firms with a data-driven competitive advantage are not only in the position to dominate their own sector, but also to take over adjacent fields. In data-driven economy, indeed, border between industries has become increasingly blurred.

There is an ongoing debate on how to tackle these inefficiencies of the data economy, between who is in favor of more proactive antitrust enforcement, and others considering antitrust inappropriate for regulating big data. According to the latter, data economy should be regulated

by consumer protection laws[40]. There is also a third possible way, where antitrust and consumer protection laws cooperate together. Indeed, the relationship between antitrust law and consumer protection law is usually seen in terms of complementarity: antitrust focus on the supply side of the market, worrying that consumers can benefit the widest range of products/services at the lowest price possible, through fair competition between firms; consumer protection law, instead, focus on the demand side of the market, worrying that consumers are able to exercise consciously the choices that competition allow them to freely exercise. Hence, antitrust law is not interested in how much the choice of consumers are mindful and responsible, since this is the role of the consumer protection law. However, it is antitrust authority's duty understand how the exploitation of data can lead to illicit anticompetitive conducts, in order to stop or at least mitigate them. There are two steps that antitrust authorities should follow in order to analyze the relationship between anti-competitive conducts and Big Data:

1. Firstly, they need to understand how Big Data affect the business activity

2. Secondly, they need to understand how and if these business activities can exploit big data for anti-competitive behavior.

5.2.1 How Big Data affect the business activity

Big Data can give rise to network effects, and these phenomenon plays an important role in an antitrust analysis. Network effects, or network externalities, are a type of economies of scale in consumption rather than production, thus they are referred also as demand-side economies of scale. Technology giants have always benefited from network effects: the more users sign up on Facebook, or the more users utilize the Google query, the more attractive and convenient (in term of quality of the service) signing up become for other users. Moreover, data are characterized even by *extra network effects*, also called *data-network effects*: using data to attract more users lead to more data generated by those extra users, which in turn help to improve the service and, eventually, the improved service attract even more users. However, the presence of network effect itself does not automatically results in anticompetitive behavior.

[40] D. Daniel Sokol, *Antitrust and Regulating Big Data*, University of Florida Levin College of Law (2016).

There are different types of network effects that can come into play in online platform that exploit big data. They can be categorized in three groups:

1. Direct network effects: they occur when a product or service becomes more valuable to a single user as more people use that product or service. This is the case of social network or photo sharing platforms, and also chat applications.

2. Indirect network effect: they occur when more users make the product or service more valuable, but not through the direct interactions between users, as in the case of direct network effects. Search engines are those who more exploit the indirect network effects, since more users allow the search engine to gain insight from user clicks into what are the users' wishes, therefore improving the quality of search results.

3. Cross platform effects: they occur to firm operating a two-sided platform[41], where more users on one side of the platform makes the platform more attractive to users on the other side of the market. However, the theory of network externalities cannot analyze accurately the behavior of multi-sided businesses, since it does not take into account multi-sidedness into account[42].

The network effects are based on the *feedback loop*, with which big data can lead to economies of scale. There are two ways this feedback loop can generate economies of scale:

1. The user feedback loop: according to this theory, a platform that gains more users can also collect more user data (feedback indeed), leading to better insights into customers and their needs.

[41] 'Two-sided platform' are those platforms where one provider caters to two different customer groups on different sides of the same platform. An example are the social media platforms, that give users free access to social networking on one side, and rely on the provision of advertising services to businesses on the other side of the platform for revenue.

[42] I. Graef, *EU Competition Law, Data Protection and Online Platforms*, Wolters Kluwer (2016).

2. The monetization feedback loop: this loop focus on the monetization from advertising on online platforms, and it claims that with more users on the platform the firm is better able to target and sell ads, thanks to the disposal of more user data. With the greater revenues coming from advertising, the firm can invest more in the improvement of the quality of the service, thereby attracting more users.

The network effects, together with economies of scale[43], characterize data-driven market, and they usually lead to a *winner takes all* result, creating insurmountable barriers to entry.
Nevertheless, there is a doctrine that consider the feedback loop, and its consequences, an overstated phenomenon. According to these theories, the non-rivalry and ubiquity of data weaken the strength of the loop, and data alone are not sufficient to improve the quality of the product or service in order to gain more users. Online providers can gain scale in users in ways that do not involve accumulating user data, such as gathering data from data brokers or making strategic alliances. Moreover, according to this line of thought, innovation can be more powerful than network effects, that have proven different times to be insufficient to avoid that newcomers disrupted established market leaders. An innovative product is enough to cause users to switch to other complementary products or services. Thus, even if scale is crucial to reach a competitive success, smaller rivals still maintain the ability and the incentive to compete. The mere existence of economies of scale and network effects does not, by itself, establish that tech giants have monopoly power and that they are acting anti competitively, or that competition have been harmed. Regulating large firms only for exploiting economies of scale can be dangerous, because it can become a tax on competitive success. Therefore, these thoughts highlight how difficult is for antitrust authorities, in data driven market, to draw a border between what is anticompetitive and what is only a licit competitive success, and the necessity of proceeding on a case by case basis. There is no *one-fits-all* solution for antitrust authorities, while they should analyze the singularity of each case, since Big Data are an upstream resource for business activities. In other words, data themselves do not create competitive advantage, while it is derived from the knowledge extrapolated through the analysis of them. From the same data set, indeed, firms can infer myriad of diverse information, that

[43] In network economies, in order to enter the market, firms have to made substantial investments. Once the initial investment is made, the incremental costs of creating additional units decrease until it becomes negligible. This is also the case of search engines, social networks and e-commerce platforms, which are characterized by high fixed costs and low marginal costs.

then can be used for the development of different products or services. The duty of antitrust authorities, therefore, is to understand every context and market separately, analyzing which characteristics of the data considered are crucial for the businesses on that particular market.
Moreover, the nature of data makes the antitrust remedies of the past less useful. For example, a classical remedy for merger harming the free competition of a market for antitrust authorities would be breaking up the monopolist. However, breaking up a firm like Google into different companies is useless, because one of the them would become dominant again with time. Therefore, as concerns merger and acquisition, a new approach is required, and two ideas stand out[44]:

1. The first idea is that antitrust authorities should consider the extent of firms' data assets when assessing the impact of merger and acquisition, instead of taking into account only the size of the companies, as they have historically done.
2. The second principle is to increase transparency, i.e. companies must reveal to consumers what information they hold and the economic value of it (how much money they make from them).

5.2.2 How businesses can exploit data for anticompetitive behavior

The second point is understanding how the results of the inferences made from data can be exploited by businesses for anti-competitive ends. The international debate on this topic is discussing the possibility that the exploitation of Big Data facilitates anticompetitive conducts, such as collusions, abuse of dominant position, price discriminating practice and concentrations that raise barriers to entry. However, before analyzing the single case of data-driven anticompetitive conducts, it is important to describe the first element to be considered in the event of a suspected infringement of the competition rules: the relevant market. The relevant market combines the product market and the geographic market, they are defined as:

- a relevant product market is composed by all those products and services which are considered as interchangeable or substitutable by the consumers

[44] *The World's Most Valuable Resource is no longer oil, but data,* The Economist (2017)

- a relevant geographic market comprises the area in which the businesses are involved in the supply of products and services and in which the competition conditions are homogeneous

In the data economy, an important tool useful to identify the relevant market is the Big Data Relevant Market (or simply BDRM). BDRM explains the whole picture of a given data market on its different stages. It is used to understand the structure of the relevant market and how the players act on each stage. BDRM therefore divide the Big Data Market in three parts:

- Big Data capture: it is the first level, where data are captured.
- Big Data storage: in this second level the data already captured are stored, which will then be accumulated and aggregated in datasets ready for the third phase.
- Big Data analytics: it is the final phase where the data are analyzed and combined with other information in order to find patterns and the developments of profiles, records, macro trends, applied for a variety of purposes.

Despite its simplicity, the BDRM can help competition authorities in analyzing more accurately the whole market structure and estimate the undertakings' market power. Particularly, it can be useful in signalizing that the Big Data cycle not only deals with overlaps on horizontal bases, but also on vertical ones, revealing existing or potentially dominant position. Identifying the BDRM under a competition law perspective may lead to verify precisely competition issues as market power, barrier to entry and abuse of dominant position.

5.2.2.1 Big Data and collusion

Big Data, machine learning and artificial intelligence are all exploited through algorithms. An algorithm can be defined as series of instructions for carrying out an operation or solving a problem. In other words, they are instruments that companies can use in different ways for different purposes.
However, there are certain kind of algorithms which have proven to be used by firms in an anticompetitive way: the "pricing algorithm". Pricing algorithms, as the name indicates, are used by firms to determine the price of the goods or services sold, after the collection and

analysis of a large amount of relevant data about the market. With those algorithms a firms can react instantaneously to price movements by competitors. They pose serious challenges for antitrust authorities, since they may facilitate collusion and complicate detection of unlawful agreements. Specifically, these algorithms are considered dangerous in those markets where there are the best conditions for collusion agreement among incumbents. Collusion agreements are those deals between two or more companies that would typically compete but who conspire to gain an unfair competitive advantage, restricting the supply of a good or service, or accordingly deciding to fix a higher price to maximize profits (price fixing). Collusion may occur in different forms, but the outcome is always the same: one party, the consumers, is eventually disadvantaged. According to economic theory, there are some type of market that favor collusion agreements. Specifically, those markets where there are few players, with similar dimensions and costs, high barriers to entry, high transparency and where goods or services exchanged are homogeneous. Those conditions also characterize the markets dominated by algorithms. For example, search engine and social network markets have few rivals and high barriers to entry (in particular: economies of scale and network effects). However, it is difficult to establish if algorithms are the cause or the consequence of these conditions, and therefore, antitrust analysis aims to understand if algorithms are influencing those variables in an anticompetitive way. Before the advent of Big Data, humans have for many years been the moving force behind these practices, colluding on every kind of economic activity. Nowadays, Big Data are providing new and more sophisticated means to foster collusion. In particular, there are four scenarios in which pricing algorithms may promote collusion[45]:

1. Messenger: in this scenario humans are still the masters behind the collusive agreements, while the algorithms are only an intermediary (the messenger indeed) that the cartel members program in order to help reaching the final aim. In this case, competition law applies straightforwardly, and prosecutors with sufficient evidence of the humans' agreement will have little difficulty in condemning them. Therefore, in this scenario algorithms are a mere extension of the human will.

[45] A. Ezrachi, M.E. Stucke, *Virtual competition – The Promise and Perils of the Algorithm-Driven Economy*, Harvard University Press (2016).

2. Hub and spoke: in this case, pricing algorithms are used as a central hub to coordinate competitors' pricing. This scenario is not present only on the online environment, since it is how price-fixing traditional cartels work, a central hub who controls numerous 'spokes', also called secondary co-conspirators, which do not need to communicate with each other. In the algorithm-driven hub and spoke, the algorithms are no more merely executor of humans' order, but rather, it is the use of the same pricing algorithm that stabilizes a cartel price and dampens competition. In the digital environment, this scenario happens when competitors outsource their pricing to an upstream supplier's pricing algorithm, because creating and refining an own algorithm is too expensive. Thus, competitors do not interact directly but they use all the same pricing algorithm to determine the market price and react to market changes. As a result, the market behavior could be aligned, since they use all the same "brain" to determine the price strategy. Another case of algorithmic hub and spoke conspiracy involve platforms, which bring together sellers and purchasers. These platforms, indeed, sets the price for all the competing operators, which can dampen horizontal competition. For antitrust authorities, it can be challenging to prove the illegality of these arrangements when there is not a clear evidence of anticompetitive intent.

3. Predictable agent: in this third scenario, each firm programs its own pricing algorithms with the final aim of maximizing profits. Those algorithms are programmed also to monitor price changes and to react to any competitor's price reduction (dynamic pricing). Therefore, algorithms engage in predictive analytics, i.e. the study of patterns in pricing and commercial decisions through the combination of real-time, historical and third-party data, in order to build forecasts of what will happen in the market in advance. Firms rely on their pricing algorithm without manually checking it, because it would be ineffective for humans to independently analyze all the underlying market data to calculate the prices on many different products. Speed, indeed, is critical in this scenario. Before algorithms, pricing decisions took weeks to be implemented, now algorithms can adjust prices within milliseconds. Therefore, the industry-wide use of

these algorithms increase the likelihood of tacit collusion[46], since no agreement between competitors is needed to fix prices.

4. Digital eye: the last scenario represents the next frontier of algorithmic collusion, which will amplify collusion to a new level of stability and scope. The conditions to reach this level are the further advancement of two aspects of algorithms: (1) their ability to analyze and process high volume of data in real time, (2) their sophistication in autonomous decision making and learning through experience (i.e. advancement in artificial intelligence). These two advancements will enable algorithms to have a wider and detailed view of the market ("God-like view"), and a faster reaction time in decision making. As a result, also markets less susceptible to tacit collusion, according to the economic theory, will be affected, and it will also be more difficult for antitrust authorities to detect it. In the Digital Eye scenario algorithms can anticipate and react to any competitive maneuver, beyond pricing strategy, and deciding autonomously how to retaliate. Consequently, when all competitors in a market quickly react to each competitive strategy of the others, the overall results eventually would be the deprivation of the incentive to undertake these initiatives in the first place. Therefore, in this scenario, not only the harm to competition is greater, there is also a murkier collusion, since human intervention is further detached from tactical and strategic decision. Antitrust authorities have to decide to whom assign responsibility for the illicit conduct: to the artificial instrument or to those who created and use them? A possible solution is the so called "antitrust by design", i.e. authorities should punish firms for how they have programmed their algorithms, not preventing them to avoid collusive conducts. However, such strategy is difficult to implement since the advanced and complex level of the algorithms.

