The Open Data challenge in the Digital Economy

Summary

Abstract

The technological innovation that has developed since the birth of the Internet has allowed economy and society to change in a way that would have been impossible to predict. The new infrastructure and tools contribute to the constant development of the digital economy, which doesn't represent only a consistent growing part of the global economy, but also pervades people daily life.

In the context of the digital economy, a key role is played by data. Thanks to the power and variety of tools available today, data can be used with innovative and functional methods for the needs of citizens and enterprises. Computers' ability to process big amounts of data, i.e. "Big data", is one of the reasons of such change, because it has allowed the generation of new businesses, products and services. In this new scenario data becomes relevant as an inexhaustible resource that can be used by anyone who possesses the necessary instruments for its processing; it represents not only a source of information anymore, but also an opportunity for business and for improving the quality of life, because of its different possible uses. For this reasons the public debate has begun to address such change as the "data driven economy", i.e. based on a pervasive use of data.

Such new phaenomenon has also seen the public sector to take part, instead of remaining confined to the private sector. Open Government Data has been discussed from the beginning as a tool for transparency and accountability of the action of the public sector, but today it is also a functional mean for socio-economic development. However, the Open Data theme hasn't known a uniform development among countries during the years, but it is affirming with different characteristics based on the context. Today the worldwide experiences are various; the Anglo-Saxon countries occupy the first rank in this type of initiatives (*okfn.org, 2015*). In Italy the development has been slower and remains not homogeneous at this moment.

This work is a research on the Open Government Data's role in the digital economy; the main goal is to understand if socio-economic value is derived or can be derived thanks to Open Data, in order to generate an impact that justifies the investment of the public sector in the opening of its databases.

In the first part the general context of the digital economy, and in particular of the data driven economy, is presented. The first chapter describes the characteristics of the Internet and the impact that such communication platform has on the economy. The second pillar of the digital economy is addressed in the second chapter: data is analysed as a source for value generation within economy and society, discussing how the technological revolution has allowed such resource to be more effectively exploited. The third chapter is dedicated to the projection of the digital economy to future scenarios, trying to understand which effects on the economy of tomorrow could be observed.

The second part focuses on the Open Government Data phaenomenon, in particular when, where and why it takes place; in the fourth chapter definition and characteristics of the concept are presented, along with policy challenges. The fifth chapter is entirely dedicated to the Italian context; a general description of the digitalisation aspect and the Open Data policy is brought. The goal of the last chapter is to map some experiences of reuse of Open Data in the private sector, collecting

information and insights from the research “Open Data 200”, in order to understand if opening government data can represent an opportunity for the Italian business.

SECTION 1 New technology-driven economies

CHAPTER 1

An enabling infrastructure: the Internet

1.1. What is the Internet?

The Internet can be defined as *"A global computer network providing a variety of information and communication facilities, consisting of interconnected networks using standardized communication protocols"* (*oxforddictionaries.com*). Even from this definition alone, anyone can understand that the Internet has been one of the most relevant technological breakthroughs of all time. Since its spread throughout the world, society has changed in almost every aspect. Basic social interactions such as talking to a friend or buying a product are very different from twenty years ago. Knowledge has become accessible for anyone who can afford a personal computer and an Internet connection. Goods that could be exchanged only in physical form such as music and movies today are accessible on the Internet for purchasing and sharing. Companies and governments provide goods and services through the Internet in order to reduce costs and reach as many people as possible. Such impact on society has been one of a kind not only because of its proportions, but especially because of its speed (Fig.1.1). Not much time had to pass for the new platform to reach millions of people, and today we are looking at a world were nearly three and a half billion people use the Internet (Tab.1.1)

Internet was developed in the United States during the second half of the XX century as a communication system, mainly for research and academic purposes. His potential grew exponentially as more devices were connected to it and more information was put and shared through it. The turning point of this growth was represented by the invention of the World Wide Web by Tim Berners Lee in 1989. The World Wide Web is "*an information space in which the items of interest, referred to as resources, are identified by global identifiers called Uniform Resource Identifiers (URI)*" (W3C, 2002-2004); it was developed pursuing the goal of finding a universal format through which information could be used by anyone in the world. Thanks to the World Wide Web, people started to navigate in this web of contents such as documents, images, and videos in a very simple way and with very little expenses.

TAB. 1.1.Internet users in the world

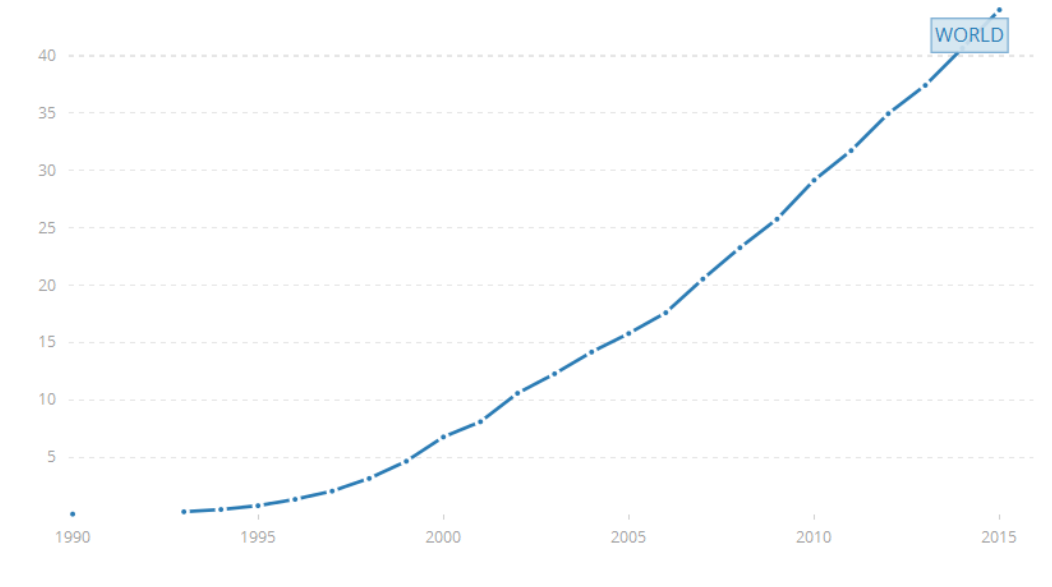
Year	Internet Users**	Penetration (% of Pop)	World Population	Non-Users (Internetless)	1Y User Change	1Y User Change	World Pop. Change
2016*	**3,424,971,237**	46.1 %	7,432,663,275	4,007,692,038	7.5 %	238,975,082	1.13 %
2015*	**3,185,996,155**	43.4 %	7,349,472,099	4,163,475,944	7.8 %	229,610,586	1.15 %
2014	**2,956,385,569**	40.7 %	7,265,785,946	4,309,400,377	8.4 %	227,957,462	1.17 %
2013	**2,728,428,107**	38 %	7,181,715,139	4,453,287,032	9.4 %	233,691,859	1.19 %
2012	**2,494,736,248**	35.1 %	7,097,500,453	4,602,764,205	11.8 %	262,778,889	1.2 %
2011	**2,231,957,359**	31.8 %	7,013,427,052	4,781,469,693	10.3 %	208,754,385	1.21 %
2010	**2,023,202,974**	29.2 %	6,929,725,043	4,906,522,069	14.5 %	256,799,160	1.22 %
2009	**1,766,403,814**	25.8 %	6,846,479,521	5,080,075,707	12.1 %	191,336,294	1.22 %
2008	**1,575,067,520**	23.3 %	6,763,732,879	5,188,665,359	14.7 %	201,840,532	1.23 %
2007	**1,373,226,988**	20.6 %	6,681,607,320	5,308,380,332	18.1 %	210,310,170	1.23 %
2006	**1,162,916,818**	17.6 %	6,600,220,247	5,437,303,429	12.9 %	132,815,529	1.24 %
2005	**1,030,101,289**	15.8 %	6,519,635,850	5,489,534,561	12.8 %	116,773,518	1.24 %
2004	**913,327,771**	14.2 %	6,439,842,408	5,526,514,637	16.9 %	131,891,788	1.24 %
2003	**781,435,983**	12.3 %	6,360,764,684	5,579,328,701	17.5 %	116,370,969	1.25 %
2002	**665,065,014**	10.6 %	6,282,301,767	5,617,236,753	32.4 %	162,772,769	1.26 %
2001	**502,292,245**	8.1 %	6,204,310,739	5,702,018,494	21.1 %	87,497,288	1.27 %

Source: internetlivestats.com (2016)

** estimate for July 1, 2016*

*** Internet User = individual who can access the Internet at home, via any device type and connection*

FIG. 1.1.Internet users per 100 people



Source: data.worldbank.org (2016)

Internet technology is so different from the other technological improvements that history has seen and so pervasive that in twenty years half of the world decides to use it. An interesting point regards the reasons of this fast spread. Here are two aspects that seem to be worth mentioning in looking for them.

1. **Speed and efficiency**.
   The Internet has given people the possibility to search for content and find it in a very little amount of time; communication is now faster than ever thanks to e-mails and social networks. In a few words, tasks which earlier required a significant amount of time can be completed in very little time, not only thanks to the Internet, but also thanks to the improvements of the devices used to access it. This aspect is definitely crucial for the impact on people's everyday life, but it represents an opportunity, especially for companies: in order to reduce costs and increase productivity, the Internet is decisive in today's economy, and its impact will probably be even more significant in the future.

2. **Content**.
   The possibility to find, use and share a lot of different contents can be considered the main feature of the Internet. In most cases, information, goods and services are all available on the Internet at low costs. This revolution has led computers to be the most used access point to almost every type of content, changing the way most people spend their time during the day.

To sum up, today Internet represents an infrastructure which influences a large part of world's society and economy. Its impact can be compared to the impact that technologies like electricity or the telephone have had during the past centuries, maybe even more relevant, because of the variety of aspects and sectors influenced by it. Therefore, some authors define it as a GPT, a General Purpose Technology (OECD, 2013a). This term stands for a universal technology supporting all sectors across the economy; according to Bresnahan and Trajtenberg (1995), in order for a technology to be a GPT, three characteristics are necessary:

- Applicability across a broad range of uses – "pervasiveness";
- Wide scope for improvement, experimentation and elaboration, continuously falling costs – "improvement"; and
- Facilitating further product and process innovations – "innovation spawning".

The Internet seems to fit this definition of technology. It can be used in multiple contexts; it becomes different and more useful as more people use it; it has been an excellent driver for further innovation both in terms of product innovation and process innovation. The Internet has also been the main driver for the Information and Communication sector (ICT) to grow; ICT refers to the ability of computers to use data and information in different contexts, and today it is often associated with the Internet due to the opportunities that this new technology brought to people and

in particular to the economy. Opportunities to collect and share huge amounts of data that can bring more efficiency and can also lead a firm to growth, thanks to costs decrease.

1.2. Internet of Things

As analysed before, the Internet represents an infrastructure capable of connecting people through different access points. Access to the Internet has been limited to a few devices, such as Personal Computers, but in the past few years, things have radically changed: technology is going towards a scenario where different devices and objects other than PCs are connected to the web. People realised that the impact of the Internet could be even more relevant if extended to the largest number of devices possible.
This scenario is now called the Internet of Things (IoT); this expression was used for the first time by Kevin Ashton, a technology pioneer, in 1999, referring to the change that internet-connected devices could bring to the world (Wood, 2015). Internet of Things is defined as *"The interconnection via the Internet of computing devices embedded in everyday objects, enabling them to send and receive data"* (*oxforddictionaries.com*).

IoT represents basically a network where devices can communicate through electric signals they send to each other. This web is composed by two types of elements that can interact: sensors and actuators. Sensors can detect physical changes and produce a signal which is sent to the actuator; the actuator responds to this signal performing an operations based on the nature of the signal received (Dutton, 2013). Such interaction allows a device to respond to an event that took place and was detected by another device. This process brings to machines the possibility to function in intelligent ways without the constant supervision of a human being.

IoT consists not only of device-to-device interactions: human-to-device interactions are also enabled. Those interactions are often made possible by connecting devices to smartphones, where the owner can monitor the connected devices and interact with them. An exemplifying case study of this innovation is the thermostat produced by the American company 'Nest'. The product has obviously all the functions of a traditional thermostat, but it works in a very different way: while setting the temperature has always required some type of human action, the Nest thermostat can learn users' temperature preferences during the day and the week, based on the setting's inputs it received. Additionally, it can sense the presence or absence of people in the environment and consequently set the temperature, thanks to different sensors. If it is connected to the Internet, the thermostat can be controlled also from the distance using the smartphone. Basically, the device constantly generates data in order to make the heating system efficient, saving both energy and money.

While an application as the one described can have little impact from a broad point of view, but also important from the point of view of the consumer (Nest Lab, 2015), using this type of technology in

a firm can make the difference in terms of competitiveness. In fact, IoT allows companies to optimize some processes, from machine monitoring to designing distribution systems, integrating their tools in the web and making them function in a coordinate way. Seeking efficiency is at the centre of the company's actions in every sector, and technological development is today one of the most useful instrument to reach it. Because of the multiple aspects technology has an impact on, it is interesting to understand which of these are the most affected. Lee and Lee (2015) identify 3 sectors in which IoT could bring more benefits, based on technology trends and literary review:

1. **Monitoring and control**
   Monitoring and control systems allow generating and collecting data on different aspects, like equipment performance, energy usage, and environmental conditions. Owning a technology which allows managing these aspects could lead to higher productivity and cost savings. One example of this innovation is the smart home: people are able to protect their property and save energy thanks to remotely control lights, climate and security systems.

2. **Big Data and business analytics**
   Thanks to the huge amount of data generated by sensors, it is possible to collect databases which, if analysed through business intelligence tools, can improve decision making processes. An example of these improvements regards the healthcare sector: patients need high personalized services, especially those who are affected by certain diseases, so data generated by medical devices are certainly of high importance. The result of the transmission of these data can be the activation of another device which acts directly on the patient to resolve the problem, or the intervention of medical personnel who becomes aware of the situation and can act faster and more efficiently.

3. **Information sharing and collaboration**
   IoT technologies allow three different types of information sharing and interaction: human-human, human–machine and machine-machine. Using data generated by devices connected to people or objects can be very important in a field like the supply chain, because delays and problems can be avoided with better control and monitoring. For example, devices can alert a store manager about malfunctions of some of his store equipment; also in the case of human to human interaction, the same store manager is allowed to better manage his or her employees using information about their real time position or activity.

Possibilities and benefits for the private sector result obvious from this analysis. Research for maximum efficiency by companies could be satisfied with tools like the ones just described. Future perspectives seem to involve a constant increase of IoT units on a global scale (Gartner, 2014); this increase will generate an impact which is difficult to understand now, both in terms of volume and nature. Nevertheless, it is clear that the more objects will be connected to the network through IoT tools, the more this impact will be relevant for society and economy. Those who will be involved in this revolution won't be only companies, but anyone who will find himself/herself living in contact

with a more connected world; those subjects will likely be spread over all society. Moreover, IoT could be fundamental and useful to deal with some of the global challenges that still lack of solutions, such as reaching sustainability of the human action to preserve the environment for future generations (Nonnecke, Bruch & Crittenden, 2016).

1.3. Internet Economy: definition and characteristics

As the Internet became widespread in the world and started to have a crucial role in many sectors of the economy, researchers started to talk about an 'Internet economy', referring to the huge impact the new technology was having. Internet economy was defined in the OECD Declaration for the Future of the Internet Economy as covering *'the full range of our economic, social and cultural activities supported by the Internet and related information and communications technologies'* (OECD, 2008a). The definition is exhaustive, but poses also certain difficulties when identification of such economy is required.

- Almost every aspect of the economy includes a certain amount of Internet's use; however, it cannot be said that certain activities are supported by it. A company which develops a simple website where only notions like how the company started or how many the employees are doesn't seem to be considered 'Internet supported'. Activities that include online buying-selling processes are clearly part of such economy, because the Internet represents the key of that type of activity. Although it is very difficult to define how much the Internet is important for an activity to operate, and identifying its degree of use can represent a difficult task. If different kinds of cultural and social activities are included in this analysis, the level of criticality of the issue increases, and represents a problem in terms of quantification of the Internet economy.

- The speed of Internet users growth can represent also an issue while identifying Internet economy. In Fig.1.1, the number of people using Internet is represented by the number of internet users per 100 people from 1990 to 2015; the curve seems almost an exponential function, with a linear grade in the last years. Because of the proportions of this growth, the Internet is used in more and more ways every day; mapping all the activities that are internet supported becomes a serious challenge as the spectrum of these activities changes and new ways of using information technologies are introduced.

- One essential feature of the Internet is that all the activity which takes place online is non-material. Transactions, communications, and all the interactions between people, firms and institutions are not tangible, as so are not many goods and services provided. This feature makes Internet economy very different from traditional economies, and very difficult to identify in its complexity. Therefore, collecting information about all Internet activities is one of the main obstacles, and it is necessary to overcome it in order to understand today's

economy. It is said that we live in the information economy, but collecting information about the shape and nature of the economy is now more difficult than ever (Brynjolfsson & Saunders, 2010).

The issue of identifying Internet economy is stressed out because sometimes it is difficult to understand what the reality is, based only on a theoretical concept. Giving a broad definition of a sector of the economy is one thing, understanding in reality what that definition indicates is a completely different operation. A general definition is not the only thing required for identification; decisions and assumptions are to be made, and it is never easy to find general approval on those decisions. To analyse and quantify Internet economy, it is necessary to decide which activities are exactly Internet supported and which are not; only after this discrimination, the volume and impact of Internet activities can be analysed. One of the challenges of our time is to understand how this new economy works and try managing it in the best way possible. This challenge is for every actor of society, because everyone is involved in this process. Also this challenge cannot be avoided; there is enough evidence about the ICT revolution to understand how actions, both from the public and private sector, are needed. Internet is not a 'ghost' only because it is mostly non-material and complicated; it is a reality with a huge impact on a vast amount of people. Trying to dodge the challenges that the new technology poses only increases the problem that those new technologies can create, if not seriously managed.

1.4. Internet economy: the impact

The issues explained above get in the way of measuring Internet economy. For this reason, many researches focus on the degree of Internet usage and on indexes like the number of Internet users within the population or the number of Internet connections operating in the world, in order to measure the impact of Internet economy. Therefore, the Internet phenomenon is often represented with tools that do not really explain how society is influenced by it; instead, they provide different information, for example on how many people are active online. However, that does not explain how many people are influenced by the Internet and which type of influence it has on them. Estimates on Internet's use are very common, estimates on Internet economy are not. This is also because there is not a generally recognized methodology for calculating these estimates, and debate is still alive.

The OECD (2013a) tried to give some new insights about measuring Internet economy. OECD conducted a detailed research with the goal of measuring the impact of Internet economy in the USA from 3 different perspectives:

1. **Direct impact.** This impact attempts to measure the share of GDP value added that is generated by Internet-related activities. This includes value added generated by:
    - Activities supporting the Internet (*e.g.* ISPs, Internet equipment manufacturers, etc.)

- Activities purely based on the Internet (*e.g.* search engines, e-commerce services, etc.)

Estimates: Internet activities add up to USD 165 billion in 2011 in the United States.

2. **Dynamic impact.** It checks the *i*) *aggregated* and *ii*) *net* impact that the Internet has on GDP. Studies within this approach take into account:
   - All possible industries that generate value added *thanks to the Internet* (not only industries supporting the Internet and operating purely on the Internet); and
   - The *net* economic effect of the Internet on the GDP.

   Estimates: In 2011 up to 7.2% of US gross domestic product was generated thanks to the Internet.

3. **Indirect impact.** This approach looks at the economic impact of the Internet that goes beyond the GDP. It studies two main impact areas:
   - The impact of the Internet on consumer surplus, and
   - The broader welfare gains generated thanks to the Internet (*e.g*. welfare gains derived from non-monetary transactions, impact on the environment, social capital formation, etc.).

   Estimates: it's very difficult to quantify this type of impact, because of the volume of data required and the high number of assumptions that are to be made. Existing studies on indirect impact are only partial.

This research also shows results from different works with the same goal; obviously different methods and assumptions led to different estimates, so more contributes to this exercise are needed for Internet economy to be understood and measured in a way that is commonly accepted. The Internet is challenging also because of the way it is conceived: roles of traditional economies, for example producer/consumer dichotomy, are not so clear online; everyone can become producer of various type of content, so the consequence is that the amount of producers has risen with proportions never seen before. The resulting economy is complicated and requires more attention when analysed.

1.5. Internet as part of the digital revolution

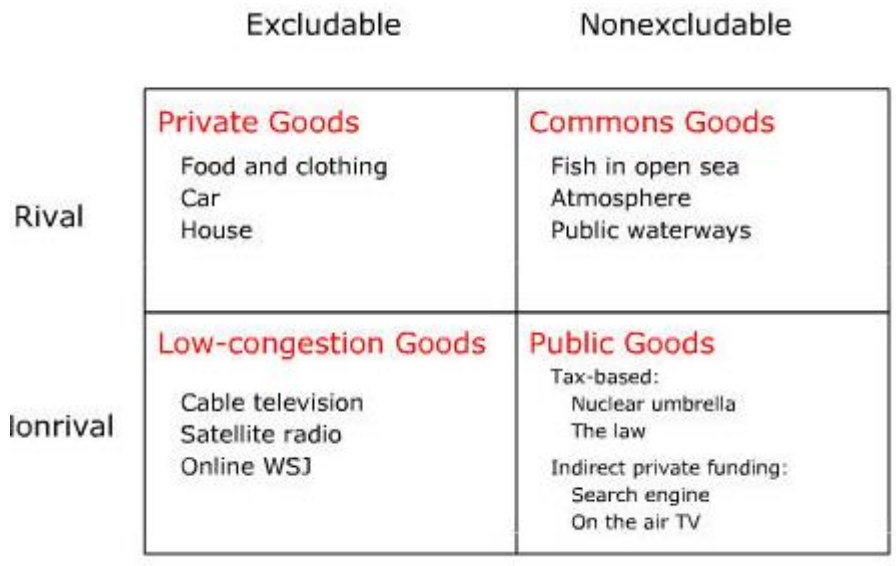
In a well-known paper, Intel co-founder Gordon Moore (1965) said that from 1965 to at least 1975 the number of transistors in a dense integrated circuit would have doubled every year. The prediction was adjusted to two years, but it was correct, and that trend kept going after 1975, as speed of transistors was included. Today, Moore's law is not representing the evolution of

computer's capabilities as well as it did in the past (Simonite, 2016), but the amount of uses which the new technology has allowed is still strongly rising, because of the variety of possibilities power machines give to people. As power of computers grew, devices costs also decreased with almost the same pattern. Minor costs allow technology to be used and developed by a greater amount of users, which bring the possibility for even more innovation.

These two tendencies, technological development and costs reduction, were essential for the success of digital revolution. "*Digital revolution refers to the advancement of technology from analog electronic and mechanical devices to the digital technology available today*" (*techopedia.com)*. The Internet is part of this revolution, being likely the most relevant element in it. In a world where digital technology shapes everyone's life, the Internet is fundamental for technology to be more effective. Digital information, goods and services can be more useful if only spread through the Internet, because it allows to reach the largest number of people possible; the more people it reaches, the more significant their impact is. It is clear how the Internet represents the fundamental infrastructure in the years to come, but it is also uncertain how the new economy will work in the digital revolution.

One of the challenges of contemporary economy concerns production. Economic theory classifies goods in four categories based on the presence or absence of two characteristics: rivalness and excludability (Fig. 1.2) (Samuelson, 1954; Buchanan, 1965).

FIG.1.2 Goods classification



Source: livingeconomics.org

The debate on these definitions has always been strong, especially on the concept of public goods and the ways they need to be efficiently produced from a Paretian perspective. In fact, public goods

suffer from free-riding, which prevent the market from being efficient. The common solution proposed to this problem is the intervention of the public sector.
In digital economy, a relevant part of the goods produced is digital goods. These are, for example, media contents such as audio or video files; obviously, there are many different types of digital goods, but taking music and movies as example helps to better understand the issue, because they represent an important fraction of the online activity. One essential feature of digital goods is their reproducibility. Reproducing a file without losing part of its quality is not one of the latest achievements of technology, since CD-ROMs were replicable already in the 1990s.

Non-excludability and non-rivalness were already characteristics of such goods, and free-riding behaviours like piracy began when the technology for practising such behaviour became available for the masses. The possibility to reproduce such goods an infinite amount of times is the key that enables goods to be non-rival and non-excludable, but for them to be accessible to everyone, a technology which enables fast sharing is required. If someone would want to listen to a song a few years ago, they should have asked to someone who owned that song on a physical support, like a CD, and ask them a copy of it. But, what if they would not know anybody who possessed that song? This issue left a good margin of excludability which the market could have taken advantage of.

The Internet overcame this issue. It can be imagined as a huge square where people can communicate and share or exchange goods, and distance does not represent an obstacle in any way. Thanks to the Internet, digital goods can be fully non-excludable, because a physical support to obtain such goods is no more needed. Once the first unit of a media file is sold, it becomes potentially available for everyone. There are also other factors that could influence the nature of digital goods, such as government policies, consumer behaviour, and firm strategies. Digital goods are also durable, therefore time cannot reduce their value. All these features reinforce the theory that digital goods are evolutionary public goods (Rayna, 2008).

The nature of digital goods makes it difficult for the private sector to provide them and to obtain sustainable profits. Since piracy behaviours took place, especially music companies have tried develop new business models, but free-riding remains the most convenient choice for consumers today. Internet has also erased the need for commercial brokering between the producers of music contents and the consumers. It is common for musicians today to sell their work directly without a record company which distributes it to the public; the full chain of music production has been influenced by new technology, creating a framework where everyone has the chance to be noticed because of their work and not because they are affiliated with an important record company. From this point of view, the Internet makes music market more competitive.

1.6. Internet issues

If it is true that the Internet provides tools for obtaining new benefits and possibilities, it also generates new challenges for the economy. It is significant highlighting these challenges, because of the speed which technological development changes society with: a scenario where the Internet represents the most relevant element of the economic activity requires deep analysis to govern it.

Technology itself is neutral, and the type of impact it has depends on how the global community decides to interpret it: consequences can be both good and bad. Below are some issues about the Internet deducted by evidence and literature:

- As the Internet has acquired the role of a General Purpose Technology (GPT), many traditional sectors of the economy are today very different compared to what they were in the past. It is hard to find a particular sector in which the Internet is not used, even in small part, but there are some of them that have been fully reshaped by that. For example:

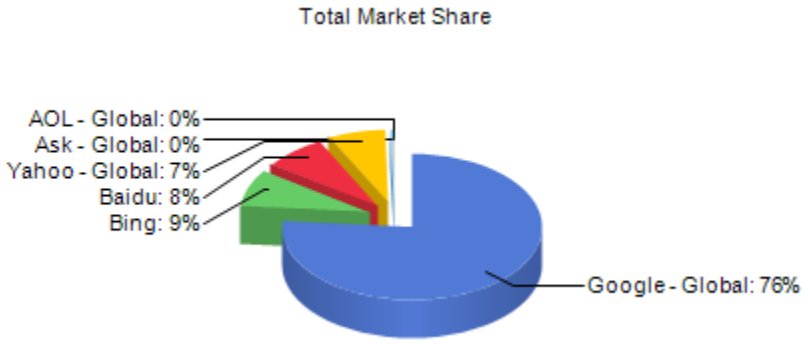
    - The telecommunication sector has changed in a way that communication is almost free if there is the chance to be connected to the Internet; services such as Skype allow to talk for free and make telephone technology obsolete;
    - The media industries suffer the digitalization of goods and the sharing technology, which lowers costs and sets the challenge of producing the media contents discussed earlier;
    - The knowledge market has to deal with almost the same problem of media industries, thanks to online services that provide encyclopaedias, newspapers, books, researches, etc.

    The problem with these changes is not so evident from the consumer's perspective, who obviously benefits of the access to more goods for cheaper prices, but it emerges seriously from the companies' perspective. The transition to new markets is ordinary in an economy where demand and supply meet in an equilibrium which is the optimal point from the Paretian point of view, but it represents a rather dangerous phenomenon when new activities do not require as much labour force. That seems the case of Internet activities, which not only replace other activities, but require less labour force than the latter ones (Lanier, 2013). The consequence of this process seems clear: unemployment. The challenge that not only the Internet but all the digital revolution poses, will be further analysed in Chapter 3.

- Another issue concerns Internet stakeholders. Even though the Internet is a community where everyone can contribute, there are some companies that possess a great amount of power, especially in the World Wide Web. The Web is a place of great monopolies, where companies like Google (Fig.1.3) and Facebook (Fig.1.4) control a big portion of Internet activity; therefore, they can manage their business in ways that advantage some users more than others, based on convenience. There is a strong debate about Internet neutrality: someone supports the idea that public intervention for regulation is required, in order to avoid abuses of power or the emergence of too strong centres of power; on the other hand, others argue that public intervention could be a barrier to innovation and Internet growth (Yoo & Wu, 2007).

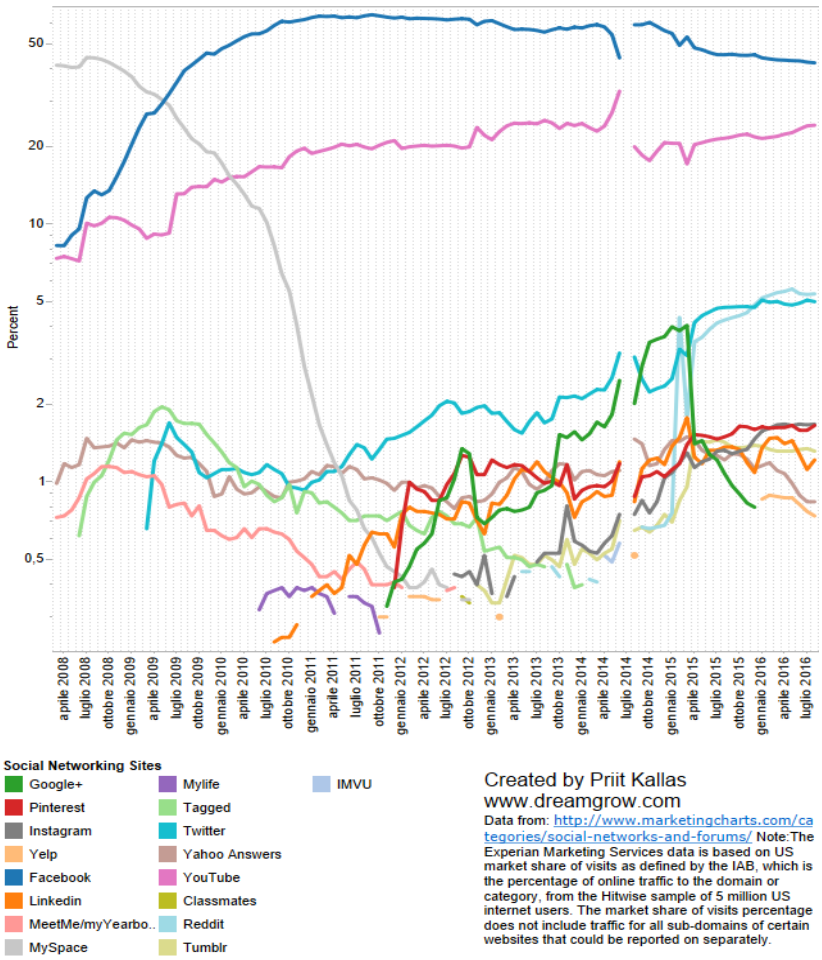
- Low access costs for the Internet allow the masses to be connected all over the world. Therefore, it can be a tool also for the low-income segment of the population for improving their social and economic condition; this could be a useful key for a decrease in inequalities that are present in contemporary societies (Oxfam, 2015).

FIG. 1.3.Desktop Search Engine Market Share



Source: www.netmarketshare.com (2016)

FIG. 1.4.Social Network Market Share by visit



Source: Kallas (2016)

But a link between the Internet and inequality reduction is yet to be demonstrated: it is necessary to understand which individuals benefit the most from being connected to the web. A research of the London School of Economics in The Netherlands, which has a developed digital infrastructure and has almost universal access, shows a completely different tendency: high income and education segments of the population take more benefits than poor and less educated individuals (Alexander et al., 2016). If such dynamic was confirmed through more studies, the Internet would represent a tool which increases social gaps instead of reducing them.