In conclusion, we must also notice that algorithms can have also positive effects for antitrust analysis. Indeed, there are always more sophisticated investigative algorithms which are useful

[46] Tacit collusion describes the process by which firms in a concentrated market might share a monopoly power, setting their prices at a profit-maximizing level, without putting it in writing explicitly.

instruments for the investigation process, monitoring firms' behavior and collusion equilibrium automatically.

5.2.2.2 Algorithmic consumer price discrimination

The exploitation of massive amount of users' data by Internet platform, together with the already mentioned pricing algorithms, lead to another anticompetitive conduct: price discrimination. Differently from the collusion scenario, price discrimination is a competitors' unilateral strategy, and the outcome, instead of uniformly higher price, in this case, on the contrary, is personalized prices. Price discrimination, indeed, is a price strategy that charges to customer different prices for products or services whose costs are the same, according to their economic possibility. Thanks to the availability of data about consumers, like purchase habits and financial information, and the algorithmic assessment of each customer's reservation price (i.e. their predicted willingness to pay), Internet companies have nowadays the capability to charge each customer a personalized price, that is the maximum price he or she would pay, eventually maximizing profits. This type of discrimination is called fist degree price discrimination, or person-specific pricing, or even perfect price discrimination, and it was considered only a theoretical case before the advent of Big Data. Before data economy revolution, indeed, traditional businesses lacked the detailed, accurate information to evaluate each customers' willingness to pay, thus, there were an imperfectly price discrimination (or third-degree price discrimination), where businesses only differentiate price into macro groups, based on generic variables like the age, the occupation or the profession of the customers.
In recent years, person-specific pricing practices have spread extensively among e-commerce retailers. The biggest of them, Amazon, is able to update its prices for each customer every 10 minutes[47]. The critical aspects of this process is that the digital consumers only see their own personalized prices, and they are unaware of the discrimination that is taking place. Digital retailers defend themselves asserting that this personalization of the shopping experience is a benefit to consumers, and they call it "price optimization" or "dynamic differential pricing" process. In that case, in dynamic pricing process, prices simply vary in response to changes in supply and demand. However, whether this process ultimately benefits consumers or instead

[47] N. Howe, *A special price just for you*, Forbes (2017).

harm their welfare, is still an open question, what is sure is that discrimination can be potentially dangerous and antitrust authorities should analyze on a case by case basis.

5.2.2.3 Concentration and barriers to entry

The degree of concentration in a determined market is commonly measured by the *concentration ratio,* which indicates the size of the firms in relation to their industry as a whole. Concentration and competition are inversely proportional, if one of the two decrease the other increase, and vice versa. The concentration ratio is calculated as the sum of the market share percentage held by the firms in an industry, and it ranges from 0% to 100%. Low concentration, and therefore a competitive market, is indicated in a ratio that ranges from 0 to 50%. Antitrust authorities' concern is to condemn firms that operate with the aim of increase the concentration in a market, with the direct consequence of increasing the barriers to entry of that market. In data economy, increasing barriers to entry means make access to certain data costly for competitors, or even impossible. Generally, barriers to entry can be of various type:

- Natural barriers: they form naturally as the dynamics of an industry take shape.
- Structural barriers: they refer to all the necessary infrastructure and skills needed by a firms in order to operate on a market.
- Legal barriers: they refer to legislative constraints. For example, in data markets, when only one firm has the exclusive right to exercise a determined activity, and due to this activity, it has the exclusivity on certain type of data.

These three above mentioned barriers define the structure of a market. For this reason, antitrust authorities do not focus on any of them, since they cannot intervene on the market structure, while they utilize them only for descriptive ends, i.e. to understand more deeply the market they want to analyze. The actual focus of antitrust authorities is to unveil the so called "strategical barriers": the barriers erected by incumbents of a market on purpose, in order to impede the access to new competitors.
The harm caused by the erection of these strategical barriers to entry in the data market does not result in higher prices, as usually happens in this cases in other sectors, but rather in a loss of quality, innovation and privacy:

- Loss of quality: smaller firms cannot adequately compete with larger firms because they lack access to the same volume of data. Consequently, together with the data gap, also the quality gap widens between the dominant firm and the smaller rival. Therefore, the competitive constraint the rival poses to the dominant firm in terms of quality and innovation is diminished and the larger firm is not driven to innovate and to maximize the quality as it would do in a competitive scenario. The best example of this reduction of quality can be seen in search engines markets. Large search engines have the ability to prioritize paid advertising over more relevant, better quality, organic results, in order to have a greater likelihood of a pay-per-click conversion for the platform provider, and this in turn means greater profits to the detriment of quality. This became an antitrust problem when the gap between large and small providers is so wide, due to the different availability of data, that the dominant search engine can afford to sacrifice a higher level of search quality, in order to increase its profits, than a smaller search engine could, since it is already struggling on quality due to lower data levels.

- Loss of innovation: firms with value proposition built on collecting and monetizing user data, have also the power, through those data, to eliminate potential competitors in advance, preventing smaller rivals from accessing necessary data. Therefore, eliminating potential challengers by limiting or preventing their access to necessary data, or by acquiring them, prevent those firms to innovate. Where market leaders with greater capital and higher data levels acquire new entrants in the market, the degree of innovation is reduced overall, and competition suffers. This issue is strictly intertwined with the debate about the data-driven mergers. The big data merger cases have increased over time, and the transaction rationale driver of these mergers is to acquire the underlying data set of the target undertaking. Besides the possible harm to competition, these data mergers have an impact also on the privacy of users. Hence, the merger of different data sets through acquisitions between Internet business players gives to the acquirer the control of consumer data, without the permission of the users of the undertaking, which did not choose to interact with the acquirer. These considerations imply that merger in any information-based market must be examined for its potential effects on all dimensions of competition, including privacy.

- Harm to privacy: privacy protection can be seen as a form of non-price competition, and it is especially important in industries where the product or service is offered for free. Antitrust concern, indeed, is not only about price issues, but all what concerns consumer's choice, and price is a type of choice. Some policymakers have proposed a good way to regulate privacy violations using competition law, and these proposals can be grouped in four categories[48]. (1) The first group would evaluate transactions like mergers that involve large data sets, determining if the deal will reduce the merged firm's incentives to compete on consumer privacy protection. (2) The second group would balance the costs and benefits of consumer protection against the impact on competition when there is a conduct that implicate both consumer protection and competition principles. (3) The third group would control, under antitrust law, if companies mislead or deceive consumers about data collection practices that helped the companies achieve or maintain monopoly power. (4) Eventually, the fourth group, the most radical, would look for possible harm to privacy from transactions, not just analyzing the harm as an existing dimension of competition, but separately.

We will now analyze more in details the relationship between competition and privacy issues, and the possible interaction between competition law and consumer protection law.

5.2.2.3.1 Competition and consumer protection relationship

Until now privacy and antitrust concerns have been explored from two different perspectives. However, separating competition and consumer protection has been defined as an "artificial dichotomy"[49], which no longer make sense to maintain, since privacy is an increasingly important dimension of competition. Indeed, in markets where data are collected in exchange for free services, low level of privacy could signal high levels of market power. The reason why

[48] M. K. Ohlhausen, A. P. Okuliar, *Competition, Consumer Protection, and the Right Approach to Privacy*; (2015).

[49] P. Jones Harbour. T. Isa Koslov, *Section 2 in a Web 2.0 World: An Expanded Vision of Relevant Product Markets,* Antitrust Law Journal (2010).

antitrust agencies have serious difficulties in incorporating invasions of privacy interests into their competition analysis, is that privacy violation lead to non-economic and intangible harms to consumers. On the contrary, competition violations are easier to detect, since they cause tangible and economic harms to consumers. This non-economic and intangible nature of privacy explain the misleading belief according to which privacy issues are unrelated to competition concerns, and that competition policy concerns and consumer protection should be left to their respective agencies. A clear example of this bias, is the European Commission investigation of Facebook's acquisition of WhatsApp in 2014 (which we will see in more details in paragraph 5.3), in which the Commission recognized that Facebook could be a threat for WhatsApp's privacy commitments. Nevertheless, they eventually focused only on Facebook's position in the online advertising market, and not on individual privacy concerns, as if privacy-related effects were unrelated to consumer welfare.

Privacy concerns can cause a deadweight welfare loss, increasing distrusts in the market. Trust is what market economies rely on, and therefore it must not be ignored by competition law. Indeed, as a consequence of distrust in a market, transaction costs would increase, and eventually the market do not grow as it can. In online market, in particular, trust is even more crucial, otherwise, without it, consumers may no longer use the technology out of privacy concerns. On the contrary, when the consumer protection law is effectively enforced, consumers will trust firms' privacy promises, and they will divulge their personal information more easily. A more active, informed role by the competition authorities can foster an effective competitive process between firms on privacy, which will use privacy safeguards as a brand differentiator. Competition authorities, privacy and consumer protection officials must therefore coordinate in order to reach this kind of competition based on the level of privacy offered. Rather than unrelated silos, they intersect and inform each other, in order to consider opportunities, synergies and their respective weak points, with the final aim of promoting competition and individuals' privacy interests. There are two initial areas on which the coordination between competition and consumer protection law should begin: first, moving beyond the notice-and-consent paradigm, and secondly, creating the preconditions to incentivize privacy competition.

The notice-and-consent regiment refers to the consensus consumers give to Internet services. This procedure does not adequately safeguard data users' privacy. Consumers, indeed, have little inclination to read lengthy, detailed and opaque privacy terms. A 2008 study calculated

that if internet users want to read all the privacy policy they agree on the Internet, they would spend 244 hours per year on average, i.e. more than half of the time that average user spends online[50]. Nowadays, ten years later, that time probably has climbed. Beyond the excessive length of the privacy terms and conditions, the content of them are incomprehensible for the majority of the Internet users, and even if they do understand, consumers cannot negotiate better terms. This imbalance of power between consumers and data users is a kind of market failure, that creates a locked-in effect for consumers. Therefore, simply increasing transparency is not enough, because transparency is a necessary, but not sufficient, condition for competition for as concern privacy parameters. For competition, to deliver the privacy safeguards that individual desire, there are some preconditions required:

- Consumers must have the possibility of choice. Currently, they are in a locked-in situation, instead, there must be mechanisms that allow consumers to have greater choice and control over their data.
- For privacy competition to flourish, an alternative business model to the advertising-dependent revenue model must be find. With the current dominant model based on targeted advertising, the economic incentive is towards the accumulation of always more personal data. Moreover, the advertising business model, due to its dominant position in the ecosystem, blocks innovative, alternative subscription-based model with more privacy-friendly services. Therefore, competition, privacy and consumer protection officials should coordinate and promote alternative revenue models where privacy is not sacrificed.
- Competition and consumer protection authorities should coordinate in order to identify ways in which consumers obtain more control over their personal data and data portability. If consumers cannot easily transfer their data to alternative services, the competition related problem is the higher switching costs, which facilitates the exploitation of locked-in consumers. Article 20 of the GDPR is a fundamental first step in this direction, since it consecrates the principle of data portability, forbidding data controller to oppose to the transfer of users' data to other systems if they request the transfer. Data portability is an important concept, because it treats data as something

[50] A.M. McDonald, L.F. Cranor, *The Cost of Reading Privacy Policies*, I/S Journal of law and policy for the information society (2008).

that is owned by consumers, and not by companies. We will see in the next chapter how this is an important step, toward the next and more radical one: the recognition of ownership rights to data.

- Another important aspect to consider is the value of data. Indeed, for an effective privacy competition to subsist, more transparency about data's value is needed. Users, without knowing the value of the data they are sharing, are unable to understand the transactions in which they engage. Instead, as they better understand how much their personal information is worth, they would likely demand more from collectors, and firms will have an incentive to become more efficient.
- Antitrust authorities should better understand the different forms of privacy degradation undertaken by Internet firms. Nowadays, they focus only on the economic aspect of competition, analyzing the small, but significant, and non-transitory increase in price (SSNIP) in a market. A cooperation with privacy and consumer protection officials can help competition authorities to analyze also the degradation of quality protection, through a SSNDQ analysis (i.e. small, but significant, and non-transitory decrease in quality). Technologies potential benefits, even if significant, should not mask the potential risks for our privacy.

Summing up, competition and consumer protection law should coordinate in order to: identify and understand all the potential harms arising from a data-driven economy, either economic and non-economic; understand firm's incentives to compete and invest in privacy friendly technologies and services; provide a legal framework to incentivize new business models, different by the current dominant advertising model, ultimately empowering consumers with more choices. Eventually, the synergies between the different legal frameworks will promote both competition and consumers' privacy interests, and broadly the consumers' well-being.

5.3 Case studies

In this last paragraph we are going to analyze three specific case studies that will help understand the theoretical aspects debated in this chapter. Specifically: European Union Commission investigation on the Facebook's acquisition of WhatsApp in 2014, the AGCM procedure against Facebook for the infringements of the Consumer Code, and the current

investigation of the Bundeskartellamt (the German Competition Authority) on Facebook for abuse of dominant position.