1.7. Conclusions

As quantification and understanding Internet economy remain difficult goals to achieve, a few statements about the technology behind this economy, the information technology, can be made with a good degree of certainty.

- Information technology has been the main driver of growth in many countries since the mid-1990s. New companies, new products, new services are developed every day thanks to the technology available; more are the people reached by this technology, more are the opportunities for the Internet sector to become a relevant part of the global economy. Using IT technologies has represented an advantage for companies: as Brynjolfsson & Saunders (2010) observed in the United States "*Until the early 1980s, the size of differences in profit margins did not vary much with IT intensity — that is, leading firms were only a few percentage points better in profit margin than lagging firms in those industries. However, since the mid-1990s the interquartile range of profits for the heaviest users of IT has exploded. The difference between being a winner and being a lagging firm in IT-intensive industries is very large and growing. Using technology effectively matters more now than ever before*."

- Having a place where anyone in the world can express himself and be listened to has always been a utopia. With information technologies, everyone without discrimination can share content with the online audience. The quote "*in the future everyone will be world-famous for 15 minutes*" by Andy Warhol in 1968 has become potentially real, thanks to the tools we possess today, like social networks, for instance. The Internet gives equality in terms of the possibility to express something.

- One of the goals of contemporary society is to achieve sustainable development, being the resources of the planet less available every day and because of the fact that the current models of development in western countries come at huge costs for other countries and for future generations. To achieve this goal, new ways to organize society and economy are still

to be found. But new technologies could represent the main infrastructure this new social and economic paradigm could take place around (Internet Society, 2015).

> *«It is now taken as a given around the globe that sustainable development is only possible if information and communication technology (ICTs), and particularly broadband, are deployed as a cross cutting catalyst for all three pillars of sustainable development.»*
>
> *Ban Ki-Moon, United Nations Secretary General (2016)*

CHAPTER 2

Data driven innovation

"*Information is not knowledge, Knowledge is not wisdom, Wisdom is not truth"*
Frank Zappa (1979)

2.1. What is data?

There has always been confusion around the concept of data. From a general point of view, data is defined as "*factual information (as measurements or statistics) used as a basis for reasoning, discussion, or calculation*" (*merriam-webster.com*). Data is basically a tool that people use for many operations, in order to better performs a task. Today the word 'data' is often associated only with its digital form. This association is caused by technology: today computers are the main devices where data is collected, stored and elaborated; therefore, it is easier to think of data as spreadsheets of economic indexes instead of a geo-localisation of a mountain path. A lot of different types of data exist, but commonly we think of data as the one which is more often used and showed by the media.

A distinction worth discussing is the one between data and information. Data is a fact about reality, it cannot be erased or be wrong; while an information can. Information is a container of data and people can use it to grab the data contained. For example, looking at a picture, we can identify data on the people portrayed, extract it, and store it in our brains to gain knowledge. The data itself (*e.g.* the colour of someone's hair) has always been there, but only thanks to technology we can capture it and generate information, which can later be shared with other people than the ones who actually see the person directly. Data is used to enrich knowledge, which is the basic material for the human brain to take decisions. For data to become something useful, it is necessary to convert it in information, as information is data processed in a meaningful form.

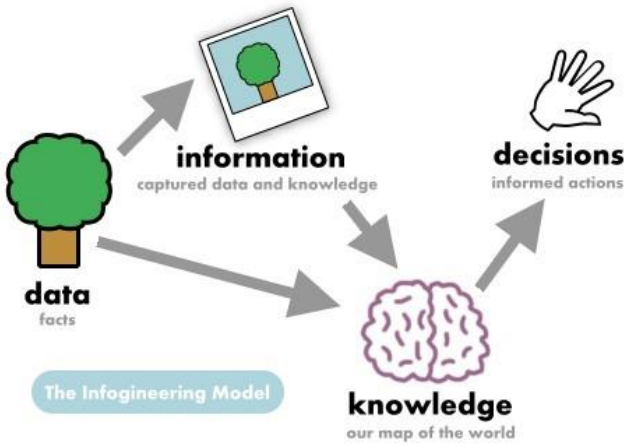
Data can be considered as a raw matter which it is possible to obtain information and knowledge from. Since it is facts about reality, data has always existed; but it is now more easily understood, because technology gives more chance to collect it and store it.

2.2. Data value

From the advent of innovative data process technologies, data itself has become a source of value. Financial data, budget data, consumer data are used by companies to be more efficient and

competitive; governments use statistics about population to design effective policies; the media collect data to provide audience with accurate information.

FIG. 2.1.Data, information and knowledge



Source: www.infogineering.net

The value of data depends on many different aspects, such as the data type, the available technology for processing, how data is used and for what purpose. These aspects can be summarized by the data value chain, which represents all the phases of data life cycle:

FIG. 2.2.Data value chain and life cycle

Source: OECD (2013b)

Generation of value depends on how and by whom every step of the data life cycle is conducted. Differences of this process lead to different value in terms of generated volume of impact and subjects who can benefit from it. Focus should be also on obstacles which prevent value from data to be generated. Every step of the data-value chain poses several challenges. Some of them are summarized below.

1) **Excessive focus on data processing.**
   It is understandable how generation, collection and storage are the fundamental tasks to be performed in order to achieve the basic prerequisites for value generation. If such operations are not well carried out, discussions on the ways data can be used lose their importance. This concept may seem obvious, but it is worth to highlight these aspects: a lot of enthusiasm circulates around all the possibilities that data processing, distribution and analysis can give, but less attention is paid on how data is generated, collected and stored. Few guarantees on the first phases of the process result in fewer possibilities on an exponential scale: if one dataset contains bad data, multiple operations, and eventually value generation, are

prevented. The consequence is that value of data is largely determined by data providers (Chatain, 2011), which play the most important role in the data value chain.

2) **Technological gap.**
   Another issue is represented by the technological gap between the tools necessary to exploit the data life cycle phases. Such gap is well represented by divergences between data generation and data storing and processing. Since technology, especially tools like sensors, gave the possibility to generate vast amounts of data, processing and storing it has become a challenge. This is caused by the speed which data is produced at, which is higher than processing speed (Géczy, 2015). The gap represents a technological challenge, which needs to be resolved for maximising the efficacy of data.

3) **Data culture.**
   To somebody who is used to work with data, the concept of value generation from it seems surely obvious. Whoever daily gains benefits from using data or has the chance to study such value can easily understand why a more data-intensive future scenario is not only desirable, but also necessary. But for those whose life is not directly influenced by data, in countries or sectors where these aspects have never been a concern, linking data to value is not immediately absorbed. From a geographical perspective, it is clear how data culture is more developed in some countries, especially in USA and UK. From a sectorial perspective, there are certain sectors of society which are historically more data intensive, so they can faster and more accurately identify benefits from data usage. Company size too could be a factor in this process: firms of relevant size and with branches dislocated in multiple places have always had the need of more data possible to control and develop a sustainable business. On the contrary, small companies could not realise the value of data because they are not accustomed to its use.

An interesting question could be raised on the reasons of this data driven innovation. The answer lies in the improvements that technology has brought; these concern every step of the data value chain: with more data generated, more capacities in collecting and storing, new ways of processing, and an infrastructure like the Internet that allows fast distribution, using data is now easier than ever. Thanks to technological development, data driven innovation has become one of the most discussed topics both in the private and public sectors, due to the fact that it is seen as a fundamental driver of growth, better quality of life and sustainable development.

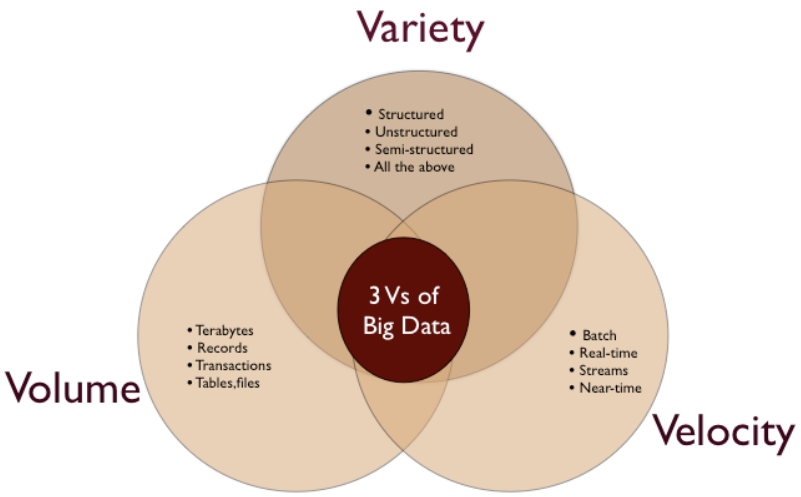
Data peculiarities go beyond its multiple uses. One interesting feature can be observed in the context of data as ‘raw matter’. Raw matters such as coal and oil, which allowed economic development in the two industrial revolutions, are exhaustible resources; only when mankind succeeded in the challenge of using renewable resources to produce energy on large scale, a way to produce and live without condemning future generations to a poorer existence, was found. Data

represents one of these renewable resources: its production cannot be stopped and it does not generate any type of impact on the environment. It could be said that using data has obviously an impact because the devices needed are material objects, but that does not mean data itself has an impact: it is a non-material raw matter; therefore the material environment is not influenced by it. Its potential is yet to be fully exploited, since it is used only in a marginal fraction of human activities.

2.3. Big Data

A keyword that is often brought up today is 'big data'. It is interesting noticing how a commonly agreed definition of the concept is yet to be found (Press, 2014), despite the fact that the expression is largely used, mainly in the context of new emerging markets. **Big Data** is generally referred to as a big **volume** of data; the problem is that the concept of *big* is relative, it varies as technology develops and sets new standard; therefore, it is susceptible from person to person and from time to time. Volume is not the only characteristic that can be considered; also velocity and variety are suggested by literature to be Big Data features (Douglas, 2001; Beyer and Laney, 2012). **Velocity** is referred to as the time necessary to process a certain amount of data; collecting and using data with real-time technologies is the best example of this concept: as data is constantly stored, huge volumes are reached in very little time, also depending on the amount of data produced. **Variety** concerns the different types of sources where data comes from: for relevant volumes of data to be used, it is not always possible to collect it from one source only, but different ones are required. Obtaining a database which is the more complete possible, could go through linking different data and to organize it in the most useful way. If these three characteristics, volume, velocity and variety (the three 'V'), are met, data can be defined without doubt as 'Big Data'.

FIG. 2.3.Big Data 3 'V'



Source: Di Girolamo (2014)

When technology made possible to manage Big Data, many sectors began to embrace the Big Data challenge. Companies saw new ways to create business and to be more efficient. Three types of data can be classified, which correspond to many more uses (Vale, 2013):

a) Social networks (human-sourced information): This category includes data regarding experiences of people. They are put online and can be used by companies to obtain information on someone and, based on this information, propose certain products or services. This type of data can derive from audio and visual content, maps, messages, blogs and documents;

b) Traditional Business systems (process-mediated data): This data comes from traditional business events, such as buying-selling interaction or client registration; both the public and private sectors contribute to collecting this data, as they can be generated from classic transactions or even medical records;

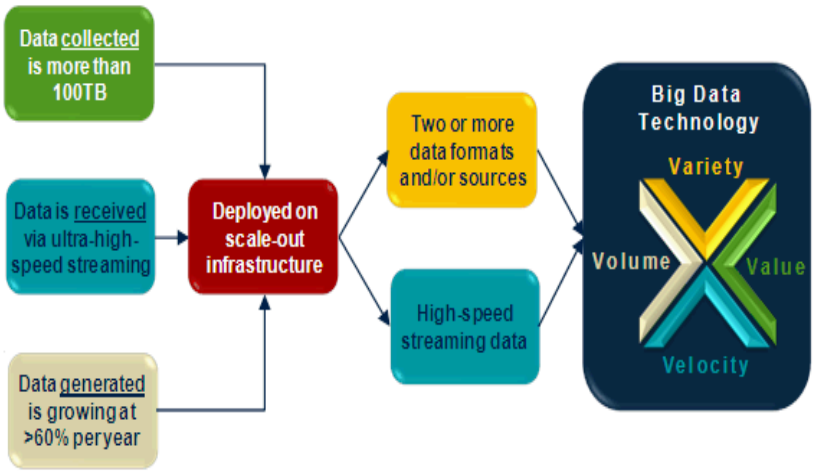
c) Internet of Things (machine-generated data)**:** Data which comes from machine-machine interactions through IoT technologies like the ones described earlier; monitoring and controlling machines thanks to sensor systems could generate data in multiple sectors, for example environment data or traffic data;

Understanding the value of data led to the emergence of a Big Data market. A research by International Data Corporation (IDC) analysed the evolution of this market forecasting its evolution from 2010 to 2015 (IDC, 2012). Criteria to define this market were summarised in 3 steps:

1. Evaluation of quantity of data collected (more or less than 100 TB), usage of ultra-high-speed messaging technology for real time, streaming data capture and monitoring and rate of growth of datasets at 60 % or more
2. It is evaluated if, for every criteria of the first step, technology is deployed on scale-out infrastructure
3. Evaluation of usage of two or more types of data or data sources and usage of high-speed data sources

The three ‘V’ remain the way to identify Big Data activities, but another V is added, standing for **value**: it is analysed and evaluated the socioeconomic value that is or could be generated by Big Data. It represents the reason that drives Big Data use; therefore, in this case, the definition goes beyond the technical aspects which are the core of the first three ‘V’, in order to get a more comprehensive view of the potential of the data.

FIG. 2.4.Big Data Market identification methodology



Source: IDC (2012)

More recent studies observed the dimension of this market and made forecasts on the future impact of Big Data on the economy. Wikibon observed that 2015 represented a breakthrough year for Big Data, because the growth rate of the market was high (23.5%), many big companies began to be interested in data technologies in order to become more data intensive, and start-ups which have Big Data at their core (Wikibon, 2016) emerged. Wikibon's forecast is represented in Fig. 2.5.

Big Data revenues are expected to grow from USD 18.3 billion in 2014 to USD 92.2 billion in 2026. Three sectors are identified in the Big Data market: Big Data software, Big Data Hardware and Professional Services. The software sector is expected to be the one with the faster rate growth (20%) per year, until representing in 2026 the biggest part of the market (46%), despite the fact that in 2014 the professional services sector represented the largest part of it. Professional services will decline in their growth rate as more actors will gain experience in Big Data technologies, to the point where services are less required (Tab. 2.1).

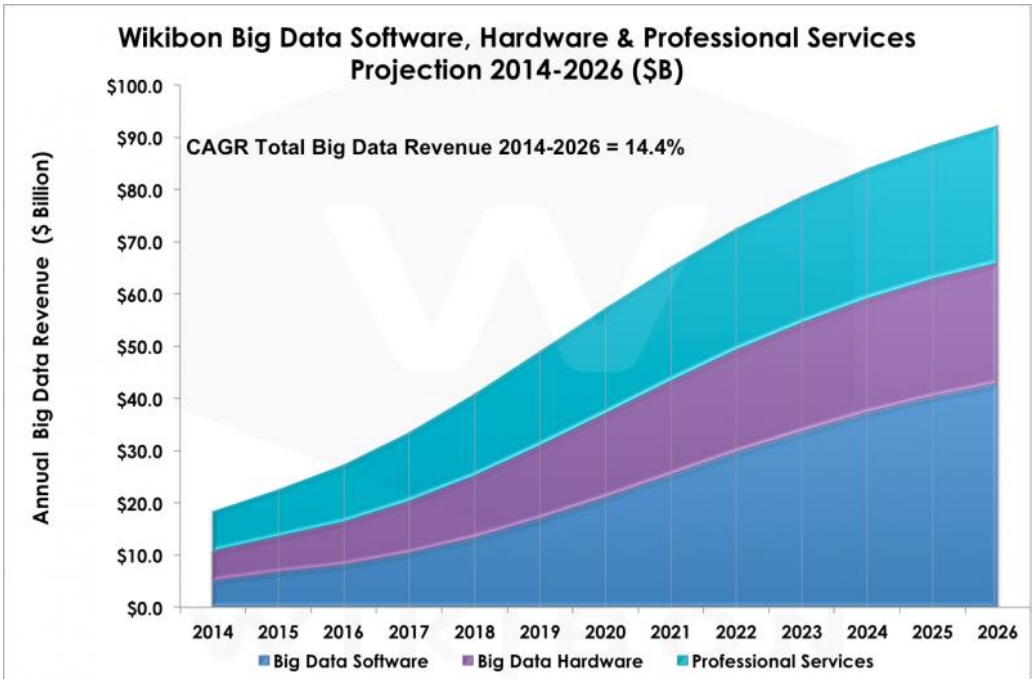
TAB. 2.1.Big Data market segments dimension and forecasts forecast

	% of Total market (2014)	% of Total market (2026)	% CAGR* (2014-2026)	Total revenues from Big Data (2014) (USD)	Total revenues from Big Data (2026) (USD)
Big Data Software	29.0	46.0	20.0	5.31 bn	42.41 bn
Big Data Hardware	31.0	25.0	12.0	5.67 bn	23.05 bn
Professional Services	40.0	29.0	11.0	7.32 bn	26.74 bn
Total market	100.0	100.0	14.4	18.3 bn	92.2 bn

**Compound Annual Growth Rate*

Source: personal elaboration from Wikibon (2016)

FIG. 2.5.Big Data Software, Hardware & Professional Services projections



Source: Wikibon (2016)

2.4. Value throughout sectors

Another fundamental matter regarding value generation can be identified in the search of the sectors where using data is most valuable. As said before, the value can differ from sector to sector depending on various factors; for this reason, selecting some of them represents a difficult task. OECD tried to identify 5 sectors that could be subject to a relevant impact by using data (OECD, 2013b); these sectors were selected by analysing previous studies and evidence in literature. The main criteria used for the selection was the level of data-usage among different sectors: where data is not exploited as a resource yet, more potential for using that data exists.

1. Online advertisement

Today, firms can collect a vast amount of data through online services and tools. Consumer data is used to improve products and develop effective ways to advertise. The Internet allows companies to collect a lot of personal data about people who use online channels to buy a product. Also social networks often include personal information which can be useful for better targeting. The more data about its customers a company has, the more it can design efficient channels for distribution and advertisement. Firms are not the only ones influenced by this new way of using data; consumers can take advantage of the data used by companies, wasting less time in looking for a certain product or finding more products which they could be interested in purchasing.

2. Governments and public sector agencies

The public sector is a relevant producer, user and supplier of data. Information in the public sector is referred to as PSI (Public Sector Information) and it is proved to be a significant source of value. Although PSI seems far from its full-exploitation, using PSI represents a key to various benefits, especially for citizens: transparency, efficient public services, and development of better policies. A more intensive public sector could be more efficient by saving costs and wasting less time in the policy making process. Statistics play an important role, but also instruments for connecting citizens with the public administration are essential. In this perspective, spread of Open Data initiatives has helped the change taking place. Further analysis on this aspect are discussed in the second part of this work.

3. Health care

Advantages derived from data-using in the healthcare sector are linked to the previously discussed matter regarding the benefits of IoT technology. Big Data transmission from medical devices allows medical personnel to observe the status of patients almost in real time, with the possibility to act faster and provide better treatments. Thanks to generation and storing of this data, diseases can be analysed, and therefore understood, in a more effective way; through this understanding, both services and products in the healthcare sector could be improved. Also cloud technology allows to share data so that all the individuals involved — medical personnel, researchers and patients — can benefit from it. Costs saving is another fundamental goal that could be achieved using Big Data; the healthcare sector represents a large part of public expenditure in European countries, and a service which not everyone can afford in countries where it is taken care of by privates, such as the US. Using data in the healthcare sector in the United States could bring a benefit of USD 300 billion savings (MGI, 2011).

4. Utilities

Like the healthcare sector, the utilities sector could benefit a lot from IoT. The new ways of collecting data available are a real breakthrough in this case: for example smart grids, electricity networks that can produce data, are tools for better managing the electricity supplying process and avoiding energy waste. Another example regards using Big Data in the domestic field, like the Nest thermostat mentioned earlier, which allows cost and energy savings. A critical point of the water supply sector is represented by leaks; now technology allows, through sensors systems, to control water pipes so that leaks are identified and dealt with faster. Thanks to these applications, Big Data are a turning point in the way natural resources could be managed, and a relevant step to reaching energetic efficiency.

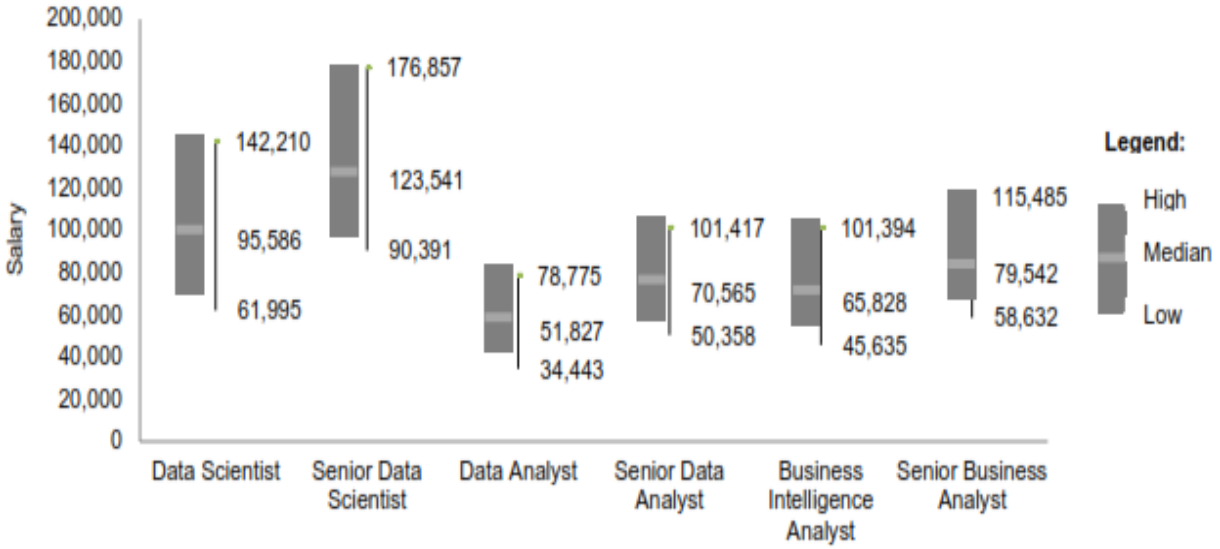
5. Logistics and transportation

A type of data that has seen a rapid and consistent growth in terms of volume is geographical data; smartphones possess a GPS system which is used mainly used by applications like Google Maps: data generated by these devices can be used to find the best itineraries in the transportation sector, both for goods and people, and consequently to save time and money. If this data is integrated in

Big Data databases, they can be used to identify and communicate the presence of traffic on a certain route, or others obstacles which can slow down the vehicle; also accidents can be localised thanks to data generated by mobile phones. Another application of Big Data regards specifically vehicles, which can be monitored by sensors that receive signals from different parts of the vehicle. All these possibilities are very useful in creating a scenario where transport companies can manage traffic and the condition of their vehicles in a more efficient way, saving energy, time and money.

One relevant aspect influenced by the fact that organizations realized how data could have been a source for value, concerns labour and education. The need of workers who could use data for creating value, generated a demand of highly-skilled labour in data profession; the education system has tried to intercept this demand, starting education programs focused on data and its uses. Despite this, the demand has not been fully satisfied, because it has risen so fast that the education system has not had the possibility to respond with the same speed. This aspect created a gap between demand and supply of data highly-skilled labour, and as a consequence a strong rise in data professions salary (Géczy, 2015). For example, the figure of 'data scientist' emerged, which has now in the US a median salary of USD 95,586 and USD 123,541, respectively for data scientists and senior data scientists, in contrast with data analyst workers for whom the median salary is USD 51,827 and USD 779,542 (Fig. 2.6).

FIG. 2.6.Annual Salary Ranges of Data Oriented Professionals in the United States (2014)



Source: PayScale Inc. (Updated: December 11, 2014)

Demand of this type of professions could represent an opportunity for creating occupation in a moment when, especially in Europe, unemployment is an issue, particularly in Mediterranean countries (*statista.com, 2017*). Nevertheless, intercepting this demand remains a serious challenge: the data scientist jobs require multiple skills, besides the technical ones; therefore, training programs for data scientists require also a relevant amount of time, and they could be accessible only to a small portion of the population, needing a high level of education. It will be interesting to

observe if in the next years the gap between demand and supply will diminish or be erased, or if demand growth will continue to be too fast compared to the supply growth.

2.5. Data society criticalities

In the previous sections, only the benefits derived from a society where big amount of data are collected, stored, processed and shared to generate socioeconomic value, are being highlighted. Even though these benefits are undeniable in many sectors, it is important not to omit the issues which this society could generate. The benefits earlier described can be seen very clearly from the perspective of a firm, not necessarily from that of a citizen; certainly, some of these benefits involve people too, in particular the development of markets where products are accessible in a faster way, more personalised and cheaper, a healthcare system with more effective treatments, or an environment with less pollution. But with regard to other issues, the perspectives of companies and citizens are likely to collide. One of the most important issues concerns privacy.

In the past, the issue of privacy was always moderately confined from the fact that acquiring personal information required more resources in terms of time and money, in addition to a level of intrusion in the private sphere of individuals that was more easily identifiable and avoidable. There are many examples of this fact: in order to locate a person, a complex system of cameras that could identify the face of an individual was needed; getting information like address, family composition, type of occupation, the network of people which someone interacts with, required a more or less direct contact with the subject, who possessed more control on it. These obstacles were in large part overcome thanks to the digital revolution which has brought us in the information era.

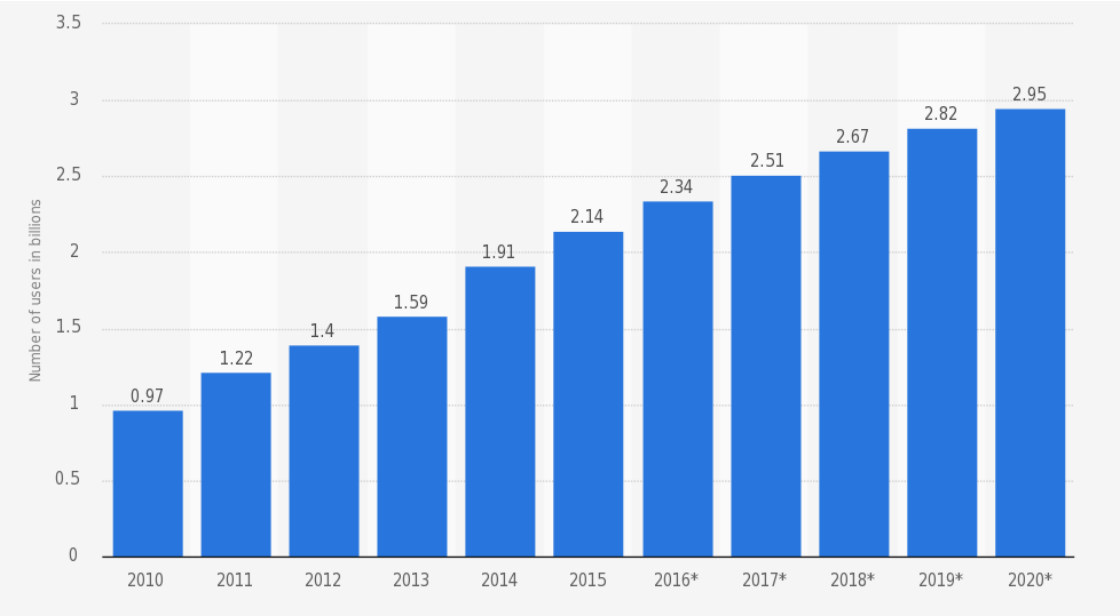
The platform that has allowed this transition is the Internet. From the development of the infrastructure that took place in the third millennium, tools like social networks were born, in which users can share different types of content, which is basically data, with other people. The possibility to share this content has generated a large interest, to the point that it has become one of the most common practices in the world (Fig. 2.6). The utility of the practice is immediate from the users' perspective; however, it is hard for them to understand the implication of this operation at once. In fact, many users think that what they share online will remain within the circuit in which they shared the information (*e.g.* a social network like Facebook), whereas that information can be taken by somebody else and published on other platforms, or even used for other purposes that the user cannot really imagine (Aware & Obama, 2009).

In the best case, personal data is used by companies to target certain groups of people in advertising a product, based on characteristics and preferences expressed through the Internet (Kadri-Liis Piirsalu, 2012). In the worst case, that data is used for criminal purposes: the more an individual has exposed himself/herself through the Internet, the more extortions, robberies, menaces are easier to be made, because of the sharing of all the information on their life and that of others. The reason is that whoever today is interested in obtaining personal information about someone, does not

encounter relevant obstacles, because that information is shared on purpose by the individual concerned on a public space.

Different types of data represent different grades of threats; bank account data is probably the most important one, and even though it is not easily traceable thanks to banks' security systems, it is constantly under attack by hackers. Thus, the web represents not only a harm to privacy, but also to security: the digitalisation of data has led to a situation where more precautions and caution are needed, because little mistakes and lack of attention could result in consequences of significant dimensions for individuals (see the social experiment conducted in Brussels by Safeinternetbanking.be). Furthermore, sometimes necessity of privacy and security cannot go along: data can help improve security if it is used by institutions which are called to law enforcement; on the other hand, it exposes the subject, due to the availability of this information on digital circuits. Therefore, more data can mean more and less security at the same time, but undoubtedly less privacy.

FIG. 2.7.Number of social media users worldwide from 2010 to 2020



*estimates

Source: eMarketer Statista (2016)

Regulation on data and on the Internet is crucial in the information revolution. Yet, if every country decided for a certain type of regulation, it could not suffice to have control on some activities. A lot of internet-related companies, especially the biggest ones, do not manage all their activities in a single country, so it is hard to control them effectively, being in a supranational dimension. Also the legislation on this matter varies from country to country; for example, the system of the European Union has a stronger focus on privacy than that of the USA (IFLA Trend Report 2013); it is instituted as a human right.

Article 8 – Right to respect for private and family life

1. Everyone has the right to respect for his private and family life, his home and his correspondence.

2. There shall be no interference by a public authority with the exercise of this right except such as is in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others.

Source: ECHR, European Convention on Human Right (1950)

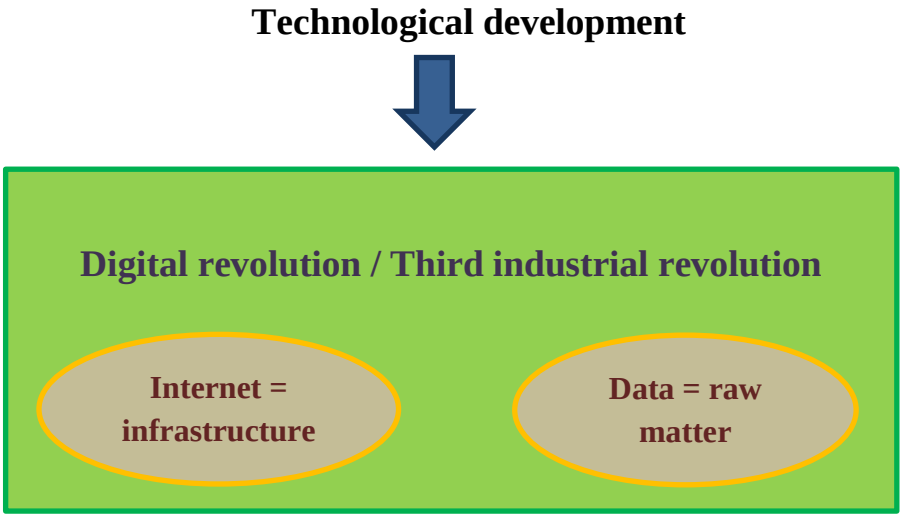
Differences between legal systems represent an issue when, for example, the data life cycle does not take place entirely in a country: data can be collected in one country and processed in another, generating obstacles in applying rules and pursuing illegal behaviours. To overcome these difficulties, more integrated and common rules and constant worldwide debates are required.