5.3.1 Facebook-WhatsApp merger

On August 29th, 2014, Facebook Inc. notified the European Commission of its plan to acquire WhatsApp Inc. for 19 billion dollars in cash and stock. Facebook, with the texting app Facebook Messenger, and WhatsApp, are both consumer communication services. They offer popular texting applications for smartphones, which consumers use to communicate by text, photo, voice and video messages. However, differently from Facebook, WhatsApp did not sell advertising space or collect personal data on its users at the time of the acquisition. Therefore, before the merger, consumers could choose between two popular texting apps, with completely different price-privacy tradeoffs.

The Commission had the duty to assess the merger according to Article 1 of the Merger Regulation, which aims to prevent concentration that would significantly impede effective competition in the European market. From the notification, the Commission has a total of 25 working days to decide if granting approval, or otherwise start an investigation. On October 3rd, 2014, the Commission approved the acquisition. According to the Commission, Facebook and WhatsApp were not close competitors, and consumers would have continued to have a wide choice of alternative consumer communications apps. The investigation, according to the Commission press release, focused on three areas:

1. Consumer communications services: those services are multimedia communications solutions that allow people to reach others in real times. The Commission did not consider the two apps close competitors in this sectors, because Facebook Messenger's user experience was considered specific, given its integration with the Facebook social network. Furthermore, users seemed to use the two apps simultaneously and in different ways. Moreover, it was also considered the dynamism of the market, that offers several competing texting apps. Eventually, the Commission in its valuation also considered the network effects mitigated by the fast growing market of consumer communications, characterized by short innovation cycle and lower switching costs for consumers, which can easily switch to different communication apps.

2. Social networking services: in this case there is not a clear definition of these kind of services. The boundaries that define what can be considered a social network are continuously evolving. The Commission decided to not consider WhatsApp and Facebook competing on the social network market, because of the substantial richer experience offered by Facebook's platform. Further, even if the Facebook's position in social networking services could have been strengthened, the net gain for Facebook would have been limited due to the significant number of overlapping users between the two platforms. For all that, the Commission considered that the merger was compatible with the internal market as concerns the provision of social networking services.

3. Online advertising services: Commission examined whether the transaction could have strengthened Facebook's position in the online advertising market. During the time of the investigation WhatsApp did not sell any form of advertising, and it did not store the users' data that would have been valuable for advertising purposes. Therefore, the Commission raised two possible theories of harm to verify if Facebook could have strengthened its position in online advertising: the first scenario was whether Facebook could have introduced advertising on WhatsApp, the second was if it could have used WhatsApp as a source of user data for improving the targeting of Facebook's advertisements. They concluded, for the first theory that: "*regardless of whether the merger entity will introduce advertising on WhatsApp, there will continue to be a sufficient number of actual and potential competitors who are equally well placed as Facebook to offer targeted advertising*[51]." In the second theory of harm, they concluded that. "*regardless of whether the merged entity will start using WhatsApp user data to improve targeted advertising on Facebook's social network, there will continue to be a large amount of Internet user data that are valuable for advertising purposes and that are not within Facebook's exclusive control*[52]." Therefore, according to the Commission valuation, in both cases, the transaction would have not raised concerns, because of the

[51] Case nº COMP/M.7217, Facebook/WhatsApp, §179.

[52] Case nº COMP/M.7217, Facebook/WhatsApp, §189.

sufficient number of alternative providers in the market. In addition, when Facebook notified the acquisition, it also informed the Commission that it did not intend to implement automated matching between the two platforms' users accounts. In other words, Facebook officially declared that it had no intention to use WhatsApp users' data to further targeting advertisement on Facebook platform.

Based on the above considerations, the Commission concluded that the merger would not have raised competition concerns. In the press release, the Commission stated an important final note, which stated:

"*In the context of this investigation, the Commission analyzed potential data concentration issues only to the extent that it could hamper competition in the online advertising market. Any privacy-related concerns flowing from the increased concentration of data within the control of Facebook as a result of the transaction do not fall within the scope of EU competition law".*

Since this decision was taken, three other investigations were opened by European competition authorities: one in Germany by the Buneskartellamnt in March 2016 on suspected abuse of dominant position by Facebook; another in Italy, by the AGCM in May 2017, concerning an infringements of the Consumer code; and a third one, by the same European Commission in 2017. As concerns the Italian and German cases, they will be analyzed in the next paragraphs. As concerns the Commission investigation, it arose from the misleading Facebook's promise regarding the use of WhatsApp users' data (the second theory of harm analyzed during the approval of the merger). Facebook officially stated both in the notification form and in a reply to a request of information from the Commission, that it had no intention to merge WhatsApp users' data into Facebook users' data for better targeted advertising on the social network. Notwithstanding, in August 2016, with a simple update of WhatsApp terms and conditions, Facebook started to match the data of the two services, associating each phone number and contacts from WhatsApp users to the corresponding person's data on Facebook. This led, in May 2017, the Commission to fine Facebook with 110 million euros fine for providing incorrect and misleading information during the 2014 investigation. Nevertheless, this infringement did not impact on the Commission's decision to authorize the merger, since it considered the

misleading information provided by Facebook not relevant on the outcome of the latter. The Commission further stated:

"*Today decision is unrelated to either ongoing national antitrust procedures or privacy, data protection or consumer protection issues, which may arise following the August 2016 update of WhatsApp terms of service and privacy policy."*

Hence, the two above highlighted final notes of the European Commission, clearly reveal how during the investigation process, they analyzed only potential data concentration issues, which could have dampened the overall competition. While, the privacy-related effects, that can arise from increased concentration of data in one single player, were considered unrelated to competition analysis, and thus, considered not a matter of EU competition law but of the EU data protection rules. This consideration, together with the two subsequent investigations on Facebook, confirms that the EU Commission approval decision of the transaction, without a real investigation, offers some doubts about how precisely it was taken, and that the Commission lacked a long term vision of the issue.

On the grounds of these considerations, the Commissions approving the Facebook-WhatsApp deal, wrongfully considered two aspects: firstly, it did not evaluate the privacy related issue deriving from the increased concentration of personal data in a single player; secondly, it did not consider appropriately the Big Data Relevant Market Structure (BDRM)[53]. As concerns the former, the Commission erred in stating that the concerns of one firm controlling so much data were strictly a privacy issue and not a competitive one, in this way it considered only half of the overall picture. Data concentration, instead, can affect multiple sides of a multi-sided market like the social network one. As concerns the BDRM, as already analyzed previously in this chapter, it can be an efficient and effective instrument for a better comprehension of the performance of firms in the Big Data market. It could have been used to verify more precisely the implications of the Facebook-WhatsApp deal to market power, barriers to entry and abuse of dominant position. Using this theoretical instrument, the Commission could have noted that Facebook and WhatsApp were already competitor at that time, specifically in the Big Data

[53] V. Bagnoli, *The Big Data Relevant Market As A Tool For A Case By Case Analysis At The Digital Economy: Could The EU Decision At Facebook/WhatsApp Merger Have Been Different?*, , 12th ASCOLA Conference Competition Law For The Digital Economy.

Capture phase. As a consequence, the merger led to a vertical integration which strengthened the market power of Facebook.

This Commission decision represents how limited is the current analysis of data-driven mergers in multi-sided markets where the product/service is free, because competition authorities do not have the effective tools to address their concerns regarding data accumulation. Competition authorities cannot estimate an overall data market value, since data are not fungible, and then simply assess the merging companies' share of this market. The value of data, indeed, depends on the context of its use, search data useful to Google may be not valuable to Facebook. Likewise, even if other firms can acquire the same type of data Facebook gained through the acquisition, they would probably be unable to analyze and capitalize the data as Facebook can. However, at the time of the acquisition it was not harder for the Commission to understand that the economic rationale behind that transaction was not for economic efficiencies. Facebook, indeed, acquired for 19 billion dollars a company that in 2013 has earned 10 million in revenues and suffered a net loss of 138.146 million dollars[54], and it did not collect huge amount of data given its privacy policy. Therefore, the data-driven merger must somehow have provided a sustainable competitive advantage, but the Commission predicted it would not.

5.3.2 The Italian WhatsApp case

Related to the Facebook-WhatsApp merger, as already anticipated before, the Italian competition authority, the *Autorità Garante della Concorrenza e del Mercato* (AGCM), on May 2017, closed two separate proceedings concerning alleged infringements of the Italian Consumer Code by WhatsApp Inc. The innovative aspect of the two proceedings was the recognition of the economic nature of personal data. Indeed, despite the free nature of the service, the information of users given to the messenger platform have high economic value. Thus personal data are considered a sort of consumers' counter-performance of the communication service[55], and when they accept WhatsApp terms of service, according to European consumer law, they enter into a business to consumer relationship, consequently, the

[54] M.E. Stucke, A.P. Grumes, *Big Data and Competition Policy*, Oxford University Press (2016).

[55] A. Cervone, *Unfair Contract Terms and Sharing of Data with Facebook, Towards A Better Protection Of Social Media Users: The WhatsApp Cases* (2017).

contract is subject to the requirements of the Directive on Unfair Terms in Consumer Contracts (UCTD), and therefore, for Italian users, to the requirements of the Consumer Code (the Italian Decree that implements the Directive).

The first procedure addressed imbalances in the relationship between user and provider, due to the application of some unfair contractual clauses in the Terms of Service. The Consumer Code prohibits the use of certain contractual terms, stating that "*in contracts entered into between consumers and professionals, terms shall be considered unfair where, contrary to good faith, they cause a significant imbalance in the rights and obligations arising under the contract, to the detriment of the consumer*" [56]. In the WhatsApp investigation, the AGCM assessed as illicit the contracts terms concerning:

- Limitation of liability: AGCM considered insufficient the 100 dollars liability cap of WhatsApp terms.
- Unilateral termination of contract: Italian competition authority considered unfair the WhatsApp provision on contract termination, which was discretionary, making difficult to consumers to know when and why they can be no longer allowed to use the service.
- Unilateral changes of terms: Consumer Code declares a term unfair when it enables the firms to alter the contract unilaterally, without a valid reason. Based on this, AGCM declared illegal the right granted to WhatsApp to introduce modification, of economic nature too, to the terms without indicating the reasons and without notifying consumer about the changes.

The second procedure concerned WhatsApp illicitly obliging their users to accept entirely new Terms, including the automatic transfer of consumer personal data to Facebook. The mere act of merging the consumer personal data from WhatsApp to Facebook itself was not illegal. However, according to the AGCM, WhatsApp induced its users to believe that without granting such consent they would not have been able to use the service. The practice has been implemented through an in-app procedure for obtaining the acceptance of the new terms, emphasizing the need to subscribe within the following 30 days, otherwise they would have lost the opportunity to use the service. In addition, there were scarce information on the option of denying the consent and the procedure to opt-out, once the terms were accepted in full, was

[56] Art. 33, paragraph 1, CC

really difficult. On this basis, WhatsApp users have been unduly conditioned to give a consent to new terms.
Since these illicit conducts significantly impaired the consumers' freedom of choice, obliging them to accept a transactional decision that they would not have made otherwise, AGCM eventually adopted a prohibition decision that banned the commercial practice and its continuation, and it also imposed a fine of 3 million Euros.

5.3.3 The German investigation on Facebook dominant position

The Bundeskartellmt, the German Competition Authority, from now on indicated as the GCA, started an investigation of Facebook in March 2016, concerning an alleged abuse of its dominant position in the social networks market in Germany. The proceeding is still in progress. The GCA is assuming that Facebook abuse its dominant position by making the use of the social network conditional on the possibility to limitlessly amass every kind of data generated by its users, in particular when they use third-party website, in order to eventually merge those data with the user's Facebook account (the so-called "Facebook package"). The third-party sites include either services owned by Facebook (WhatsApp and Instagram) and websites and operators with embedded Facebook APIs (Application programming interfaces). Through APIs users' data are collected and transmitted to Facebook when they visit other websites, embedding Facebook products like the 'Facebook login option' or the 'like button' into third-party websites. Data collected on third-party websites are indicated as "off Facebook" data, and the GCA procedure is mainly concerned of this kind of data, overlooking the "on Facebook" one (data collected directly on the social network).

The premise on which the investigation is based, is that Facebook has a dominant position in the social networks market in Germany. In its preliminary assessment, the GCA confirmed this hypothesis: Facebook has a dominant position due to its massive amount of users (around 30 million per month, of which 23 are daily users), and also because of the scarce substitutability of the products (social networks). Indeed, other social networks, according to the GCA, exists, but they all serve complementary needs, such as, for example, professional networks (Linkedin) or video streaming networks (YouTube). The GCA accuse Facebook of imposing unfair conditions to its users, forcing them to choose between accepting the whole "Facebook package", or otherwise they cannot use Facebook at all. In this way, Facebook creates a locked-

in effects for its users, and simultaneously forecloses rivals in the market, dampening competition in the overall advertising market. Moreover, the damage to users consists in a loss of control of their personal data, since they are no longer aware and able to control how and when their data are collected and processed. Therefore, thanks to direct and indirect network effects, Facebook exploit its dominant position taking advantage of superior access to data, which in data-driven market is an essential factor for competition.