CHAPTER 3

The digital revolution: characteristics and future perspectives

In the first chapters two aspects are analysed: Internet widespread and innovation through data, which are hard to discuss in an exclusive way, because they are part of the same phaenomenon. This phaenomenon is called the digital revolution (Fig.3.1), which took place from the Eighties and saw a increasing of digital circuits and in general a high growth of ICT sector. To this revolution is also associated a new industrial revolution, the so called Third industrial revolution (Rifkin, 2013), because it is shaping some crucial aspects of the contemporary economy and society. Technological innovation is the engine that has allowed the digital revolution to take place; in the digital context Internet has the role of an infrastructure that allows communication between potentially every object, and also changes the ways people interact; data takes the role of an essential raw matter from which a more efficient and smart economy can be generated. The several aspects that are influenced by Internet and data are the demonstration of how important they could be for the present and future society.

FIG.3.1.The Digital Revolution



Source: personal elaboration

Digitalisation is associated to a new industrial revolution because the change that is happening has some common features with the first two revolutions. Jeremy Rifkin (2015) argues that industrial revolutions take place thanks to a bond between a new communication model and a new energy model; if in the second revolution the telephone and oil, along with the internal combustion engine, have embodied these new models, in the new revolution Internet and renewable energy seems to take over as the new ones. Internet is already the leading infrastructure for communication in a global scale, as explained in the first chapter. Renewable energies represent every year a higher

quote of the global energy production (REN21, 2016) because of the investments that were made to find a more efficient and less environmental-affective way of sustaining society (Fig. 3.2 and 3.2).

FIG.3.2.Global new investment in renewable energy by asset class, 2004-2015, $BN

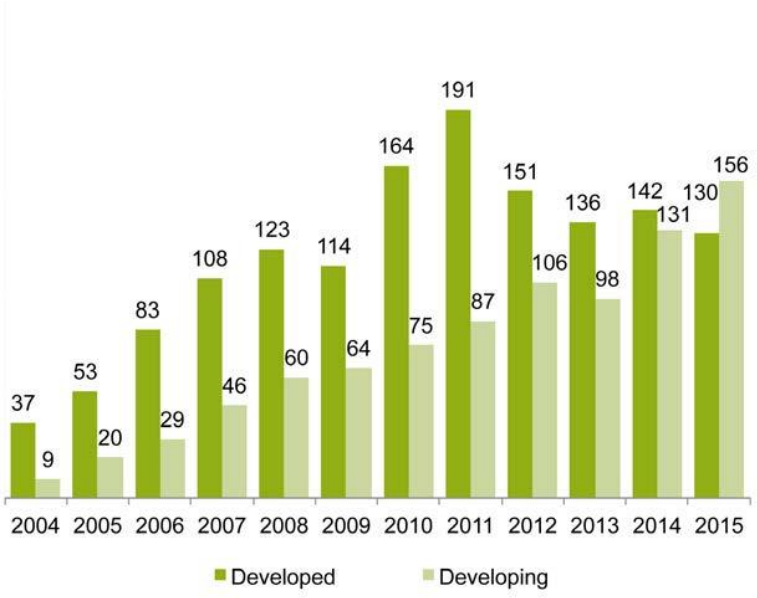
* Asset finance volume adjust for re-invested equity. Total values include estimates for undisclosed deals

Source: UNEP & Frankfurt School of Finance & Management (2016)

The consequences of this revolution are determined for the most part by the impact that contemporary technology has brought to production and to prices of goods; efficiency that companies can achieve through Internet, data, robots and Artificial Intelligence shapes a socio-economic scenario very different from the one we've lived in since the last industrial revolution between XIX and XX century.

Rifkin defines this scenario a "zero marginal cost society"; an economy where the cost for the production of one more unit of a good is almost zero, therefore the fixed costs represent the major share of the total expense in the production process. Some researchers like Ayres (2010) analyse the fundamental factors of growth and claim that the most important factor in order to achieve higher productivity is energy, or better said, thermodynamic efficiency. Capital per machine and work revenue represent only the 14% of growth, as Robert Solow (1963) discovered. Technological development is taking thermodynamic efficiency to high levels, because energetic sources to which the investments are concentrating on, like solar energy and wind energy, are seeing an exponential growth in the capacity of technology to capture and use them.

Figure 3.3.Global new investment in renewable energy: developed vs developing countries, $BN



Asset finance volume adjust for re-invested equity. Total values include estimates for undisclosed deals. Developed volumes are based on OECD countries excluding Mexico, Chile, Turkey.

Source: UNEP & Frankfurt School of Finance & Management (2016)

Moreover the marginal cost of energy production from these sources is destined to be very low (Sunshot Vision Study, 2012; Deligianni, Ahmed & Romankiw, 2011; Naam, 2011). Internet could also contribute to this new energetic scenario, allowing communication between different facilities and exchanges of energy units; for this reason in the field of Internet of Things the "Internet of energy" has a crucial role.

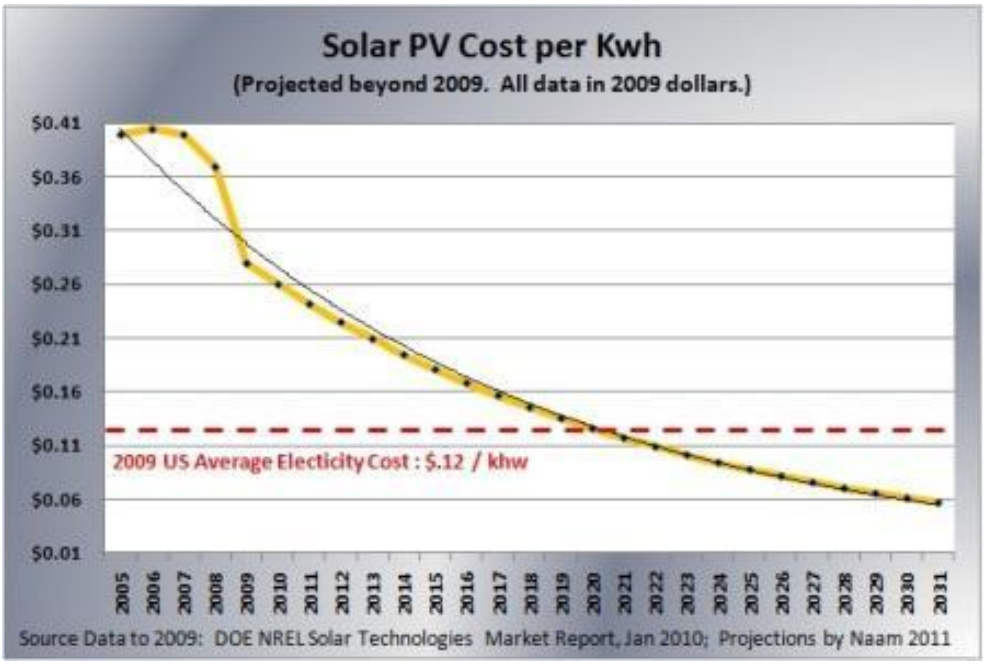
Simultaneously labour costs dramatically decreased due to the substitution of human work with machines and computers, which are more productive, efficient and without negotiating power. Such substitution regards important sector of the economy like services, entertainment, knowledge and manufacturing. Switching to machines led the production process to be even more efficient, pushing marginal production costs to lower levels.

Inputs received from the new energetic, technologic and communication paradigms have an impact on the whole society and on different sectors; if the medium-long term is considered, entire aspects of our life could change. Some of these changes are here presented, referring to Rifkin perspectives.

- **Energy**
  Thanks to the sign of COP21 agreement on 21st December of 2015, 195 countries stated their effort to stop and prevent the global warming that is happening, to use renewable sources of energy to reduce environment costs derived from human activity and to pursue resiliency as a general goal (UNFCCC, 2015). Investments which are made in this direction,

which can bring result like the ones showed in figure 3.4, could reduce energy production costs to levels near to zero. Furthermore future could present changes from the point of view of the subjects who produce energy; thanks to technologies like solar panels every building is potentially a small power station in which electrical energy is produced, stored, consumed. Such energy could be exchanged too, thanks to the network that Internet can create. The result will be the same as the previous, the lowering of private expenses for energy sustenance, in addition to a lower impact on the environment of the human activity. A world where everyone is auto-sufficient in energy production can be very interesting both for privates, who can manage independently their needs and give spare energy to those in need of it, and for governments, which can spare public money to provide energy to citizens and companies after the infrastructure for this revolution are built. For now the installation of these implants to national scale remains far from reality; therefore, strong public incentives and investments need to be provided for this scenario to be realised.

Figure 3.4.Solar photovoltaics price trend



Source: Theel (2013)

- **Goods**

  The phaenomenon of automation of production processes brings a decrease of goods marginal costs and as a consequence a decrease of prices. Mass production could go through a new spring due to this phaenomenon, but with some limits to the sustainability of this scenario that could emerge and are presented later in this paragraph. For sure consumption could increase in the periods immediately next to the widespread of those production processes, because of the greater supply of many products. However other technologies are pushing towards another production model, similar to the one of the energy. The keyword is

again self-production, which can spread thanks to 3D print technology, consisting in modelling materials like molten metal or plastic and use them to obtain products; since 2007 the uses of this technology are grown, along with production costs reduction (Pîrjan & Petroşanu, 2013; Sniderman, 2012).

A less costing production of objects of daily use, more personalised and efficient, could represent a revolution for consumption. A new role will emerge from this process: the so called "prosumer", an individual who produce and consume goods produced by him. This technology allows common people to meet their needs thanks to the possibility to personally design the product. Personalisation is a very important feature, because it makes demand and offer to meet in the most effective way.

The software needed for production helps the transition from manufacture production to digital production, because the role of human beings is almost non-existent. Also if open source model is implemented in contrast with license models, 3D printers will be even cheaper instruments, because paying expensive licenses for producing an object won't be necessary. Internet could allow project sharing throughout the world, creating a global digital market in which people share and exchange ideas and products in a faster way. The transition from a mass production to the production by masses could be a reality (Hoda, 1997; Gandhi, 1936)

- **Education**

Education is probably one of the sectors what will see some relevant changes too. One of the issue regarding education is accessibility; in major countries of the world, higher education remains accessible only to the high-income segments of the population (IREG, 2010). This issue reflects in socio-economic inequalities. In fact direct relation between education level and income level has often been demonstrated by empirical evidence in both USA and Europe (Strauss, 2012; Rodríguez-Pose & Tselios, 2008). Therefore, it is difficult to change a situation where individuals who are born in wealthy families or countries can afford a similar lifestyle to the one they are born in, thanks to the fact that they could obtain a high-level education; while those who don't get these possibilities risk to remain in the so-called poverty trap, not being able to change their economic condition.

A solution to this problem could be found through Internet; an experience of recent development is the MOOC (massive open online courses), namely "online teaching programs in open access which offer university credits to millions of students in exchange of little no costs" (Kaplan & Haenline, 2016). Thanks to MOOC it is possible to establish a new education model through which anyone can access to high level education, democratising knowledge; democratisation of knowledge can contribute to overcome obstacles that don't allow low social classes to access to a decent lifestyle. The whole higher education sector could be redefined by this phaenomenon, because universities would have to represent a more attractive choice for students, who have the option to access to the same level of instruction but with fewer expenses. Teacher's and professor's role would have to be different in order to offer something that online courses can't.

- **Communication**

  In the first chapter Internet is presented as the infrastructure that already plays some important roles throughout society. If the infrastructure will be developed on an even more global scale, overcoming access barriers and achieving cheaper prices, traditional forms of communication between people won't be a match against it. The social network platforms already allow free text communication and content sharing, and the most common of them have earlier developed new functions for calling. Since 2003 Skype offers a free phone service that has been a very challenging competitor for traditional phone companies (Heinrich, 2014). The issue regarding the widespread of these tools is the connection infrastructure: to become the predominant way of communication constant access to the web is needed, which could be achieved through fast and open broadband lines, not limited to the building's environments but spread throughout the whole city. An ecosystem of permanently connected devices can take the place of traditional phone lines, representing a cheaper and almost free solution for people to be in contact with each other.

The consequences of technological development could be so relevant that the economic paradigm in which we live, capitalism, would need a redefinition or could be replaced by other way of organizing the economic activity. Because the concept of capitalism, along with its definition, is often debated from a theoretical point of view, it is hard to define when exactly it became a widespread economic model throughout society. To overcome this issue capitalism is here considered as the appropriation by capitalists of the means of production, which were earlier property of the working class. This shift of propriety happened around the end of XVI century, began with the enclosure of the fields and led to the aggregation of workers in factories owned by capitalists in the following centuries (Slater, 1968); the centralisation of propriety in the capitalists became common because it represented the best way to maximise productivity thanks to the verticalisation of the production process. The core force of the capitalist paradigm is the research of maximum productivity and efficiency, in other words the maximisation of the production-resources relationship which is determined by how much labour and machines can produce in relation to their cost. Thanks to the new inputs here described, such as new energy sources and more productive machines, production is rising to levels never seen before and costs for production are decreasing among various sectors.

The problem is that the achievement of the so wished maximum productivity could be the destruction of the paradigm that is founded on such goal. If the costs for producing goods will reach values near to zero, prices also will decrease, presenting a serious problem for profit. If profit can't be achieved anymore because goods are almost free, then production will stop because companies can't earn from their activity any longer. If the capitalist production process collapses because it has exploited his full potential it means that it is no longer an efficient way to provide goods and to organize economic activities. It's hard to predict when this point could be reached exactly, but for sure this could be a consequence of the changes we are witnessing.

Not so more positive perspectives can be advanced about work and unemployment. If the scenario that some researchers have predicted will be confirmed (Frey & Osborne, 2013), human work will see a season of great unemployment as machines and computers will take the place of people in some sectors that in developed countries give jobs to a relevant part of the population, like manufacture, services, knowledge and entertainment. There are some clues that this change is already happening; an interesting fact, which is representative of the things just said, concern the manufacture sector; in the time period between 1997 and 2005 manufacture production in the United States rose of 60%, while between 2000 and 2008 3,900,000 jobs were lost in the same sector. The reason could be appointed to the technology that was implemented in the production process, which requires less human work but it represents a more productive solution. Mark Perry, economics professor, analysed GDP and employment data in the United States, and found out that after the Great recession of 2008 the crisis war reabsorbed at the end of 2012 because GDP grew 2.2% from 2007 to 2012 (Fig. 3.5). The interesting fact is that while GDP saw a substantial growth, the number of workers in industry decreased of 3,840,000; so growth was achieved with less human work because companies found cheaper and more productive ways of producing (Perry 2012). The steel industry is a clear example of this trend: between 1982 and 2002 steel production grew from 75 million tons to 120 million tons, while the iron and steel industry workers drop from 289.000 to 74.000 (Joseph, Meadows & Fresco, 2009). From a broader point of view it can be observed that between 1995 and 2002 global economy production grew of 30% but industry lost 22 million jobs. If machines become even cheaper than workers in countries where manpower costs are very low, the process becomes relevant to a global scale and not only confined to some particular countries. That is what it happened in China where 16 million labourers lost their jobs (Baum, 2013).

If the automatization of production processes will expand to more and more sectors, mass unemployment will represent the most important problem to deal with. Capitalism will risk to fall if there will be much fewer jobs, because the demand of goods will be too low due to the less money disposable to people. High productivity itself won't be necessary to meet demand; the offer could outdo the demand and companies will have to produce less. A high-productivity world is sustainable only if mass consuming is the rule, a scenario that won't likely become reality if unemployment will spread throughout society. Mass production could also be useless because of self-production. 3D Printing technology could be the new system through which the population, both private citizens and companies, produce the goods they need; it is difficult to imagine that most of production will depend on 3D printer, but for sure it will be an interesting option for domestic environments.

In conclusion, the third industrial revolution that Jeremy Rifkin imagines could break the link between productivity and employment; these two aspects will become antithetical as more productivity will be achieved using machines, cheaper and productive than humans, instead of creating jobs for people. However the new industrial revolution isn't achievable if important infrastructure that will exploit the potential of Internet and new energy sources won't be built. The core mechanism required for changing society and economy is the energy-communication new system in which Internet and renewable sources play these roles.

Figure 3.5.U.S. GDP Annual Growth Rate

Source: www.tradingeconomics.com; U.S. Bureau of economic analysis (2017)

PART 1 – conclusions

I have now highlighted the most important elements and dynamics that, from my point view, are transforming economy and society as a whole. Chapter 1 describes Internet potential for the economy as an infrastructure that enables connections between people and things. Chapter 2 follows presenting some ways data can be used as a raw matter to create social and economic value. The scenario resulting from the widespread of these elements, and also others, is the digital world that Rifkin analyses focusing especially on the consequences and the challenges that might present in the future. From this perspective Internet and data can be instruments for immediate growth and development as showed in the first two chapters but also part of a future scenario that could present enormous challenges, very hard to manage; for example, mass unemployment will be hard to tackle if industry won't require human labour anymore. If work demand won't be heightened by the rise of new sectors that need human work, countries will have to face huge waves of poverty; this rise is now unlikely to happen from the facts that are known, because new economic sectors don't require so much labour compared to the traditional ones.

To sum up, this first part of the work has the role to present the scenario where a public policy like an Open Government Data initiative could take place. The public sector can certainly play an important role in the era of information and digitalisation. Understanding how this role can be best interpreted through the analysis of the context is one of the goals of this first part. Open Government Data is one of the action that public sector can pursue to be aligned with the digital revolution earlier presented, so it is interesting to reason on such theme in order understand if and how public subjects can contribute to the development of the digital revolution. The second part will focus on such theme, discussing Open Data from different points of view and focusing finally on the Italian experience.

SECTION 2 Open Government Data

Introduction

Open Government Data has become a strongly debated theme since the US ex-president Barack Obama included the publication of data in open format in the federal government agenda (Obama 2009). Because OGD became an issue of worldwide interest, experiences in such field grew over time and rose as popular practices. Today some countries, like USA and UK, have reached a certain degree of maturity concerning Open Data, while in others data publication processes are yet to be fully implemented (*opendatabarometer.org, 2015*). In this part of the work the Open Government Data phaenomenon is explored and analysed, identifying the theoretical aspects on which it is based and presenting some examples and case studies. Impact derived from OGD is also described, and some quantitative and qualitative insights are given, but challenges and issues are not eluded (Chapter 4). In Chapter 5 the focus is on the Italian experience, comparing the peninsular context with the international context and trying to assess the state of Open Data in the country.

There are two different approaches concerning the way of discussing Open Government Data: the first is focused on the claim for transparency within governments and public sector; Open Data is considered a useful tool to enforce the citizen's right to be informed on what and how the public sector acts. The other approach focuses on the socio-economic value that Open Government Data initiatives can generate, being data a raw matter that can be used for different purposes. These two approaches are represented by two different movements, the Right to Information movement and the Open Government Data Movement (Ubaldi, 2013). The analysis here is more congruent to the latter approach.

Right of Information movement organisations:

CIVICUS

OPEN SOCIEY INSTITUTE

Open Government Data movement organisations:

OPEN FORUM FOUNDATION

SUNLIGHT FOUNDATION

OPEN KNOWLEDGE FOUNDATION

CHAPTER 4

Open Government Data: a theoretical and empirical analysis

4.1. Open Data and Open Government, definitions and history hints

The origin of the Open Data's phaenomenon within the public sector can be seen nearly fifty years ago, but important developments were reached only in the Nineties. The first occasion in which Open Data was mentioned in this context was in the Seventies: in order to monitor American satellites, a collaboration between international partners was put into action; this collaboration needed the adoption of an Open Data policy by the participants to the program, so that data was available to all the subjects for a more effective collaboration (Department of State, 1976).
One of the pioneers of Open Data science was Jim Warren, open government activist and West Coast Computer Faire founder, who "*show(ed) California Assembly Member Debra Bowen how public access to state legislative records could be accomplished via the Internet at low cost and high benefit to the public*" (Warren, 2001).
But the time when Open Data really began to play a role in the political agenda corresponds to the election of Barack Obama as president of the United States. During the presidential campaign he already promised to "*restore the American people's trust in their government by making government more open and transparent*" (Change.gov, 2012a). He created a group called TIGR, Technology, Innovation & Government Reform Policy Working Group, to implement his Innovation agenda which aimed to:

> *"[…] create a 21st century government that is more open and effective; [that] leverages technology to grow the economy, create jobs, and solve our country's most pressing problems; [that] respects the integrity of and renews our commitment to science; and [that] catalyzes active citizenship and partnerships in shared governance with civil society institutions"*

(change.gov 2012b)

During his first day in office he also issued a memorandum for federal public agencies in order to have them make public the biggest quantity of information possible based on FOIA, the Freedom of Information Act approved in 1966. In the Obama's administration commitment to open government status report is said that, thanks to this commitment, since Obama has become president in 2009 strong progresses in terms of transparency and collaboration of the public sector took place; this was achieved through the publication of data and new technology usage that have guided innovation in the sphere of open government, one of Obama's main concerns (Obama, 2009).

Defining Open Government Data is not an easy task. Different definitions proposed over time have blurred the essence of the concept, because of the different points of view from which it can be seen. Social scientists, economists, businessmen, computer scientists, public sector experts and politicians have all showed interest on this matter, so they all have added different perspectives to the concept of Open Government Data. In order to better understand how a precise definition can be found, it is useful to start from the Venn diagram realized by Joel Gurin (2014) (Fig. 4.1). The conceptual framework designed by Gurin allows to identify the fundamental basics of OGD.

1) **Open government**

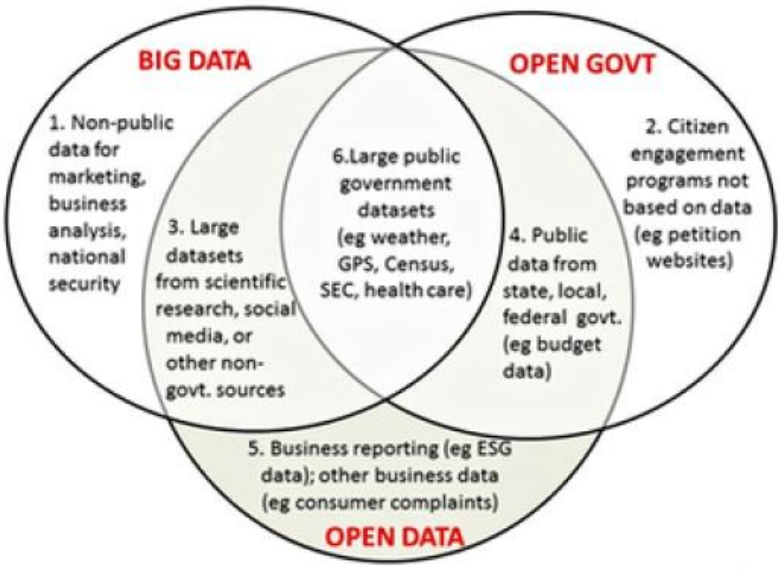
Open government is a philosophy concerning the government activity and the relation between government and citizens; it is based on three pillars: transparency, participation and collaboration (fig. 4.2).

Transparency means that citizens should be informed on what a government does, especially how it spends public money and which policies it promotes. The fundamental assumption is that government activity should be subject to citizens monitoring because governments are elected by citizens, so they have the right to control what the public sector does with the powers and the resources provided by the community.

Open government can be pursued also because it can be seen as a source for social and economic innovation. The more a government is open, the more citizens can monitor public policies and better understand how they are governed. With access to public documents, data and information, a different model of governance could be created, with more participation and collaboration of citizens, who can help to develop better choices for the common good. When governments are able to create spaces and moments for discussion with citizens, they can learn about their needs and inform them on how they want to resolve public issues. In this context, all actors are involved in the public debate and can share their opinions, evaluations, and insights to contribute to a better way of managing the community. One interesting aspect concerning collaboration is that different types of actors — such as companies, public authorities or simple citizens — have all a different know-how to share, so that a matter could be discussed considering different point of views; therefore, the debate is enriched by different insights and could lead to more effective policies.

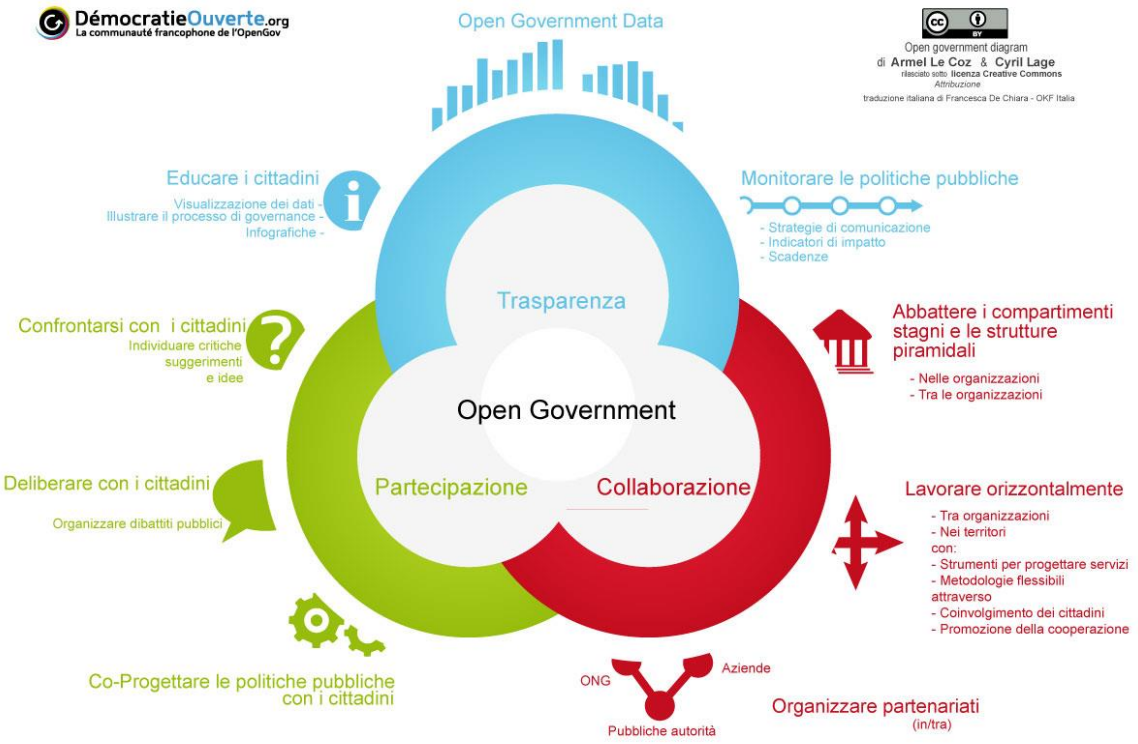
The origin of the Open Government concept dates back to the Enlightenment period, when state secrecy was questioned by some philosophers (Habermas, 1962). The expression became popular in the second half of the 20th century when some countries began to promote freedom of information laws. The most popular example is the USA with the approval of the Freedom of Information Act (FOIA) in 1966 (*foia.gov*). The concept has been subject to new definitions due to the new instruments available in the last years; the true driver for Open Government is now technology. As Internet technology rose and became the most popular means of communication, the number of possibilities to make governments open and to share information grew.

Figure 4.1 Joel Gurin Open Government Data Venn diagram



Source: Gurin (2014)

Figure 4.2.Open Government definition



Source: DémocratieOuverte.org, translated by Francesca De Chiara (2013)

Open Government developed a strong link with technology, especially with the concept of Open Data, which represents the main focus today when such themes are discussed. Making governments more transparent and open means that government data should be published in a way that it can be

reusable by everyone. It is also important to identify which data goes under the category of 'government data'. A commonly agreed definition considers government data as *"any data and information produced or commissioned by public bodies"* (Ubaldi, 2013). Not only data produced directly by governments is part of the Open Government Data category, but also that produced by public agencies, so that very different types of data can be available for citizens. The more different information and data types are available for the public, the more potential for reuse is available. Public administrations cannot realize how important certain data is for civil society if they do not implement mechanism of discussion through which citizens' demand is understood.

From the evolution of the concept of Open Government, it is easier to mistake the use of new technologies by governments with real open government policies. In this case the link between Internet and Open Government could be misleading; developing web sites or modern tools to enforce transparency, collaboration and participation is not enough. If a regime wants to become open it doesn't simply need to publish more data, but it is important to focus on which types of data could foster open governments' three principles and on how those data should be published. The regime becomes more open and accountable when stakeholders are more informed and participating to public activities, not necessarily when they have more data (Yu & Robinson, 2012). The concept of Open Data clarifies this confusion between open government and technology, and poses solid bases for the creation of innovative tools for socio-economic development.

2) **Open Data**

There are different points of view from which the concept of Open Data can be discussed; the goal of this section is to identify the essential characteristics of the definition and to understand what Open Data means from a technical perspective. First, a definition has to be given in order to analyse the concept. The Govlab of New York has done the most useful work in this sense, analysing different definitions and comparing them to reach a comprehensive definition. The resulting concept from this work is the following:

> *"Open Data is publicly available data that can be universally and readily accessed, used, and redistributed free of charge. Open Data is released in ways that protect private, personal, or proprietary information. It is structured for usability and computability."* (Verhulst & Caplan, 2015)

The first important characteristic of this definition is public availability of data; it is obviously necessary that data be accessible to everyone, being a public resource from which anyone can benefit in pursuing their goals. No one can be excluded from the use of such resource; this is exactly one of the innovative elements of the concept: making data open doesn't mean providing databases after a specific request from a particular subject or institution, but to leave freedom of access to anyone who wants that data, without asking formally to the data owner. This condition

makes data usage faster and freer, from which a faster development of services, products and innovation can arise.

Another fundamental characteristic regarding Open Data is the free circulation, flow and redistribution of data. The fact that data doesn't need to be supplied in exchange for money and that it can reach third parties, increases possibilities for reuse in an exponential way, it helps enrichment and quality improvement of data, thanks to the amount of users which can benefit from and contribute to it. As a matter of fact, users cannot only take the data and use it, but if it presents some errors it can be corrected by them and re-published by the data provider. This possibility doesn't change the fact that data providers must provide good quality and complete data; otherwise, the policy could result in failure instead of representing a resource for innovation.

The two features just presented concerned the issue on what 'open' means in terms of the characteristics that data should have when published, in order to fulfil the concept of Open Data. The definition of Verhulst and Caplan adds more elements which aren't taken for granted when speaking about Open Data, and can potentially overcome some of the issues raised in these years on such matter. These elements concern more the way that data should be published, rather than the specific characteristics of the data.

The first element consists in a fundamental condition: data should be published following private and personal information protection criteria. This aspect is worth underlining, because it faces and clarifies a latent ambiguity which lays in the concept of Open Data; transparency and the process of data opening has been sometimes interpreted in an erroneous way, by mistaking them for the disclosure of private information, based on an understandable concern about privacy. Instead, it is necessary to highlight that from this comprehensive definition, data destined to opening are not the personal ones; only data that doesn't clash with privacy protection legislation is potentially subject to opening.

The other element defines data as "*structured for usability and computability"*. Release of data and information in less structured ways makes them less usable, especially by data elaboration softwares. More efficient usage requires standardisation of data; therefore time can be spared in the process of collecting, merging and constantly controlling the dataset structure. This goal presents practical obstacles and is very hard to pursue when data needs to be collected from different sources, because standards, captions, models and formats of presenting data change depending on the geographical contexts and the organisations that divulge it.