The innovative aspect of this investigation is the GCA different approach, with the simultaneous analysis of the case under the perspective of both competition law and data protection law. Bundeskartellamt president Mundt argued that "*data protection, consumer protection and the protection of competition interlink where data are a crucial factor for the economic dominance of a company"*, and also the European Commissioner for Competition Margrethe Vestager said that this investigation fell into the "grey zone between competition and privacy"[57]. Furthermore, in the note about the background information on the proceeding, it is stated that:

Monitoring the data processing activities of dominant companies is an essential task of the competition authorities which cannot be fulfilled by a data protection authority. In its assessment of whether the company's terms and conditions on data processing are unfair, the competition authority does, however, take account of the legal principles of data protection laws. For this purpose, the Bundeskartellamt works closely with data protection authorities.

In sum, according to GCA preliminary assessment, Facebook terms and conditions are unfair both under competition law standards and under data protection principles. However, the GCA main focus remains on the antitrust issue of Facebook's strategy, while also applying data protection principles. Specifically, under a legal point of view, the Bundeskartellamt is concerned with two antitrust issues:

1. Facebook's data accumulation is strengthening its dominant position to the detriment of competitors in the social networks market.

[57] P. Lugard, L. Roach, *The Era of Big Data and EU/US Divergence for Refusals to Deal,* Antitrust Magazine (2017).

2. Facebook's request for a single grant of consent for the whole "Facebook package" is considered as an unfair trading clause (Art. 102, point a, TFEU).

Since the proceeding is still in progress, we cannot examine the conclusions to which the GCA will arrive. Nonetheless, we can consider the analysis made by the doctrine[58] about these two legal characterizations of Facebook's conduct. As concerns the first particular case, according to both European Union law and US law, a dominant firm's conduct is anticompetitive when: it is capable of excluding rivals and responsible of a reduction of the consumer welfare. In the Facebook case, none of the two conditions applies. Indeed, Facebook API's embedded in third-party websites do not prevent that website to incorporate the products of other social networks, and the fact that one firm receives increasing quantities of data does not prevent its rivals from doing the same. Therefore, embedding APIs into third-party websites does not reduce competition and it does not constitute an exclusionary and anticompetitive conduct. As concerns the second theory of harm, it could be argued that by collecting data through APIs, Facebook lower the quality of the services offered by third-party websites., i.e. making those services less privacy-friendly. However, this point would be difficult to be proved within the meaning of Article 102 point b, which states that abuse of dominant position consists in "limiting production, markets or technical development to the prejudice of consumers". In order to prove that, it would be necessary to demonstrate that a reduction in privacy is directly related to an overall reduction in quality and eventually in consumer welfare. On the contrary, as long as data accumulation bring to new personalized services, it will be hard to demonstrate that it is an anticompetitive conduct, even if it is not disputed that the accumulation and processing of personal data may strengthen the Facebook dominant position. Consequently, the GCA should follow a different path, and attempt to apply Article 102 point a, which states that an abuse of dominant position may consist in "directly or indirectly imposing unfair purchase or selling prices or other unfair trading conditions". Therefore, since the collection of users' data take place under the terms and conditions imposed by Facebook, one could argue that the act of data collection becomes an exploitative abuse where the conditions imposed are unfair. Clauses are considered unfair when they are unjustifiably unrelated to the purpose of the contract, when they unnecessarily limit the freedom of the parties and when they are disproportionate,

[58] G. Colangelo, M. Maggiolinio, *Data Accumulation and The Privacy-Antitrust Interface: Insights From The Facebook Case For The EU And The US,* Standford Law School and the University of Law TTLF Working Paper No, 31/2018.

unilaterally imposed or seriously opaque. Thus the GCA could be right in accusing Facebook, particularly because of the opacity of the Facebook's practice. However, this has two consequences. First, focusing on the opacity of the contractual clauses recall more to privacy concerns and not to competitive ones. Second, even once a possible abuse is detected, another issue comes out: what would it be the right remedies and sanction? A typical privacy-related remedy would be the disclosure of the unfair practice in order to ensure user awareness. However, such remedy would not prevent further data accumulation and the alleged entrenchment of Facebook's dominant position, but it would only guarantee a fair data collection.

In conclusion, this investigation can teach us two lessons. Firstly, the privacy-antitrust coordination presents different pitfalls, indeed, it is questionable if the GCA through this approach will resolve the concern about data accumulation. Secondly, it demonstrates that when there is a mismatch between the goal pursued and the tools available to pursue it, there are two possible options: prosecutors refrain from intervening, i.e. giving up as the European Commission did in the Facebook-WhatsApp merger case, or impose a forced interpretation of a particular provision, as in this case with the Article 102 (a), but eventually ending up with remedies that cannot realize the goal pursued.

6. DIGITAL CAPITALISM

What it was described in this work, until now, is the data economy as it is today, with an emphasis on the critical aspects of it (namely, antitrust and privacy). In this chapter, it will be analyzed how we are far from the social and economic optimum. For this reason, we will define the current phase as the era of *digital feudalism*, because of the parallelism with the feudalism period during the Middle Age. Therefore, we will analyze what is necessary to change in order to solve, or at least mitigate, those critical aspects, moving toward a phase of *digital capitalism*.

In the last century, capitalism was vital for the rise of a global and developed society, with much less wars and more cooperation between states. This was possible thanks to the absolute trust posed on growth at every costs, through an approach based on the mutual benefit. In other words, capitalism stopped the common belief of the pre-modern societies that economy is a zero-sum game, in which the profit can come only at the expenses of others, so that the eventual net change in wealth is zero. This mutual benefit approach should be applied also to the data economy, while now it is still anchored to the zero-sum game logic, in which Internet companies make huge profits at the expense of users.

Capitalism is now living its second phase, the digital one. Digital capitalism is no longer based on international flows of goods, or on financial flows. What links the world together in this new environment is cross-border data flows. They have grown by a factor of 45 only during the last decade[59], and they are expected to grow more in the coming years. As a result, digital companies go global in a leaner and less-capital intensive way compared to the old multinationals, who need to build physical structure in each country where they operate. This has allowed little digital startups to expand into new markets without losing their agility and flexibility, fostering their growth, until, in some cases, they have become tech giants which dominate the global market.

[59] S. Lund, J. Manyika, J. Bughin, *Globalization is Becoming More About Data and Less About Stuff,* Harvard Business Review (2016).

6.1 Digital feudalism

Data users of today can be compared to the serfs of the Middle Ages. During the Medieval period, serfs had no property rights, and they had to hand over most of what they produced to the landowners (the feudatory). In exchange of their hard work, serfs received services, like protection from conflict or access to a village oven. Moreover, they had little or no choice to opting out those unfavorable deals, since there were no other options available for them, and they had no contractual power to negotiate better conditions. If we consider serfs as data users (i.e. us while using digital platform), the digital platforms and Internet as the land, and the landowners as the tech giants, such as Google and Facebook, the parallelism is clear. Nowadays, we are living in the dark ages of data[60], a sort of digital feudalism, in which consumers have no choice and no contractual power to negotiate their position against Internet firms. Data users agree, with one click, the terms and conditions of the platform they use, which subjects them to constant monitoring, and the collection and sell of their data to many more actors. In exchange for these data, data users have free-of-charge services. i.e. free access to sites, social media, search engines or music streaming platforms. Accepting those deals, data users also entrust their data to feed algorithms owned by Internet companies, which in return give a more personalized service. They share voluntarily most of their data only because it is a necessary condition of use of those free-of-charge services that gather all the data. In this bargain there are also some hidden clauses, such as:

1. A cultural clause: Internet users end up seeing what they want to see, hearing what they want to hear and reading what they want to read; increasing the subjective biases of everyone. In other words, the Internet imprisons users into their filtered bubbles.

2. A social clause: accepting the clauses, users implicitly accept the imposed systems of norms of each system, such as the norm that banish any naked image by Facebook. Jaron Lanier, a futurist and researcher of Microsoft, stated: "*Free inevitably means that someone else will be deciding how you live*."[61]

[60]G. Koenig, *Leaving the Data Dark Ages,* Project Syndicate (2018).

[61] J. Lanier, Simon & Schuster, *Who Owns the Future?* Penguin Books (2013).

For Big-Tech firms these terms of the deals are extremely lucrative, like it was for landowners in the Middle Ages. In numbers: the value of users' personal data is expected to reach is 8% of European GDP by 2020[62], roughly 1 trillion euros, according to the European Commission. Therefore, we live in an optimized extractive economy, characterized by an infinite availability of free raw materials (our personal data), that is enriching enormously few companies at the expense of consumers and society.

Given this overall picture, there are three conventional options of political philosophies to tackle this situation:

1. Nationalization: in this scenario, data are considered as a *res communis*, as air or water. Therefore, it would be necessary a national data agency, that brings together and encrypts all the data of the population, and then make them available to companies under stricter conditions. In other words, it is a sort of digital communism, but it would be really complex to realize, since it would create a bureaucracy diametrically opposed to the culture of the Internet.

2. Fundamental rights: this is based on a personalist logic, since it is attached to the rights of the person. It is the philosophy embraced by the European Commission, based on the concept that everyone should be able to decide about the use made of their personal data, and it is the philosophy that had guided the General Data Protection Regulation. To the user would be granted additional rights: the right to allow data to be circulated and to know the use made of those data, or the portability of data from one platform to another, or also the strengthening of consent procedures. As for the platforms, they would be subject to new obligations. The risks of this logic is the hampering of the innovation process, and at the same time, not offering real guarantees to data users.

3. Ownership rights: the third option is a way to make the second one a viable solution. Assigning ownership rights to data, i.e. making users the legal owners of their personal

[62]T. Vasudha, Arvind G., *The value of data,* World Economic Forum (2017).

data. This would be a revolution for the market as we know it today. Nowadays, indeed, nowhere in the world there is such property rights on personal data.

Property rights are what had revolutionized the feudalist economy of the Middle Ages, protecting and empowering individuals, and it is what could revolutionize and eventually abolish the digital feudalism in which we live. This topic will be the focus of these final paragraphs.

6.2 Property rights for data

Historically, technological change has been the driver force for radical economic transformation: the printing revolution brought intellectual property rights, while the Industrial Revolution brought the patent system. Nowadays, the digital revolution must bring to property rights for data[63]. The source of the problem is the absence of a real market for data[64]. Absent market means that digital information does not have a price, consequently, valuable data may never be generated. With the creation of an absent market it is possible to address simultaneously the inequality issue, the stagnation of the economy and the sociopolitical conflicts. Otherwise, the dearth of a market for user data makes also more difficult to solve the issues deeply analyzed in the previous chapters of this work, namely antitrust and privacy.
To understand what does it mean to create this market, with true property rights for data creators, it is useful to firstly analyze the three classic elements of property rights:

1. *Usus:* data users should use their personal data as they wish
2. *Abusus:* data users can destroy their personal data as they wish
3. *Fructus:* data users can profit from the value of their personal data

The first two rights, *usus* and *abusus*, would give data the consideration of something that can be appropriated and controlled, abandoning the logic of protection and replacing it with a logic of self-responsibility. The third point, the *fructus* right, is the one that lead to the most radical shift: data would be no longer considered capital (Data as Capital, DaC), but as labor (Data as

[63] L. Lèger, *My data are mine, Why we should have ownership rights on our personal data*, Report GenerationLibre (2018).

[64] E.Posner E. Glen Weyl, *Want our personal data? Pay for it*, The Wall Street Journal (2018).

Labor, DaL). Since data users are data producers, they should be paid for the raw material that they supply to the algorithms of Big Data. This aspect will be analyzed more in details in the next paragraph.

Nevertheless, there are some moral and philosophical argument that reject this possibility. The French *Conseil d'Etat* in its 2014 report[65] on digital technology and fundamental rights, rejected the possibility of giving property rights to data, because the monetization of them would undermine the protection of fundamental freedoms, since a person cannot be valued and traded like a good. However, to overcome this moral aspect, it is sufficient to consider data, under the law of property rights, as something that remain separate from the person, just as ideas can be protected by intellectual property rights.

6.2.1 Data as Labor

Internet companies are guided by a mercantile logic when they gather and exploit data. Data, indeed, are a source of considerable income, mainly through the increase in value of the advertising spaces sold, and the resale of data to partners or to data brokers. Moreover, thanks to data, businesses can customize the service, increasing the quality of it, providing a more fluid, simple and personalized experience to the customers. Therefore, although users are simply considered as consumers, they are actually also suppliers for internet platforms. When a user on Facebook or Instagram posts and labels photos, or when he/she use Google Maps while driving or upload a video on YouTube, the user is generating data about human behavior. Then, those information are used by these companies to feed their machine-learning programs, that will use those personal data to learn patterns that allow them to imitate and understand human behavior.