The Govlab definition deals with many issues regarding Open Data which were not immediately clarified when the concept spread in the international debate. However the characteristics listed above have to be defined in a more practical way, because they need to guide to common practices in order to express a more homogeneous disclosure of data throughout countries and organisations. For this reason it is necessary to provide purposely technical standards to achieve this goal; even

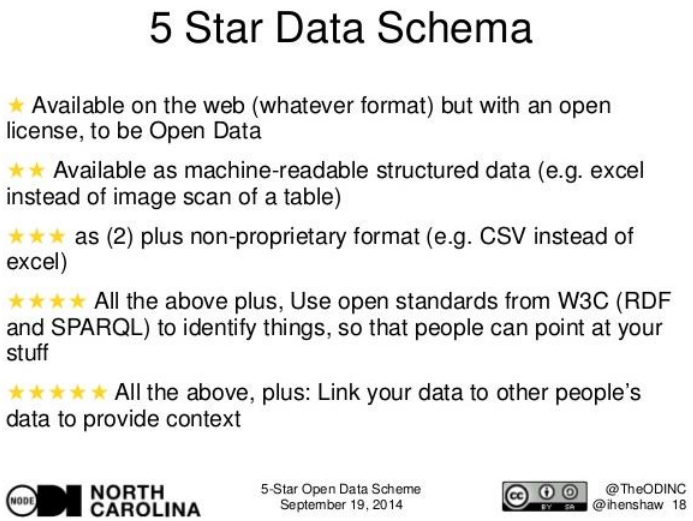
though it defines clearly the concept, the presented definition can be subject to different interpretations which endanger the required homogenisation of practices. An organisation can certainly start from such definition to set up an Open Data process, but, in order not to diverge and take a different direction from the one recommended, it is necessary that it knows and understands the precise procedures through which managing the process. The most popular standard concerning Open Data is the one designed by Tim Berners-Lee, World Wide Web inventor. Berners-Lee is one of the people who have most supported Open Data and given popularity to the concept: *"Open Data and a change of mind-set is the next step in the Internet revolution"* (Berners-Lee, 2013)

He designed an Open Data quality scale, the so called "Open Data five stars".
This type of scheme seems very useful because it provides a very clear idea on the best way to make data open. The five stars meet also the criteria of Verhulst and Caplan definition:

1. The first star defines data opening and reuse focusing on the license through which data should be published. The license that allows data to be fully open, so taken and re-used without breaking copyright laws, is the so called CC0, Creative Commons 0 (*creativecommons.org*).

2. The second star identifies the data format that should be used to be readable by machines; data should be structured so that content is identifiable. A software can analyse and process data only if this condition is met, so this feature becomes fundamental for using data in effective ways. As a matter of fact Verhulst and Caplan speak of "computability" when defining Open Data.

Figure 4.3.Open Data 5 stars



Source: The Open Data Institute of North Carolina (2014)

3. The third star suggests using a non-proprietary format like CSV instead of the Excel format. CSV format has the peculiarity of being readable by open source software too, which guarantee free using, because the license is open. The tool, being free, allows everyone to use it and as a consequence to use data that is available in a non-proprietary format. By contrast, the proprietary format allows only the ones who possess the related software that can read it to use the data. Furthermore, proprietary software can make a format obsolete, through updates and modifications that do not allow elder formats to be read. The feature of universal usability is satisfied by this star.

4. The Openness concept is stated in the fourth star too, which suggests using open standard of the W3C organisation, the World Wide Web Consortium, which aims to "*leading the web to its full potential*" (*w3.org, 2004).* Such standards make merging the data easier, even if they have different structures, through a system of connections and relations.

5. The fifth star stands for the achievement of maximum quality level for Open Data. Its criteria are satisfied when data are connected between them. This type of connection refers to the concept of "Linked Open Data". Linked Open Data is structured and connected data; semantic technology allows using these connections so that through so-called interrogations, specific data can be obtained (Bizer, Heath & Berners.Lee, 2009). Thanks to this system, data becomes easy to find, access and use.

The GovLab definition and the standard provided by Tim Berners Lee's 5 stars can represent the conceptual framework needed to understand Open Data. These are the main instruments that should be the references in every organisation if a process of data disclosure wants to be developed.

3) **Big Data**

Big Data was already discussed in Chapter 2, so it is not necessary to explain what it is and what role it has today in society. Big Data goes along with the discussion on Open Government because it is possible for the public sector to provide databases of relevant dimension, generated by a single source or from the aggregation of multiple sources. As a matter of fact, governments own the resources to produce and collect huge amounts of data, depending especially on how many activities in a country are accountable to the public sector. Being the concept of Big Data susceptible to technological development, which increases the data process capacity of machines, discussing Big Data in the future could become obsolete if using huge datasets will represent a common practice. The eventuality that the public sector will provide Big Data also represents a hope, because:

"*when the government turns Big Data into Open Data, it's especially powerful: Government agencies have the capacity and funds to gather very large amounts of data (...), and opening up those datasets can have major economic benefits*" (Gurin, 2014)

Open government, Open Data, and Big Data: these are the three pillars which sustain the concept of Open Government Data. Only through the implementation of all three a scenario where data really functions as an engine for social and economic development can be achieved. If Big Databases become of public domain, while participation in public life and collaboration between citizens are encouraged, it is possible to establish virtuous mechanisms both in society and the political and administrative system. This is possible because data can be a tool for the emergence of new entrepreneurship and innovation in companies already operating on the market, but also for more effective and efficient public services.

4.2. Why Open Government Data?

From the beginning of the new millennium, public administrations have begun to experience new technologies more intensively, using the Internet more frequently and according to the digital revolution in progress. This change has increased the possibilities for reuse of public information for different purposes, fact that has created more interest on data and information in civil society and companies, as well as in the public sector itself (Iemma, 2013).

Why any public agency or public administration should launch an Open Data project to make the information it owns publicly available is certainly not an easy question to answer, because understanding the opportunities that such policy can generate for a certain social context is not an easy task. For this reason, potentialities should be analysed and tested in order to understand the real value that data in open format provided by the public sector can be produced.

Before studying real cases implementation of Open Data policies and examine in depth the utility from a practical perspective, in this section the benefits that from a theoretical point of view can be gained from OGD, are summarized. These benefits are considered in this work the most relevant reasons which are worth focusing on to support the opening of data, even though the action of opening public information to citizens has an intrinsic value itself. Informing people on the activity of public agencies is fundamental because such agencies obtain their legitimacy and resources from the population of the context in which they operate. However, supplying data to citizens represents a further step, because it embraces the open government philosophy and overcome the traditional schemes of the administration-citizen relationship. Opening data is therefore a citizen's demand that cannot easily find response if only the common pressure to government for more transparency and accountability is brought; alternative reasons have to be found to sustain such demand. Nonetheless traditional pressure for more accountability should continue. Accountability remains a fundamental pillar of the representations system, which allows power to be delegated to public institutions; activity of such institutions should be constantly monitored and evaluated because public officials have a responsibility towards the community.
But the discussion on this matter should also evolve by putting new inputs and arguments on the field, such as the social and economic benefits that can be gained. Considering the Italian context,

this debate is taking place in a situation where public administrations cannot really have many resources to invest in innovative policies, especially if the outcomes of such policies aren't still quite clear at the beginning of the process.

For this reason, the focus here is on the effects that Open Data could generate, because they are considered more solid and valuable reasons to justify this kind of policies; reasons that are harder to find in transparency requests expressed by citizens. Dealing with Open Government Data putting forward the impact, especially on government itself, is an easier way to encourage its implementation.

The Economist (2013) has called Open Data "a new goldmine". There are different types of reasons which can support this statement; the purpose in this work is to classify them in the most complete way possible. Iemma (2013) explains which are the fundamental characteristics of Public Sector Information (PSI) and which types of value are associated to it. Value of PSI can be described with three different reasons:

- Public information is produced in a continuous way, it is unique in terms of thematic coverage and volume. Its production is linked to the public mandate; it is therefore constantly updated and based on the public needs. Data produced is not easily made available from other sources.

- PSI and its availability are not related to the demand for reuse of other subjects, but only to the internal use.

- Purposes and uses of data produced by the public sector form a large set, which contains also elements that go beyond the use originally programmed by the producer; this point was derived from the nature of data as a digital good, because it is difficult to imagine all its different possible uses, being a raw matter; the possibility to reuse data increases its value, which for this reason is hard to quantify before its actual use.

Three areas of Open Government Data's impact are here presented (Alvino, 2017; Barbera, 2016; Capgemini Consulting, 2013; Deloitte, 2012; European Union, 2015a; Zotti, 2017).

4.2.1. Open Government Data as the base for transparency and Open government

One of the main discussed aspects of Open Government Data is transparency, because such commitment is highly debated within public institutions. Transparency can be considered as a right from the point of view of citizens, if they believe they must be informed and made participants of the public administration activity; but the point of view of the public sector is different. Opening data owned by a certain agency to the public will irremediably results in more exposition on its

activity, which could cause concern and pressure on those who are accountable for such activities; the exposition can be sensed as a form of control, a control from "below" which adds to the forms of control that already exist. Sensing this constant control cannot be well received from the workers of that agency, which from their point of view see transparency as a threat that could endanger their work and their career. In fact control can be perceived as a lack of trust in the institution.

For a public agency, being accountable on its activity and resources is a fundamental principle because such activities should be developed following the public interest. This doesn't necessarily mean that public agencies have to be more transparent and publish, make available and reusable to everyone the data they constantly produce. In fact the effort demanded to a public subject when dealing with Open Government Data goes beyond accountability; it is demanded to make public information that do not concern only the agency itself, but all that which is produced and collected by it. This process can be activated only if a change of mind-set takes place, both in the public sector and the individuals which the information is destined to:

- Public subjects must see this type of policy as a provision of resources to the citizens which can set in motion virtuous mechanisms necessary to make public service more efficient and effective, instead of an alternative form of external control of their activity;
- Citizens, and in general all the subjects that can benefit from data, must be part of the public action in order to improve it, being also important suppliers of information and useful feedbacks to a better understanding of the reality they live in (Barca, 2015).

Therefore, transparency itself can produce its full potential impact and benefits only if the other two pillars of open government — transparency and collaboration — are simultaneously stimulated and put into action. Open Data is a central point of junction for the realisation of two of these aspects: it makes public action more transparent, because citizens have a more direct contact with it thanks to the data available; and it improves participation of the actors who are part of a certain social context, because their contribution is more informed thanks to the information they get, therefore more valuable in terms of quality. The most important feature of Open Data which is able to support this idea is reusability. If data is published in a way that it cannot be reused by users, a big part of the benefits that could be obtained cannot be achieved, because users cannot exploit data as a concrete resource but only as information received and not as a potential source of value. For this reason, in the implementation of Open Data policies it is crucial to keep Tim Berners-Lee's five stars as the main standard, because they have been thought to make data released in the most useful way.

4.2.2. Open Government Data for economic value generation

One of the reasons that are often presented to support Open Government Data policies concerns the potential economic value of the information produced by the public sector. Without an Open Data

policy, access to such value is limited to the information owner; in fact governments can decide to deliver their data only through a commercial system. It is clear that a limited access per se constitutes an obstacle for the possibility to use data by anyone who wants to, therefore for the creation of value. However, making data usable by anyone can represent a useful move for generating economic value; citizens, organisations and companies can take data, process it in their own way and eventually build a new business on it. From data processing can in fact emerge an intermediate good which can be translated in a product or a service; moreover, due to the fact that different types of data processing exist, from the same data different types of uses, and therefore different products and goods, can be developed (Iemma, 2013). A new road for innovative start-ups, new businesses, services, products with a consequent economic value, employment and growth can be made available. The first aspect on which Open Data can have an impact on is therefore the creation of business.

Among the economic benefits of open government, data efficiency has to be included. Data is a mean for acquiring knowledge; therefore, from the acquisition of a higher volume of data it is possible to operate in more efficient ways in a variety of sectors, both public and private. Various types of data provide knowledge which is reusable in the corporate context to reduce waste of resources. Also a policy can be more efficiently designed, reducing the waste of resources caused by misinformation, thanks to more complete data regarding the socio-economic context in which an institution operates; it is important to underline this aspect to understand why the data-resource made reusable by the public sector could bring benefits both to suppliers and users of it.

4.2.3. Open Government Data for decision making and quality of information

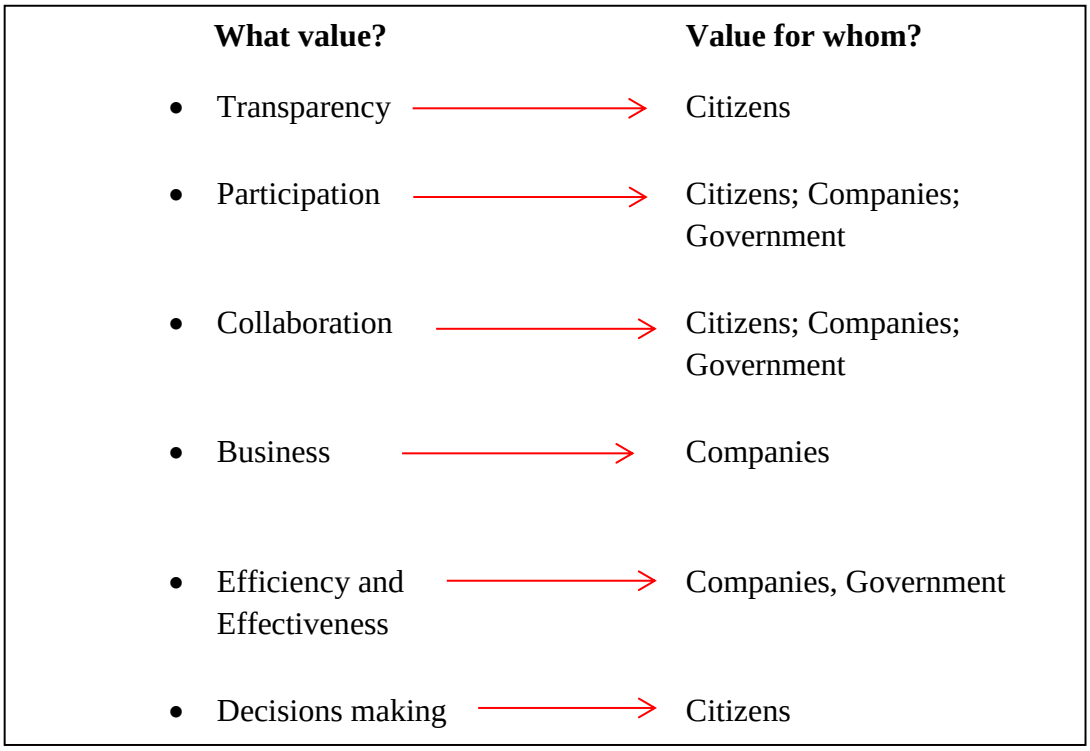
Open Government Data is a useful tool to improve decision making, both in public and private field. Citizens get information and use it in their daily life to improve the quality of their decision; thanks to Open Data, but also to the whole data driven scenario, the quality of life can improve through these better data-based decisions. Possessing information concerning aspects of daily life allows citizens to change their behaviour. For example, public transportation data can potentially help people save time for moving; citizens could choose the means and times of transport based on their own preferences: they could opt for the fastest, the most eco-friendly, or the most comfortable way. The more this data is aggregated, the more it is useful for individuals to address their choices.

The information sector is facing a radical change, because of its dimension and its speed. Most part of information comes already from the web, due to the digitalisation of knowledge and low cost that it offers. In this scenario, information channels have multiplied and for this reason, collecting true information is becoming more and more difficult. A branch of the information sector which uses data to analyse facts through the scientific method, is data journalism. Researches and inquiries of such type of branch benefit from an increased availability of data, especially from the governmental one due to their importance. The information sector can have a fundamental role to play in the

publication of facts that emerge from the analysis of data regarding a public administration, allowing democratic control to be stronger and more effective. It is hard to imagine common people to directly access huge databases to be informed on a certain matter of their interest; for this reason, the press and the whole information sector have to be the intermediary between information and citizens, because of their know-how. Transparency and accountability need a system where the subjects that possess these means for understanding information provide citizens that information in a way that is understandable. Without doubt in the next years, as the databases proliferate, it will be necessary to train journalists to acquire data skills for a better use; this proliferation of data can bring society to more accurate information, less susceptible to interpretation. Impact of Open Government Data on the information sector concerns the quality of information, improved by the presence of more databases. It should not be forgotten that also academia, but more generally all the research activities, could highly benefit from data published by the public sector. Figure 4.5 tries to summarize the insights about Open Government Data value just presented.