Notwithstanding, it is also true that consumers derive functional benefits from the use made of their personal information, and sometimes even monetary benefits (like in the case of loyalty programs). This argument is typically what big Tech firms use to rebut against critics for the unfavorable deal between them and the customers that provide data for free. However, the critical point is to understand if there is a balanced exchange of benefits, and it is not the case. The share of the value returned to customers in the form of functional, or even monetary

[65] Conseil d'État, *2014 Annual Study: Fundamental rights in the Digital Age*, La documentation française, Paris (2014).

benefits, is low compared to the total value that business derive from them. The critical aspect to solve is if data should be considered as capital rather than as labor. Therefore, it would be useful to compare the different implications of the two different paradigms[66]:

- Data as Capital (DaC) treats data as natural exhaust from consumption to be collected by firms, while Data as Labor (DaL) treat them as user possessions that should primarily benefit their owners.
- DaC benefits AI companies and platforms encouraging entrepreneurship and innovation, while DaL benefits individual encouraging quality and quantity of data.
- DaC will reserve spheres of work to workers where AI will fail for humans, while DaL sees Machina Learning as another production technology able to enhance labor productivity and creating a new class of data jobs.
- DaC encourages people to find dignity in leisure or in activity outside the digital economy, while with DaL data works can be a new source of digital dignity.
- In the DaC scenario, online social contract is based on the exchange of free services for free data, while DaL can be a countervailing power to create a data labor market.

Describing DaL versus DaC as a binary is a simplification, since the production function for data and the AI systems built on top of it is more complex. The economically and socially optimal shares of each factors depend on the details of the production function of data. However, the details of this function are still unmeasured, and data themselves are not purely created by users, since they require firms to track, record and organize them. What is already known, is that the optimal share of user data contributions is not a negligible fraction of the total value of the digital economy. There is, indeed, a misbelief, that some scholars called the elephant in the room of our tech world[67]: AI-driven technologies always involve paid humans working behind the scene in front of a computer screen. Algorithms that feed Artificial Intelligence need to be trained using reams of human-generated examples, the so called process of Machine Learning: this ranges from users granting permission to access data created during

[66] J. Lanier, Simon & Schuster, *Who Owns the Future?* Penguin Books (2013).

[67] M.L. Gray, S.Suri, *The Humans Working Behind the AI Curtain,* Harvard Business Review (2017).

the consumption experiences, to users that provide examples of translations or feedback, or to the creative content displayed on blogs and video/photo sharing platforms, and many others. Therefore, Artificial Intelligence is a misnomer, while it should be called Collective Intelligence[68]. The creation of new human tasks after a technological advancement has always been part of automation's economic history, and this mix of AI and humans is essential for the functioning of this system. This process is called the *paradox of automation's last mile*: as AI improves, it also creates and destroy temporary labor markets for new types of human-in-the-loop tasks. The power of the latest generation of Machine Learning has been its ability to tackle increasingly complex tasks thanks to the better quality and quantity of data. These tasks are impossible to be tackled without ample data, as the learning algorithms require many training examples, that only data generated by humans can give to them. Algorithms are not able to translate languages, understand speech or recognize objects in images, unless they are feed with the right data provided by humans. Thus, data provided by humans can be seen as a form of labor that power artificial intelligence algorithms. Treating data purely as capital is economically and socially irrational. Despite these facts, few users are aware of the productive value of the data they are giving and the role they are playing in enabling Machine Learning. Who is benefiting from this availability of free or extremely cheap data, are the tech giants, in particular Facebook and Google. They are struggling, through lobbyist pressures on institutions, in order to maintain the *status quo*: a DaC equilibrium where users are not even aware of the value their data create. In fact, if users will start to be aware of the value of their data, they would likely begin to ask for compensation for their contribution, and consequently, the share of value that could be captured by businesses as profit, will dramatically drop. This monopsony[69] power, through the inefficient exploitation of labor by concentrated capital, is and has been the primal force blocking the potential productivity gains from a data labor market. Employment and income distribution concerns are strictly connected to this topic, since there is a common fear that AI systems will replace human workers. Indeed, as we already see in the

[68] *What if people were paid for their data?* The Economist (2018).

[69] A monopsony is a market condition similar to a monopoly. However, in a monopsony, a large buyer, not a seller, controls a large proportion of the market and drives prices down. A monopsony occurs when a single firm has market power through its factor of production. The firm is the sole purchaser for multiple sellers and drives down the price of the seller's product or services according to the amount of quantity it demands. (Source: Investopedia).

second chapter of this work, the employment numbers of leading technology giants give little space for optimism: Facebook and Google market capitalization are more than a firm like Walmart, however, they employ 1 to 2 orders of magnitude fewer workers; and the share of income going to labor of Facebook and Google is about 5 to 15 percent, Walmart share is 80 percent instead[70].

In addition, this free data economy model is not ideal even for tech companies themselves. The absence of a market for data is slowing down the productivity growth of the AI, despite its huge potential. Indeed, this system does not reward those with the greatest expertise and context. For example, the data that Facebook collects are of pretty low quality, and for this reason Zuckerberg has hired an additional army of paid workers who are given dedicated tasks, such as content moderation. If AI systems continue to expand into new industries, more valuable data will be needed for the proper functioning of algorithms, and new data work will take place. Thus, the basic idea of this digital revolution is simple and straightforward: when users supply their personal data to tech companies like Facebook and Google, it is a form of labor, and the users should be compensated for it. Much of this work will be passive, as people engage in all sorts of activities online, but some data work will be more active, and include decision process (for examples labelling images).

Nevertheless, these tech companies do not agree to consider data as labor for one obvious reason: it would reduce their bottom line. With their current business model, indeed, using free data to sell targeted online advertising, they raked in a combined 135 billion dollar revenue. Despite that, it would be implausible that tighter margin due to a DaL model will put the Tech giants out of business. Firstly, because the same conditions will apply to all the players, therefore, every firms will see a cut in its margin, but also because the wider economy would grow, with an increased productivity and a fairer distribution of income. Substantially, the big companies would take a smaller share of a larger pie, but, at the same time, their business model would be more sustainable from a political and social justice perspective.

[70] E. Porter, *Your Data is crucial to a Robotic Age. Shouldn't you be paid for it?* The New York Times (2018).

In order to start this difficult and radical process of changing paradigm, from a Data as Capital perspective to a Data as Labor one, with all the implications that this change means, there are three macro area to consider: the legal framework, the technology and a class consciousness.

6.2.1.1 A new legal framework

The first challenge, in order to associate personal data with ownership rights, is to characterize personal data, giving them a flexible definition to ensure that they can be included in an analytical framework that will be able to cover all their characteristics. The legal framework is a necessary condition for the legitimization of a real market of personal data, ensuring the good faith in the transaction. To be property according to the law, data must be:

- an object of desire, with a value resulting from its usefulness or rarity
- an appropriable object
- a transferable object

Data are essentially information, once they are recorded and communicated they take the form of a thing which is legally interesting through its ability to circulate and to be used in different ways. The interest of this 'thing' depends on the value that economic actors are able to extract from it. Since the value of data is continuously evolving, and there is not yes a way to estimate the exact value of them, the legal approach can be trickier. Giving a legal status to data means that they can no longer be considered merely as things, but they must be treated within a legal framework. What would actually change is the treatment of data, since the same data, at different stages, can legally be dealt differently. This legal framework would be used to attribute a status to data depending on the nature of them, since not all data have the same characteristics or experience the same change in value over time. The nature of data distinguishes two broad categories, useful to determine their regime: firstly, data whose importance change over time; secondly, the sensitive nature of the data. As concern temporal consideration, data can be short-lived or long-lasting:

- Short-lived data: they are those data characterized by an instantaneous aspect. This kind of data only have an economic interest at the time they are issued. A sector that illustrate

this category is the betting and gambling one, in which data on bets have an economic value only until the outcome of the bet is known.

- Long-lasting data: the interest of these kind of data is not dependent on the fate of an instantaneous aspect. These are those data that acquire more value through the accumulation of them, since value does not derive from a single data item but from an analytical overview of them. They are personal data, or even anonymized data, that once collected and in a massive amount allow behavioral profiles to be created. The social media and search engines industries are two clear examples of sectors who exploit those data.

As concern the sensitive nature of data, in this case, data are not sorted by its short-lived or long-lasting aspect, but by the actual content of them. Sensitive data are those data that refer to the intimacy of a natural person. The technical capacity to collect these data has increased exponentially during the past decades, and new categories of sensitive information have emerged, together with new economic model for their use. These are the personal data exploited for free by social media and search engines, from which they derive an important source of income. Therefore, their economic importance should mark a change in their legal regime, allowing control of the commercial exploitation of personal data.

Through the application of the solutions provided by ordinary property law to personal data, the possessors will have means of reservation, appropriation and defense on his data. In order to deduce ownership rights to personal data, it should be firstly understood how they can be possessed under a legal perspective. What actually confer the possession are two elements the possessors should hold: the *corpus* and the *animus*. The *corpus* of possession consists in material acts performed by the possessor on the data. These material acts are acts of holding, i.e. the data item must be subject to the power and control of the possessor; and acts of possession, i.e. the economic use of them. These two types of acts prove the existence of the *corpus* of data. As concerns the *animus*, it relates to the behavior of the possessor, i.e. the will to keep them secret. Eventually, the possession of data implies the ownership of them, and the possession gives the possessor ownership rights of these data, in accordance with ordinary property law. Ownership is characterized by its exclusive mechanism, since possession allowed by secrecy or control is a form of expression of the will of the owner to exclude, ensuring that

him can enjoy their property and prevent others from doing so. Therefore, this exclusivity gives to the holders a mean of defense against infringements of their property, with the ability to claim property from a third party.

Once the ordinary property law is applied to personal data, all the property rights will apply. Moreover, ordinary law ownership lasts for as long as the possessions subsists, differently from special ownership laws for intellectual property. In conclusion, if this logic works, it means that personal data may be sold, rented or leased like any property for which the owner has *usus, fructus* and *abusus*. With the current economic model, data are monetized without a fair distribution of the benefits to the primary provider of them, the citizen, instead, they have been unilaterally appropriated. Citizens are the owner of the primary data, and data generated and aggregated, as the provider of the raw material. With this new paradigm the users are at the center of the business model for the utilization of data. This new model will be a factor of growth and it will also create more security for the privacy of personal data. The actors of the data utilization chain are: the citizen or the legal person with data assets, i.e. the primary data provider; the data collector, the data aggregator, i.e. the private or public entity which will manage and analyze the data; the platform, the data retailer or broker and the Data Privacy Officer (DPO); the guarantor instituted with the GDPR, it is responsible for defining the content of usable data.

As concern the type of business model economic actors can utilize, there are various hypothesis according to the existing legal mechanisms. There can be at least two models:

1. The trademark and license agreement: with this model, the users register their data assets in the form of a trademark. Therefore, utilization of this trademark can only be made under a trademark license agreement granted to a third party. The users are paid by a fee based on the volume of the data and their use.

2. The resale right of copyright: this model is backed by the Intellectual Property Code, in particular, Article L.111-1[71], Article L.111-2[72] and Article L.111-3[73]. According to Article L.111-1, data can be considered a work of the mind with intellectual, moral and economic attributes. Article L.111-2 allows the work to be deemed to have been created independently of any disclosure, by the sole fact of the realization, even incomplete. Therefore, users whose data are captured by cookies or other computer mechanisms, would be allowed to consider those data a work of the mind, and it would be companies' responsibilities to inform the users of the collection of their data, for which they have to pay a percentage or a fixed amount on the valuation of the database. Article L.111-3 allows this utilization since it states: "*The incorporeal property right set out in Article L111-1 shall be independent of any property right in the physical object*", therefore, it allows data to be used on many media. However, one adjustment is required to Article L.122-1, that nowadays states: "*The right of exploitation belonging to the author shall comprise the right of performance and the right of reproduction".* The adjusted Article would have in the end also: *"[...] and the right of digital collection".*

Nevertheless, the process needed to reach this change of paradigm is complex and troubled, and it will probably be a result of different gradual reforms, more than the result of a unique radical one. This gradual shift is already started in Europe, with the GDPR, that can be seen as the first step toward the identification of property rights for personal data.

[71] Article L111-1: "The author of a work of the mind shall enjoy in that work, by the mere fact of its creation, an exclusive incorporeal property right which shall be enforceable against all persons. This right shall include attributes of an intellectual and moral nature as well as attributes of an economic nature [..]".

[72] Article L111-2: "A work shall be deemed to have been created, irrespective of any public disclosure, by the mere fact of realization of the author's concept, even if incomplete."

[73] Article L111-3: "The incorporeal property right set out in Article L111-1 shall be independent of any property right in the physical object. Acquisition of such object shall not vest in the acquirer of the object any of the rights afforded by this Code, except in those cases referred to in the provisions of the second and third paragraphs of Article L123-4. These rights shall subsist in the person of the author or of his successors in title who, nevertheless, may not require the proprietor of the physical object to make such object available to them for the exercise of those rights. However, in the event of manifest abuse by the proprietor preventing exercise of the right of disclosure, the first instance court may take any appropriate measure, in accordance with the provisions of Article L121-3."

6.2.1.2 The first step: the GDPR

The General Data Protection Regulation 2016/679, already analyzed previously in details, is considered a step in the right direction. The two fundamental principles for which the GDPR was emanated are: reaffirming the free movement of data and to provide a uniform level of protection within the Union. The GDPR pushes personal data rights into the era of compliance, requiring private and public organizations to demonstrate which data they are processing, and obliging them to comply with a number of guiding principles during their practices.
The main novelties introduced in the GDPR that goes toward the direction of granting ownership rights to personal data are the following:

- Article 17: it consecrates personal data as intangible property, since the users can request to have their data erased, the so called "Right to be forgotten"[74]. According to this principle the digital citizens can request an end to the divulgation of their data when these data are no longer necessary for the purpose for which they were collected, or because they withdraw their consent.
- Article 20: it consecrates the principle of data portability. Data controller cannot oppose to the transfer of personal data to other systems.
- Article 77 to 81: they provide the right for every data subject to lodge a complaint with a supervisory authority, if they consider that the processing of their personal data infringes the Regulation.
- Article 82: it recognizes the right to compensation to any person who has suffered material or non-material damage as a result of an infringement of the GDPR. Therefore, it is the closest existing norms to recognize a *fructus* right for data users.

Overall, even if the GDPR does not create a real personal data ownership regime, allowing citizen to receive payment for their personal data, it is a first step toward a more genuine governance of data. The role played by businesses shifts from *owners* of data to *guardians* of

[74] The 'right to be forgotten' derives from the case between Google Spain and *Agencia Española de Protección de Datos* in 2014.

them, and individuals are now at the center of attention, although the value of their personal data is still out of their control.

6.2.1.3 Technological solutions

Besides the legal aspect, there is a technological challenge to be addressed in order to give data users the possibility to monetize their personal data.

The first critical issue to address is how creators of digital data (i.e. data users) can prove their authenticity, establishing that the data they create belong exclusively to them. This is an essential condition for the right functioning of a data rights market. In addition, users also need a tool that make their data available for trade.

As concerns the person's online identity, the law currently states that IP (Internet Protocol) addresses are personal data. An IP address is a series of numbers associated with each physical element connected to the Internet (computer, smartphone, tablet etc.). However, it is impossible to establish a credible correspondence between the IP address of a device and the user who is using it, since it only indicates a terminal with which the users connects to the Internet, that can be borrowed or stolen by another person. Therefore, IP address can only identify a person's way of accessing the Internet, while it cannot be the solution for ascertain the identity of data users. In order to allow a person to benefit from their data it is therefore necessary to be authenticated by something different. A possible solution is proving authenticity through electronic signature. An electronic signature is based on cryptographic mechanism, and it establishes a relationship between a person and a digital document, valid at least as a handwritten signature between a person and a paper document. Electronic signature through cryptography is based on the process of *public key cryptography*, known also as asymmetric cryptography. Public key cryptography involves two mathematically related keys, a private key and a public one. The private key is the key that allows data to be encrypted, while the public key is needed to decrypt the data. The private key should be kept secret by its owner, while the public one can be disclosed. Of course, without the private key, it is impossible (or at least very costly and difficult) to reconstitute the corresponding private key. Therefore, with this system, encrypted data are secured from third party interference. In addition, in order to prove that to a certain public key corresponds the private key of a particular person, the public key is included in a digital certificate that contains information about the identity of the data owner. This

certificate is also signed by a trusted authority, which proves the authenticity and integrity of the encrypted content. Therefore, this is a good way for data users to prove that they are the owner of their digital data, and that these data have not been altered since the certificate was signed.

As concerns the tool that can provide a solution for the economic model that we are discussing, a possible solution can be found in the blockchain technology. A blockchain is a distributed autonomously decentralized database, and it is the only place in the cyberspace that is easily accessible, that everyone can view and nobody can change. Indeed, the blockchain can be seen as a registry open to everyone, duplicated, automatically and constantly on many servers, and the digital data on them would be located in chained blocks. Once included in the blockchain, a block cannot be changed, since to change a block it is necessary to change all the previous one. This is why it is called a distributed ledger technology, and it is considered inviolable. Thus, through the blockchain, data users wishing to trade their data can certify the ownership of these data, until a transaction certifies that they have sold them, in which case they do not belong anymore to them but to the person or business who has bought the data. The contracts suitable for personal data trade are the so called *smart contracts*, and they should be the right solution to manage the market of digital identities. Smart contracts are tamper-proof and non-erasable contracts that exploit the technology of the blockchain, and they are written in computer language. Ethereum is one of the blockchains that manage these kind of contracts, in which payments are made in *ethers*, a cryptocurrency that can be converted into euros or other currencies. Nevertheless, other easier to use blockchains can be created on the Ethereum model, with simpler language and with better performance, suited for the market of personal data. In this economic model, tech giants, and internet companies in general, will not anymore collet data directly at the source (the data users), with or without their consent. Instead, data will be available on dedicated servers, subject to certain payment conditions that the owners will specify in a smart contract. However, in order to properly work, all the process should be automated, and Interne tools like browsers, should be modified to allow users to place their data directly in a blockchain of their choice. At this point, people is free to choose which data want to market, and which one want to keep confidential. In the case that data have been acquired illegally, when there is no trace of the transaction on the blockchian, it would be a case of theft or plagiarism. Indeed, a transaction on a blockchain always leaves a track of the transaction

through a publication. In this scenario, data users have full responsibility of the retaining and protection of their data, backed by an ownership right guaranteed by the State.

6.2.1.4 Data workers' union labor

The third and last challenge to be addressed in this data market reform, is the lack of what we can call a global "class consciousness" of data users, which should start to consider themselves more as *data workers*. Nowadays, users do not even realize how valuable their data are, and even if they do, an individual data worker lacks bargaining power against tech giants such as Facebook and Google. In addition, it would be extremely difficult to manage all their data privately in order to trade and monetize them. To realize gains from data as labor, data users need organization to manage their data and ensure their quality. Indeed, it would be costly and time consuming to handle all the complexities of the digital systems. The only type of organizations that can fulfill these tasks are unions. Data labor unions would create a data labor movement, which would have three main roles:

1. Collective bargaining
2. Data quality certification
3. Career development

This triple roles correspond exactly to the roles unions played during the Industrial Age[75]. Prior to the emergence of unions, workers' wages remained steady for decades, despite the technological improvements. Thanks to unions, and their ability to counter the monopsony power of industrialists, not only workers started to have higher salaries, but also the overall productivity increased. Indeed, unions guaranteed the quality of the work produced by their workers and helped them to learn new skills required by the technological progress. As of today, we should consider how the monopsony power of Internet firms is holding down wages for data workers at zero, or at the value of the services they provide to users. This system, as we already said, is suppressing the quality and quantity of data that are feeding the artificial intelligence's

[75] E. Posner, G. Weyl, *Data Workers of the World, Unite*, The blog of the Stigler Center at the University of Chicago Booth School of Business (2018).

algorithms. Through these organizations, people would have the means to engage in collective actions, and this may awaken their role as data workers.

Moreover, a data labor union would have an advantage that no traditional union have ever had: an international market completely unaffected by borders and government regulation. Therefore, the network effects that characterize digital monopolies would eventually work against them. If a data unions reach a critical mass, they would become a sort of gatekeepers of people data. Then, they could organize strikes, that would combine both labor stoppages and boycotts, since data users are simultaneously workers and consumers. For example, during a strike against Facebook, the social network would lose access both to data, the labor side, and to advertising revenues, the consumer side.

Finally, data labor union may be a tool that will foster competition, breaking the monopoly of data giants, since they would have the power to share data between many different Internet companies, rather than accumulating them in one place.

There would be, obviously, also downsides of unions, as there were in traditional ones. They, for example, could start to abuse their authority. However, these problem would come in a second phase when unions will be mature, while initially, the gains would certainly exceed the losses. Nowadays, there are already startups trying to develop these services: companies like the UK startup "digi.me" allows users to upload and store their data in a single app, or the European Union based "Wibson", and the US startup "Datacoup", all of them promise their users the possibility to trade their data with interested parties for money or discounts.

In conclusion, the techno-legal solution proposed above for a data market reform require the intervention of both the private and public sphere, private for what concern the technological advancement needed and public for the legislative reforms. However, there are at least two important limits at the present time:

1. The territoriality of the law: data are geographically agnostic, while the law has territorial limits.

2. The identity problem on a blockchain: the blockchain creates a transaction systems fully pseudonymous, if not anonymous. Since the identity is not verifiable, a person can always argue that someone has robbed their data or that they lost their private key for

the decryption of their data. These two aspects would increase the instability and uncertainty related to the contract.

Nevertheless, despite the technical and legislative difficulties of such a radical reform, the necessity of this approach grows as data accounts for more of the indispensable capital in the economy. Recognizing ownership rights to data could mitigate the mass unemployment that the global economy will face because of the further development of the artificial intelligence. This approach would recognize what people contribute even if they do not actually work for a company, and at the same time, it would give better quality data to the latter.

CONCLUSION

The advent of data economy has produced disruptive effects on every industries and on the overall society. Its development has been so rapid that it was difficult, and still is, to understand all the direct and indirect effects on the economy. The aim of this work was to analyze the broad picture of data economy, trying to connect the different dots in a single scheme.

We have seen, firstly, the intrinsic characteristics of data and the impact the data economic have on the output of the world economy. A revolution that is just began: data economy is doubling in size every two years. Consequently, the challenges we are now facing are only at their early stage. The European Union, as we already told in the second chapter, see the data revolution as an opportunity which can strengthen the Union economy, after a decade of low growth and stagnation. Indeed, the Digital Single Market strategy is one of the seven pillars of the Europe 2020 Strategy, which sets objectives for the growth of the EU by 2020. Together with the opportunities, the European Commission is also valuating carefully the risks that data economy has brought. The General Data Protection Regulation (GDPRD), issued in May 2018, is a first step to regulate the massive amount of data handled by firms, focusing on the privacy concerns. Under an economic perspective, we have analyzed in the third chapter, the difficulties in determining the true value of data, caused by the non-fungible nature of them. Their value, indeed, depends on a series of contingencies: their authenticity, the type of information they bring, their structure and also the way they are collected and analyzed. Therefore, in order to exploit their business potential, firms need a technological and organizational structure. The pioneer of data economy, that have understood before anyone else how to exploit and capitalize data, are now tech giants that dominate the global market. These platforms-companies and data-aggregators capitalize on individual data, and they become monopolists in their respective industries, and currently, the most valuable firms in the world. They are the exemplification of what is the potential disruptive power of the data competitive advantage. They had replaced in one decade the old dominant corporations, that were mainly oil and energy companies. At the same time, they have also highlighted, all the pitfalls of the new data-driven business models. Pitfalls that we have analyzed in the fourth and fifth chapter, privacy and competition concerns. The right to privacy is a fundamental right granted to every individual, and the right to data protection is strongly connected to it. However, data protection law has evolved through a

separation between the two, recognizing the increasing importance to the right of data protection, now considered a fundamental right itself. As concerns competition, in the fifth chapter, we considered the point of view of competition authorities: firstly, they have to understand how Big Data affect the business activities and then how businesses can exploit them for anticompetitive conducts. We come to the conclusion that antitrust laws alone do not have the sufficient tools to address the challenge arisen from platform monopolists, while, a strict cooperation between competition, privacy and consumer protection authorities is needed. Particularly, through the three specific case studies on different legal investigation on Facebook and WhatsApp in Europe, either by the European Commission and by the Italian and German national competition authorities, we have seen how the current sanctions and remedies are not sufficient to contain the anticompetitive conducts of the big tech firms.

Eventually, in the final chapter, after the examination of all the critical aspects of the current legislation, we analyzed the possible solutions, toward a new economic structure called Digital Capitalism. The name Digital Capitalism needs to emphasize the discontinuity in relation to the current model, which we can be called Digital Feudalism. In the era of digital capitalism, the economy would be based on a mutual benefit approach, and not anymore to a zero-sum game logic. The basic and radical change on which Digital Capitalism must start is considering data as a form of labor. Data as Labor (DaL) treats data as user possessions that should primarily benefit the owners of the data, and encourage the creation of a data labor market. A data labor market potential benefits extend beyond the economic fairness, because it would also boost artificial intelligence productivity thanks to a better quality of the data. Property rights have protected and empowered individuals for millennia, evolving as technology does. As the printing revolution brought intellectual property rights and the Industrial Revolution the patent system, the digital revolution must bring the right to personal data ownership, therefore including the three classic elements of property rights: *usus*, *abusus* and *fructus*. In the final part of the work we focused mainly on the latter of the three elements, i.e. how digital users can monetize their data, analyzing the legal and technological challenges and possible solution suggested by the doctrine. This vision is backed by many scholars and experts of the data economy, which support the ideal of a decentralized Internet, where individual digital identity has a value, and it is not reduced to a flow of data. In this scenario people would get paid for their contribution to the data economy and to the improvement of the algorithms that power the

artificial intelligence. The technical and legal challenges are difficult to overcome, but this is the right path to follow, in order to reach a fairer and more productive society.

BIBLIOGRAPHY

A. Cervone, *Unfair Contract Terms and Sharing of Data with Facebook, Towards a Better Protection of Social Media Users: The WhatsApp Cases* (2017).

A. Chirita, *Data-Driven Mergers Under EU Competition Law,* Durham University (2018).

A. Ezrachi, M. Stucke, *Algorithmic Collusion: Problems and Counter-measures,* OECD Note (2017).

A. Ezrachi, M.E. Stucke, *Virtual competition – The Promise and Perils of the Algorithm-Driven Economy,* Harvard University Press (2016).

A. Schlosser, *You may have heard data is the new oil. It's not,* World Economic Forum (2018).

A.M. McDonald, L.F. Cranor, *The Cost of Reading Privacy Policies,* I/S Journal of law and policy for the information society (2008).

Accenture, *Exploring Next Generation Financial Services: The Big Data Revolution (2016).*

B. Schneier, *Data and Goliath – The Hidden Battles to Collect Your Data and Control Your World,* W.W. Norton & Company (2015).

B. Van Der Sloot, *Legal Fundamentalism: is data protection really a fundamental right?* (2017).

Building a European Data Economy, Communication from the Commission to the European Parliament (2017).

Conseil d'État, *2014 Annual Study: Fundamental rights in the Digital Age,* La documentation française, Paris (2014).

D. Daniel Sokol, *Antitrust and Regulating Big Data*, University of Florida Levin College of Law (2016).

E. Brynjolfsson, McAfee A., *Big Data: The Management Revolution*, Harvard Business Review (2012).

E. Porter, *Your Data is Crucial to a Robotic Age. Shouldn't You Be Paid for it?* The New York Times (2018).

E. Posner, E. Glen Weyl, *Radical Markets – Uprooting Capitalism and Democracy for a Just Society* (2018).

E. Posner, G. Weyl, *Data Workers of the World, Unite*, The blog of the Stigler Center at the University of Chicago Booth School of Business (2018).

E. Posner E. Glen Weyl, *Want our personal data? Pay for it*, The Wall Street Journal (2018).

European Commission Communication to the European Parliament, *A New Skills Agenda for Europe - Working together to Strengthen Human Capital, Employability and Competitiveness* (2016).

G. Cattaneo, G. Micheletti, D. Osimo, K. Jakimowicz, *First report on policy conclusions*, IDC Italia (2018).

G. Koenig, *Leaving the Data Dark Ages*, Project Syndicate, (2018).

G. Sivinski, A. Okuliar, L.Kjolbye, *Is Big Data a Big Deal? A Competition Law Approach to Big Data*, European Competition Journal (2017).

Guida al nuovo regolamento europeo in materia di protezione dei dati personali, Garante per la protezione dei dati personali (2016).

How the power of data will drive EU economy, The European Data Market (EDM) Monitoring Tool Report (2017).

I. Graef, *EU Competition Law, Data Protection and Online Platforms*, Wolters Kluwer (2016).

J. F. Rayport, J. Sviokla, *Exploiting the Virtual Value Chain*, Harvard Business Review (1995).

J. Lanier, *Should we Treat Sata as Labor? Moving Beyond Free,* American Economic Association (2018).

J. Lanier, Simon & Schuster, *Who Owns the Future?* Penguin Books (2013).

J. M. Cavanillas, E. Curry, *New Horizons for a Data-Driven Economy, A Roadmap for Usage and Exploitation of Big Data in Europe,* Wolfgang Wahlster Editors (2016).

J. E. Short, S. Todd, *What's Your Data Worth?,* MIT Sloan Management Review (2017).
Key success factors for digital transformation in the banking industry, IDC Report (2015).

L. Columbus, *89% of B2B Marketers Have Predictive Analytics on their Roadmaps for 2016,* Forbes (2016).

L. Lèger, *My data are mine, Why we should have ownership rights on our personal data,* Report GenerationLibre (2018).

L'Economia dei Dati - Tendenze di Mercato e Prospettive di Policy, ITMedia Consulting (2018).

M.E. Stucke, A.P. Grumes, *Big Data and Competition Policy,* Oxford University Press (2016).

M.L. Gray, S.Suri, *The Humans Working Behind the AI Curtain,* Harvard Business Review (2017).

Mauree K. Ohlhausen, Alexander P. Okuliar, *Competition, Consumer Protection, and the Right Approach to Privacy* (2015).

N. Howe, Forbes, *A special price just for you,* (2017).

OECD, *Data Driven Innovation - Big Data for Growth and Well-Being* (2015).

OECD, *Human Capital – The value of People,* OECD Publishing (2005).

OECD, *OECD Digital Economy Outlook,* OECD Publishing (2017).

P. Jones Harbour, T. Isa Koslov, *Section 2 in a Web 2.0 World: An Expanded Vision of Relevant Product Markets,* Antitrust Law Journal (2010).

P. Lugard, L. Roach, *The Era of Big Data and EU/US Divergence for Refusals to Deal,* Antitrust Magazine (2017).

R. Avent, *A Digital Capitalism Marx Might Enjoy,* MIT Technology Review (2018).

R. O'Harrow Jr., *The outsourcing of US intelligence raises risks among the benefits,* The Washington Post (2013).

S. Galloway, *The Four. The hidden DNA of Amazon, Apple, Google and Facebook,* Bantam Press (2017).

S. Lund, J. Manyika, J. Bughin, *Globalization is Becoming More ABout Data and Less About Stuff,* Harvard Business Review (2016).

T. Vasudha, G. Arvind, *The value of data,* World Economic Forum (2017).

The digital universe of opportunities: rich data and the increasing value of the IoT, International Data Corporation (IDC) (2014).

The World's Most Valuable Resource is no longer oil, but data, The Economist (2017)

V. Bagnoli, *The Big Data Relevant Market As A Tool For A Case By Case Analysis At The Digital Economy: Could The EU Decision At Facebook/Whatsapp Merger Have Been Different?,* 12th ASCOLA Conference Competition Law For The Digital Economy.

What if people were paid for their data? The Economist (2017).

THESIS' SUMMARY

Data economy and data-driven innovation (DDI) are key pillars in the economy of the 21st century. The increasing digitalization of every kind of socio-economic activities and the decline in the cost of data collection, storage and analysis, has led to a new kind of economic paradigm, centered on the use of huge volumes of data, the so-called Big Data. Data are considered as the new world's most valuable resource, the oil of the new century and the engine of the digital economy. The innovative characteristics of data are conventionally summarized in four Vs:

1. Volume: the quantity of data created every year is impressive and it is exponentially growing year over year;
2. Velocity: data are collected and analyzed quickly, even in real-time;
3. Variety: there are many different types of data,
4. Value: data are becoming a valuable asset for firms and a new factor of production.

We can talk of a data revolution, or a new Schumpeterian wave, that is fostering new industries, processes and products, while replacing the old ones. However, as always happens when there are such radical changes in the economy, there are also challenges to be addressed. The capability of private firms to collect and analyze huge amount of personal data, indeed, has created a wide array of policy issues, ranging from privacy and consumer protection to competition concerns. Particularly, governments will need to anticipate and address the disruptive effects of Big Data on the economy and overall well-being. For the moment, they have not been capable of keeping pace with the technological challenges of this new economy.

In this work, I have analyzed the data economy from different point of views, with a particular focus on its critical aspects, and eventually, on how they may be addressed. The final aim of the work is to analyze the broad picture of data economy, trying to connect the different dots in a single scheme.

In the first chapter (Introduction to data economy), I have analyzed more in details the intrinsic characteristics of data with some numbers, trends and statistic of the global data economy. The digital universe is expanding at a tremendous pace, an expansion that include not only people and enterprises online, but also all the smart devices connected to the internet. This ever-

expanding universe is doubling in size every two years, and by 2020 the amount of data circulating in this universe will reach 44 zettabytes (or 44 trillion gigabytes), the world produces 2.5 quintillion bytes a day, and 90% of all data has been produced in just the last two years. As everything is going online, there are always new ways of exploiting data, the so-called *data-driven innovation*, defined by the OECD as "the use of data and analytics to improve or foster new products, processes, organizational methods and markets". This kind of innovation has allowed the transformation of data into assets, allowing enterprises to learn deeply the customer's behavior, to speed the business cycle, to flatten the organizational structure and broadly to make better predictions and smarter decisions. Indeed, data do already have a huge impact on the output of the world economy, the Mckinsey Global Institute estimates that data flows have boosted the world GDP by more than 10%. However, the abnormal bulk of data circulating in our economy is exploitable only for the corporations capable of managing this large amount of data, and the best in exploiting them are now the companies that are dominating the world economy. The corporations that had already exploited the potential of Big data at its best in the last decade, are nowadays the most valuable firms in the world. They are platforms-companies and data-aggregators: Alphabet, Facebook, Amazon and Microsoft, each one of them capitalize on individual data. These new monopolists of the 21st century are generating significant value for billions of people, that can benefit of free services, but very few reap the economic benefits. If, indeed, we compare the market cap:workforce ratio (market cap per employee) of General Motors with the one of Facebook, this situation emerges clearly: in 2016 GM created economic value of approximately 231 thousands dollar per employee, while the Zuckeberg's platform generated 20.5 million dollar per employee, i.e. almost a hundred times the value of GM. This is the disruptive power of Big data, and particularly the power of the *data-network effect:* using data to attract more users, who then generate more data, which help to improve services, which in conclusion attract more users.

As concerns the analysis of the economic characteristics, data can be classified as non-rivalrous, excludable and intangible good. The non-rivalry nature implies that data should be used as much as possible, since the marginal costs of an additional use of them is zero. The second characteristic, the excludability, is an important precondition for ensuring the possibility of revenues for the data holders, since it allows them to exclude others from using their data. Data is a capital with increasing returns, since it can be re-used as input for further production.

Eventually, data are input for multiple purposes and their value depends on complementary factors related to the capacity to extract information.
An important distinction is between personal and non-personal data:

- Personal data: any information relating to an identified or identifiable living individual (data subject). Personal data that has been rendered anonymous in such a way that the individual is not or no longer identifiable, is no longer considered personal data, but the anonymization must be irreversible. Examples are: name and surname, home address, email address, Internet Protocol (IP) address, a cookie ID or an identification card number.
- Non-personal data: conversely they are all the others, in particular they refer to machine-generated data, especially raw data created through sensors.

As concerns the non-personal data, there is less brightness in term of legislations compared to the primary one (which are protected as a fundamental right), since the European Commission are still working on finding the optimal solution for them. Until now, the Commission has proposed a new Regulation aimed at removing obstacles to the free movement of non-personal data in an official EU Communication, that together with the GDPR will ensure a comprehensive and coherent approach to the free movement of data in the EU.
Eventually, in the last paragraph of the first chapter, I analyzed the social implication of data. Data can directly contribute to inclusiveness, development and the well-being of citizens. The key issue is that a society with a high level of data innovation may still be unequal, with a few actors concentrating most of the power and capturing most the benefits. At the same time, data can have also positive social impact, thanks, for example, to the exploitation of them in science. In fact, the ability to access and combine data coming from different sources, along with the increasing power of analytics, have allowed researchers to conduct complex experiments and potentially achieve scientific discoveries and better predictions leveraging massive data sets and algorithms, especially in the fields of medical, social and environmental sciences.

In the second chapter (European Union Data Economy), I narrowed the focus on the data economy only in Europe and on the projects of the European Union to exploit this new technological revolution, through the analysis of the Digital Single Market Strategy and the

European Data Market Monitoring Tool. Among the seven pillars of the Europe 2020 Strategy, which sets objectives for the growth of the European Union by 2020, there is the European Commission's Digital Agenda. The main goal of this Agenda is to develop a digital single market, in order to generate smart, sustainable and inclusive growth in Europe. The digital revolution offers the opportunity to strengthen Europe's economy and society which have suffered, and it is still suffering, from a decade of low growth. However, this revolution needs also to be incentivized by the right legislations and policies. The McKinsey Global institute estimates that Europe has captured so far only 12% of its potential from digital technologies and it is lagging behind the United States, even though the European market has the potential to be the largest digital market in the world in size and value, only if appropriate policy and investment decisions will be made. Analyzing more in details the DSM Strategy, it has identified three main emerging challenges: (1) to ensure online platforms can continue to bring benefits to our economy, (2) to develop the European Data Economy to its full potential and (3) the protection of Europe's asset through cybersecurity. Moreover, it is built on three main pillars, and each of them has its own policy proposals:

1. Pillar I (Access): Better access for consumers and business to digital goods and services across Europe.
2. Pillar II (Environment): creating the right conditions and a level playing field for digital networks and innovative services across Europe.
3. Pillar III (Economy & Society): maximizing the growth potential of the digital economy

Further narrowing the focus only on the Italian market, in 2016, the Italian data market value was of 4.6 billion euro, expected to reach in 2020 a value of 5 to 9 billion euro (from Baseline scenario to High Growth scenario), while the value of the Italian Data economy was 28 billion euro in 2016. The Italian market share compared with the overall European economy was about 8% in 2016, and it is expected to grow to a 9% in the best-case scenario.
However, the impacts of the data economy on the overall economy is only less than 2%, highlighting that much has yet to be done in Italy to exploit the full potential of this new economic revolution.

The third chapter (The value of data), focuses on how data's economic value can be computed, through the analysis of the Data Value Chain. Data, despite their increasing value for competitive success, are difficult to price because of their non-fungible nature. The economic and social value of data depends on many different factors: the veracity (i.e. the authenticity and trustworthiness), the type of information they bring, the structure (structured or unstructured) and also the way they are collected. Their value can be better analyzed through a value chain, more precisely a Data Value Chain, that describe the flow of information through a series of steps, necessary for the economic valorization of the data sets. The data value chain is an important instrument for coordinate the different activities regarding an information system. A value chain is made up of a series of subsystems each with inputs, transformation processes, and outputs, and it can be applied to information flows to understand the value creation of data technology. In a Data Value Chain, information flow is described as a series of steps needed to generate value and useful insights from data. The following are the key high-level activities:

- Data Acquisition: the process of data gathering, filtering and selection, before they are put in a data warehouse (or any other storage solution) on which data analysis can be carried out.
- Data Analysis: the phase that transform the raw data into something meaningful for the purpose of the decision process. The aim of the analysis of data is to highlight relevant data, synthetizing and extrapolate hidden and useful information with high potential from a business decision perspective.
- Data Curation: the active management of data during their life cycle. This process aims to validate data, to ensure the necessary data quality requirements for their effective use.
- Data Storage: the persistence and management of data in a scalable way that satisfies the needs of applications that require fast access to the data.
- Data Usage: the final step that covers all the data-driven business activities that needs access to data. It is the output of the process, the reason why of all the previous phases. Data usage in business decision-making enhance competitiveness through reduction of costs and increased added value.

Corporations can extrapolate value from data through two distinct approaches: the direct commercialization of data and the approach based on data driven innovation. In the first case, with the commercialization of data, corporations aim to monetize the value of data as quickly as possible, with their diffusion and circulation. Data Driven Innovation (DDI), instead, is an iterative innovation process consisting in the exploitation of any kind of data inside the company (*in house*), in order to create new value. Data driven innovation is leading to the development or improvement of new products and services, the improvement of the marketing strategies or more in general of the business decision process. Eventually, the value of data lies in their uniqueness and modality of use. Trying to understand the value of each bit of information that should be gather, combined and analyzed is difficult mainly because corporations themselves cannot fix the value of their data until they are able to specify and understand how to use them. Potentially, data within an organization can represent only a small percentage of the total revenue, but, at the same time, they can also become a key success factor for the future growth of the business.

In the fourth (Privacy) and the fifth (Competition) chapters the focus switch on the critical aspects of the data-driven innovation disruption, and the policy issues it has brought. Specifically, in the fourth I have analyzed the privacy implications of the mass surveillance business model that dominates the new economic landscape. Privacy debates have always been at the center of a global controversial debate, in particular for what concerns the government surveillance on our life, and the question of how much we can sacrifice of our privacy for public safety concerns. In the last decade the amount of surveillance from governments and corporations has increased enormously. They both gather, store and analyze an enormous amount of data that we create. The "privacy paradox" consists in the desire of the majority of people to protect their personal information, while, at the same time, giving away this data explicitly for free. Corporate surveillance and government surveillance are not separate, they are intertwined. What has changed over the last years that has led to this mass surveillance era is that our interactions and conversations, and all our actions either online and offline, are no longer ephemeral as once were through most of our history, in other words, everything is recorded and permanently available. As technology improved and prices dropped, corporate surveillance has grown from collecting as little data as necessary to collecting as much as possible and governments broadened their surveillance as well. The common denominator and

primary goal of all this corporate internet surveillance is targeted advertising, the so called surveillance-based marketing. Therefore, considered all that, legislators are concerned to increase transparency in the data economy. Transparency is essential to any open and free society and it is a necessary condition for another important principle: accountability. Indeed, there is an inverse relation between transparency and power in the hands of corporations and institutions, the more transparency the less the power, and vice versa. A step toward a more transparent data economy is the General Data Protection Regulation (GDPR), entered into force the 25 May 2018, replacing the Data Protection Directive 95/46/EC, adopted the 24 October 1995. GDPR aim is the harmonization of data privacy laws across all European Member states, in order to protect and empower all European Union citizens' data privacy and to reshape the way corporations across the continent approach data privacy. The three pillars of this new regulation are transparency, responsibility and accountability, for the data controllers and processors. The GDPR emphasize the rights of the individual, the duties of the corporations and public administrations, and the relations between them, in order to tackle all the most critical aspects that concern the data economy. It creates new figures and roles, and for the first time it defines a distinction between the 'Data Controller' (Art.4, n.7), i.e. "the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data"; and the 'Data Processor' (Art.4, n.8), i.e. "a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller." The actual reform of the GDPR lies in the new set of responsibilities of the data controller and data processor. It has introduced new duties for who control and process personal data:

- Article 30 - *Records of processing activities*: in the case of enterprise or organization employing more than 250 persons, data controller shall maintain a record of processing activities under its responsibility.

- Article 33 - *Notification of a personal data breach to the supervisory authority*: as the title says, this article concerns the duty of data controller and data processor to notify the personal data breach to the competent supervisory authority.

- Article 34 - *Notification of a personal data breach to the data subject*: it is the same obligation of article 33, but in this case the notification concern the individual persons damaged by the data breach. The notification is required when the leak of data is likely to result in a high risk to the rights and freedoms of the natural persons involved.

Moreover, the GDPR abrogates the *country of origin principle* of the old Data Protection Directive, and expands its field of application, ratifying the applicability of the Regulation "to the processing of personal data in the context of the activities of an establishment of a controller or a processor in the Union, regardless of whether the processing takes place in the Union or not." A further framework that distinguish the GDPR from the old Directive is the one concerning remedies and sanctions. It foresees heavier administrative fines for unlawful corporations: the maximum fine is of 20 million euro, or up to the 4% of the total worldwide annual turnover of the preceding financial year of the accused corporation. In addition, in Article 82, it is enunciated the right of data subject to compensation when there is a material or non-material damage as a result of an infringement of the Regulation.

The fifth chapter focuses on the potential and actual harms to competition of data economy, and the challenges competition authorities must overcome to regulate the new industries arisen from the data revolution, with a particular focus on the European antitrust law. Data, generating revenue from user-data-based profiling and advertising, are a source of innovation that can affect the traditional levers of competition. The competition in the digital economy has its own characteristics, such as "winner takes all" situation, network effects, two-sided or multi-sided platform, fast paced innovation and huge investments. However, the existing antitrust regulations were conceived for this pre-data economy. Therefore, anticompetitive behaviors of big tech companies remain mainly unpunished. Besides the anticompetitive results, there are also many pro-competitive benefits for the economy created by the exploitation of Big Data. As a matter of fact, unprecedented consumer benefits have already been realized through the use of big data. The first and most obvious benefit to consumers has been the ability of firms to offer heavily subsidized, usually free, services, as consumers give those firms permission to monetize consumer data on the other side of their business (two-sided platform). Moreover, online data are used by firms to improve and refine products and services in a number of ways,

and to develop brand new innovative product offerings. The complications for antitrust authorities derive from the difficulty of prove how the exploitation of Big Data create a barrier to entry, while, data-driven markets are considered markets with low entry barriers, where little user data is required as a starting point for most online services. Firms can enter with innovative new products that address new customer needs, and thanks to its innovative solution, they can quickly collect data from users, which they can use to further improve the product and finally compete with the incumbents on equal terms. There is an ongoing debate on how to tackle these inefficiencies of the data economy, between who is in favor of more proactive antitrust enforcement, and others considering antitrust inappropriate for regulating big data. According to the latter, data economy should be regulated by consumer protection laws. There is also a third possible way, where antitrust and consumer protection laws cooperate together. Indeed, the relationship between antitrust law and consumer protection law is usually seen in terms of complementarity: antitrust focus on the supply side of the market, worrying that consumers can benefit the widest range of products/services at the lowest price possible, through fair competition between firms; consumer protection law, instead, focus on the demand side of the market, worrying that consumers are able to exercise consciously the choices that competition allow them to freely exercise. Hence, antitrust law is not interested in how much the choice of consumers are mindful and responsible, since this is the role of the consumer protection law. However, it is antitrust authority's duty understand how the exploitation of data can lead to illicit anticompetitive conducts, in order to stop or at least mitigate them. Antitrust authorities should follow two fundamental steps in order to analyze the relationship between anti-competitive conducts and Big Data: firstly, they need to understand how Big Data affect the business activity, and secondly, they need to understand how and if these business activities can exploit big data for anti-competitive behavior.

As concerns the first point, Big Data can give rise to network effects, and these phenomenon plays an important role in an antitrust analysis. Technology giants have always benefited from network effects: the more users sign up on Facebook, or the more users utilize the Google query, the more attractive and convenient (in term of quality of the service) signing up become for other users. Moreover, data are characterized even by *extra network effects*, also called *data-network effects*: using data to attract more users lead to more data generated by those extra users, which in turn help to improve the service and, eventually, the improved service attract even more users. However, the presence of network effect itself does not automatically results

in anticompetitive behavior. There are different types of network effects that can come into play in online platform that exploit big data. They can be categorized in three groups:

4. Direct network effects: they occur when a product or service becomes more valuable to a single user as more people use that product or service.

5. Indirect network effect: they occur when more users make the product or service more valuable, but not through the direct interactions between users, as in the case of direct network effects.

6. Cross platform effects: they occur to firm operating a two-sided platform, where more users on one side of the platform makes the platform more attractive to users on the other side of the market.

The network effects are based on the *feedback loop*, with which big data can lead to economies of scale. There are two ways this feedback loop can generate economies of scale: the user feedback loop, when a platform that gains more users can also collect more user data (feedback indeed), leading to better insights into customers and their needs; the monetization feedback loop, which focus on the monetization from advertising on online platforms, and it claims that with more users on the platform the firm is better able to target and sell ads. The network effects, together with economies of scale, characterize data-driven market, and they usually lead to a *winner takes all* result, creating insurmountable barriers to entry.

The second point is understanding how the results of the inferences made from data can be exploited by businesses for anti-competitive ends. The international debate on this topic is discussing the possibility that the exploitation of Big Data facilitates anticompetitive conducts, such as:

- Collusions: there are certain kind of algorithms which have proven to be used by firms in an anticompetitive way, the "pricing algorithm". Pricing algorithms are used by firms to determine the price of the goods or services sold, after the collection and analysis of a large amount of relevant data about the market. With those algorithms a firms can react instantaneously to price movements by competitors. They pose serious challenges

for antitrust authorities, since they may facilitate collusion and complicate detection of unlawful agreements. Collusion may occur in different forms, but the outcome is always the same: one party, the consumers, is eventually disadvantaged. Before the advent of Big Data, humans have for many years been the moving force behind these practices, colluding on every kind of economic activity. Nowadays, Big Data are providing new and more sophisticated means to foster collusion.

- Algorithmic consumer price discrimination: the exploitation of massive amount of users' data by Internet platform, together with the already mentioned pricing algorithms, lead to another anticompetitive conduct: price discrimination. Differently from the collusion scenario, price discrimination is a competitors' unilateral strategy, and the outcome, instead of uniformly higher price, in this case, on the contrary, is personalized prices. Price discrimination, indeed, is a price strategy that charges to customer different prices for products or services whose costs are the same, according to their economic possibility.

- Concentrations: concentration and competition are inversely proportional, if one of the two decrease the other increase, and vice versa. The harm caused by the erection of strategical barriers to entry in the data market does not result in higher prices, as usually happens in this cases in other sectors, but rather in a loss of quality, innovation and privacy.

After the theoretical aspects of anticompetitive conducts in data markets, I analyzed three specific case studies: (1) European Union Commission investigation on the Facebook's acquisition of WhatsApp in 2014, (2) the AGCM procedure against Facebook for the infringements of the Consumer Code, and (3) the current investigation of the Bundeskartellamt (the German Competition Authority) on Facebook for abuse of dominant position.

Eventually, in the final chapter, after the examination of all the critical aspects of the current legislation, I focused on the possible solutions, toward a new economic structure called Digital Capitalism. The name Digital Capitalism needs to emphasize the discontinuity in relation to the current model, which is called Digital Feudalism. In the era of digital capitalism, the economy

would be based on a mutual benefit approach, and not anymore to a zero-sum game logic. The basic and radical change on which Digital Capitalism must start is considering data as a form of labor. Data as Labor (DaL) treats data as user possessions that should primarily benefit the owners of the data, and encourage the creation of a data labor market. Data as Labor paradigm is opposed to the current Data as Capital (DaC) one. Comparing the two paradigms, there are several implications:

- Data as Capital (DaC) treats data as natural exhaust from consumption to be collected by firms, while Data as Labor (DaL) treat them as user possessions that should primarily benefit their owners.
- DaC benefits AI companies and platforms encouraging entrepreneurship and innovation, while DaL benefits individual encouraging quality and quantity of data.
- DaC will reserve spheres of work to workers where AI will fail for humans, while DaL sees Machina Learning as another production technology able to enhance labor productivity and creating a new class of data jobs.
- DaC encourages people to find dignity in leisure or in activity outside the digital economy, while with DaL data works can be a new source of digital dignity.
- In the DaC scenario, online social contract is based on the exchange of free services for free data, while DaL can be a countervailing power to create a data labor market.

A data labor market potential benefits extend beyond the economic fairness, because it would also boost artificial intelligence productivity thanks to a better quality of the data. Property rights have protected and empowered individuals for millennia, evolving as technology does. As the printing revolution brought intellectual property rights and the Industrial Revolution the patent system, the digital revolution must bring the right to personal data ownership, therefore including the three classic elements of property rights: *usus*, *abusus* and *fructus.* In the final part of the work I focused mainly on the latter, i.e. how digital users can monetize their data, analyzing the legal and technological challenges and possible solution suggested by the doctrine. This idealistic, but realistic, vision is backed by many scholars and experts of the data economy, which support the ideal of a decentralized Internet, where individual digital identity has a value, and it is not reduced to a flow of data. In this scenario people would get paid for their contribution to the data economy and to the improvement of the algorithms that power the

artificial intelligence. The technical and legal challenges are difficult to overcome, but this is the right path to follow, in order to reach a fairer and more productive society.