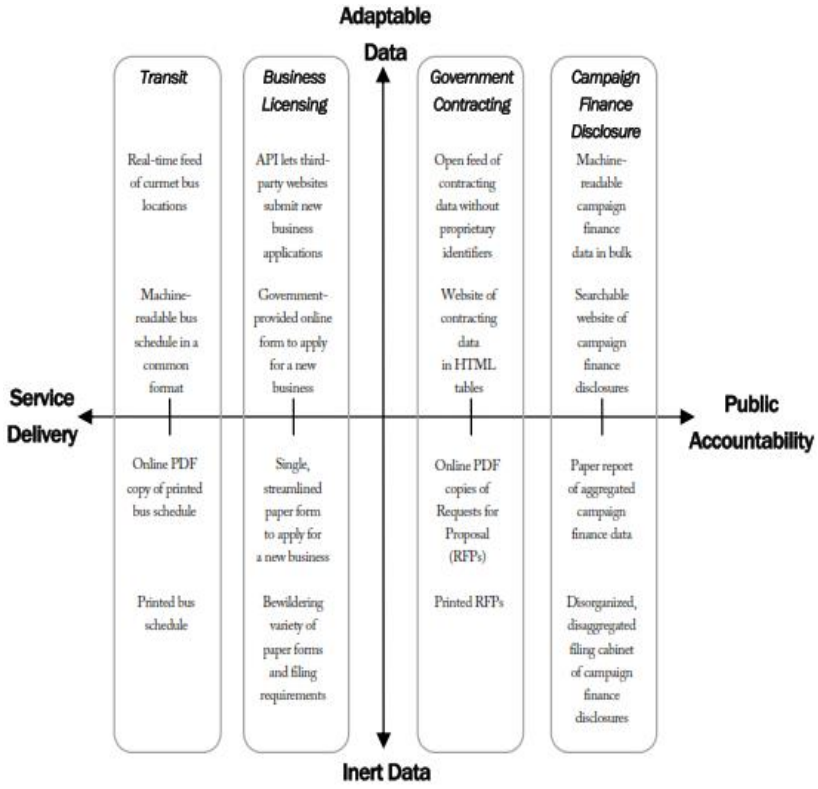
After having focused the main types of impact that Open Data is able to generate, it is interesting to look at the reasons based on which a public subject can develop a process for publishing its data. Yu & Robinson (2012) build a diagram in which the vertical axis identifies the technical aspect, and the horizontal axis describes the goal of the subject in opening data. The resulting diagram is a simplified representation of the framework in which Open Data policies can be developed; some examples of OGS policies are put in the framework to understand their positioning regarding the technical aspect and its goals (Fig.4.6).

Figure 4.5.Open Government Data value



Source: personal elaboration

Figure 4.6.Yu & Robinson Open Government Data framework



Source: Yu & Robinson (2012)

4.3. OGD types and uses

Up to this point Open Government Data has been considered as a single set of information, without any distinction for its typology, use and characteristics. This simplification has been made in order to provide a first theoretical look on Open Data and especially on its potential for value generation. It is clear that that reality is more complicated; data produced by public agencies is in fact very varied because the public sector plays an important role in society as data producer and collector. Fields of public intervention are linked mostly to the type of state considered: within countries with a wide public structure, types and quantity of government data produced are significantly high; instead, within countries that opt for a minor public intervention, areas of public interest are fewer, and as a consequence information produced. Despite this distinction, also in traditionally liberal countries like the United States relevant volumes of data can be produced and collected. There are in fact different factors that influence the ability of a public system to become a relevant data producer, but regarding the number of data types produced, the degree of public intervention across society is for sure a relevant one.

Now an excursus on the various typologies of data is presented to understand which type of value can be actually generated through Open Government Data. Indeed, such value is linked to the type of data that is made available in open format; for this reason, it is important to base the analysis on

these typologies, and to understand which benefits they can bring. Ubaldi (2013) considers the following data types:

- **Business information**, including Chamber of Commerce information, official businesses;
- **Registers, patent and trademark information** and public tender databases;
- **Geographic information**, including address information, aerials photos, buildings, cadastral information, geodetic networks, geology, hydrographical data and topographic information;
- **Legal information**, including decision on national, foreign and international courts, national, legislation and treaties;
- **Meteorological information**, including climate data and models and weather forecasts;
- **Social data**, including various types of statistics on economics, employment, health, population, public administration;
- **Transport information**, including information on public congestion, work on roads, public, public transport and vehicle registration.

Source: Ubaldi (2013)

Every typology of data can generate different sorts of value, because data can be used in different ways; also beneficiaries of this value change based on the type of data, but it is difficult to identify ex-ante which typologies can be more useful for citizens, instead of companies and governments. But concerning the production phase it is clear that the central administration role is different from the local administration's role; both central and local governments are information producers, but of different types and volume. The cause of this difference is the set of means available to a certain public subject to fulfil its mandate, as well as the entity of the subject; a local administration is able to collect more complete data on its territorial context, especially on its inhabitants and on the structure of its territory. Subjects depending on the central administration are instead able to collect a bigger volume of information, due to the more powerful and wider means. Furthermore, an agency that depends on the central level can carry out the role of data aggregator more effectively: it can collect all the information that is generated on its territory, in order to create one or few access points for all the data. It is also desirable that such role is effectively played by central governments, because it can avoid the fragmentation problem: fragmentation of access points in addition to the lack of a point of conjunction between them makes data access less efficient, especially if different platforms use different standards of publication modality and data characteristics. The process for value creation has to be made the more efficient possible through aggregation and standardisation of data.

Continuing the research that this work is taking forward to understand how Open Government Data is able to generate value, it is now necessary to identify which possible uses can be developed from it, starting from a technical perspective. Basically, data can be used in 5 ways, listed exhaustively by Davies (2010):

- **Data to fact** – often underestimated in accounts of 'data for developers' – individuals may seek out specific facts in a newly Open Dataset. These facts
  may support their engagement in civic or bureaucratic processes, or in business planning. Facts could be found through online interfaces, but also by browsing downloaded Excel spreadsheets.

- **Data to information** – creating a static representation and interpretation of one or more data sources. Leading to visualizations, blog posts, infographics and written reports.

- **Data to Interface** – creating a means to interactively access and explore one or more datasets. For example, creating a searchable mapping mash-up, or providing a tool to browse a large dataset and crowd source feedback or scrutiny. Interfaces often also include 'static' interpretations of data (data to information) – showing particular summary statistics or algorithmically derived assessments of underlying data.

- **Data to data** – sharing derived data (either simply an original dataset in a new format, or data that is augmented, combined with other data, or manipulated in some way). A whole dataset may be shared, an API onto a dataset created, or an interface that makes it easy to download subsets of a large dataset.

- **Data to service –** where OGD plays a 'behind the scenes' role in making some online or offline service function. For example, the use of boundary data to route messages reporting potholes to the responsible authority.

Source: Davies (2010)

Despite the fact that publishing data, as already said, is a task for the public sector, an important role in the provision of data is played by private parties. Third parties can process raw data in many different ways, and so provide a more effective service, because data itself could hardly reach and inform them without any sort of mediation. The role of the public sector should be limited to providing data in the more reusable way possible, instead of address data usage by citizens only to reach certain goals (Robinson et al., 2009). Here's a list of the operations that private parties can perform using Open Government Data, in order to reach citizens with it in a more effective way:

- *Advanced search:* The best search facilities go beyond simple text matching to support features such as multidimensional searches, searches based on complex and/or logical queries, and searches for ranges of dates or other values.
- *RSS feeds:* RSS, which stands for "Really Simple Syndication," is a simple technology for notifying users of events and changes, such as the creation of a new item or an agency action. The best systems could adapt the government's own feeds (or other offerings) of raw data to

offer more specialized RSS feeds for individual data items, for new items in a particular topic or department, for replies to a certain comment, and so on.

- *Links to information sources:* Government data, especially data about government actions and processes, often triggers news coverage and active discussion online. An information service can accompany government data with links to, or excerpts from, these outside sources to give readers context into the data and reactions to it.
- *Mashups with other data sources:* To put an agency's data in context, a site might combine that data with other agencies' data or with outside sources.
- *Discussion fora and wikis:* A site that provides data is a natural location for discussion and user-generated information about that data; this offers one-stop shopping for sophisticated users and helps novices put data in context.
- *Visualization:* Often, large data sets are best understood by using sophisticated visualization tools to find patterns in the data. Sites might offer users carefully selected images to convey these patterns, or they might let the user control the visualization tool to choose exactly which data to display and how.
- *Automated content and topic analysis:* Machine learning algorithms can often analyse a body of data and infer rules for classifying and grouping data items. By automating the classification of data, such models can aid search and foster analysis of trends.
- *Collaborative filtering and crowdsourced analysis:* Another approach to filtering and classification is to leverage users' activities. By asking each user to classify a small amount of data, or by inferring information from users' activities on the site (such as which items a user clicks), a site might be able to classify or organize a large data set without requiring much work from users.

Source: Robinson et al. (2009)

The point that is stressed out here is that opening data is not the only operation required to make data an innovative tool for transparency and development: if these goals are part of the public agenda, a more complex process that requires private parties' collaboration must be put into action. In this process, public sector and private actors play different roles. Enabling public data usage by third parties should be the main focus of governments, representing the provider of the raw matter that it produces and collects. Privates can mediate between governments and the civil society by developing instruments to make data more useful, and a real source of innovation both for citizens and companies. This process is necessary because processing raw data could represent a hard and long task for subjects that don't have the time or the competence to do it, so they could need easier and faster ways to actually benefit from it.

4.4. OGD, from theory to practice

In the previous sections Open government was analysed as a source of value from a theoretical perspective, which is useful for identifying the possibilities that an Open Government Data ecosystem could create for citizens, companies and governments. It is clear that in theory all the mentioned actors can benefit from Open Government Data, and that value can be generated from data, but a more interesting line of research regards the relation between theory and practice: how data is being used and which subjects take actual advantages from its use around the world constitute two fundamental questions of such research. Literature still suffers from a lack of material on this inquiry, due in part to difficulties in the identification of evaluation methods and impact indicator that can analyse and evaluate Open Data activities. The different kinds of impact that an Open Data policy has on society, requires different types of measurements to be taken, in order to include all the aspects that are influenced in a certain way by such policy. Therefore, there is a technical challenge for Open Government Data to be evaluated.
Furthermore, it should be said that Open Data policies are very different throughout the world, and so it is sometimes hard to make a comparison, especially because of these aspects:

- Open Government Data policies involve different levels of government: an Open Data policy can be promoted by both local and central governments, having, as a consequence, a different potential impact in terms of intensity and beneficiaries. A local government Open Data policy can enable value mostly on its context, because data that is collected concerns for the most part its territory and its population. A local government doesn't usually possess the instruments to generate and collect big amounts of data; therefore, its policy alone cannot affect the other local context of a certain country. On the contrary, a more comprehensive policy on a national scale can produce effects on a bigger scale, because of the instruments that are available at the central level for creating a more useful data value chain.

- Data which is made open by governments can be published in different ways, for example developing low or high quality Open Data policies. Certain governments don't understand the value of making data reusable by everyone, so they publish it in formats that are useful only for browsing and reading. Other governments, on the contrary, follow the literature's suggestion and make data reusable, enabling value to be generated. Therefore, the differences in terms of quality of Open Government Data policies make comparisons more difficult.

- Governments around the world produce and collect different types and amounts of data, depending on various factors. The impact of Open Data initiatives depends most of all on which type of data is made open, so results on impact measurement could be different depending on which public subject is involved. Countries have different potentials in

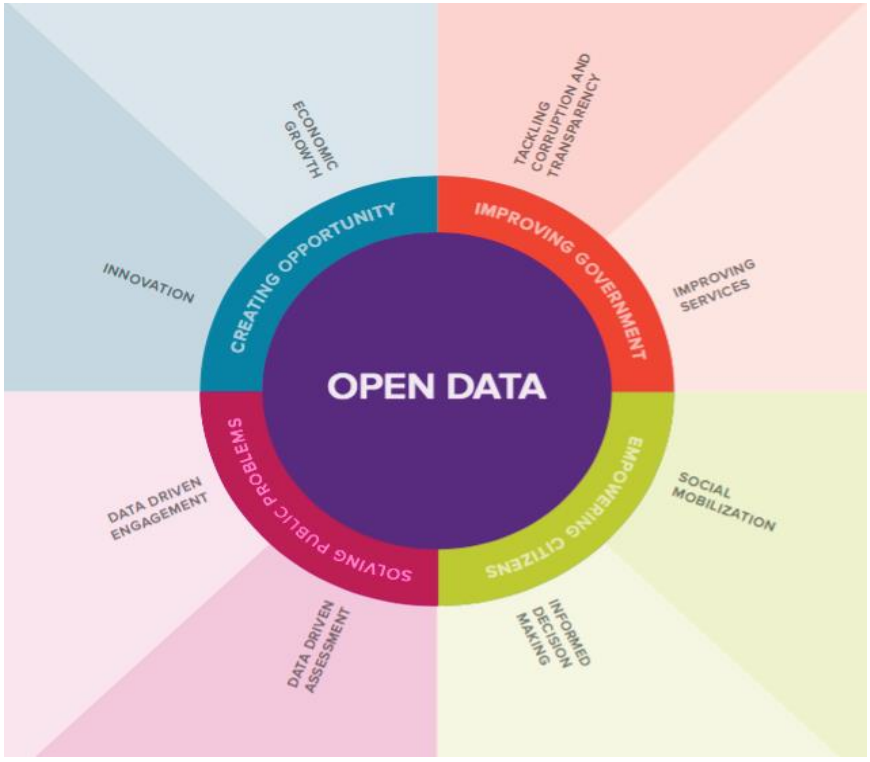
obtaining benefits from such policies, so that it cannot be said that they can generate unanimously a certain amount of value from opening data.

Instead of focusing on looking for a reason to support the idea that Open Data policies should be implemented by all public subjects, it is probably more useful to take a step back and observe how value is and has been created from Open Government Data during the last years. Few are the examples of mature Open Data policies that can be evaluated in a scientific manner, because most initiatives are rather recent in this field. Here a unique global study on Open Data value and impact is used, discussing the value generated by practical experiences from a broad point of view and also through 4 case studies.

4.4.1. Open Data case studies around the world

A few systematic studies exist on value generation through Open Data. The field of Open Government Data has been explored in detail from a theoretical point of view, but how data is actually used still remains a question without a concrete answer. One study elaborated has tried to fill this gap between theory and practice; Young and Verhulst (2015) search for insights on Open Data usage, examining in depth 19 case studies of projects related to Open Data around the world. From this research they build a taxonomy of the ways in which Open Data is used:

Fig. 4.7.Open Data Taxonomy



Source: Young & Verhulst (2015)

It's interesting to compare this taxonomy with the classification of impacts that is presented in section 3.3, which wasn't based on the observation of real examples of Open Data's initiatives.

1. **Creating opportunity**: economic growth & innovation
2. **Improving government**: Tackling corruption and transparency & Improving services
3. **Empowering citizens**: Social mobilization & Informed decision making
4. **Solving public problems**: Data driven engagement & Data driven assessment

The classification made by Verhulst and Young presents four areas of impact, three of which mostly regards the public sphere: "**improving government**", "**empowering citizens**", "**solving public problems**"; the fourth focuses instead on the benefit for the private sector. The difference between the "solving public problems" and the "improving government" areas is not very clear: for example, the impact on tackling corruption could easily go under both categories, so it is hard to identify which impacts are better represented by one or the other category. This issue is understandable if we use figure 4.5 to interpret the pen data taxonomy.
After having a broad look on these two different elaborations, it is clear that the former elaboration bases his classification on the beneficiaries of Open Data value, using the three general categories of 'citizens', 'companies' and 'government', while the latter apparently tries to overcome this distinction. But with a deeper look to the beneficiaries, it appears clear also in the taxonomy, because the three types of beneficiaries are simply presented with different terms: "**creating opportunity**" stands for the 'companies' category, "improving government" stands for 'government' and "empowering citizens" is obviously linked to 'citizens'.
It seems that Young and Verhulst basically add one more general category, "solving public problems", which doesn't identify a specific type of beneficiaries but want to suggest that from the evidence there is a kind of impact that is hard to classify and depends on the particular context in which the Open Data initiative takes place. In fact, the case studies that are mentioned to represent this sphere go from the improved information that helped fighting Ebola in Sierra Leone to the support that Open Data brought in New Zealand to deal with the reconstruction process after the earthquakes. The characteristics that define the belonging to such category could be identified as:

- The public relevance of the issue due to the involvement of the whole community;
- The magnitude of the issue; a humanitarian crisis could well fit this definition.

Therefore, even though the classification by beneficiaries of such Open Data activities is at first sight present only in the first classification, also the second use the same element to divide the various impacts, but with the addition of a fourth category that can be summarized as 'emergency for the whole community'.

Furthermore, both classifications outline the economic value that can be created through Open Data activities. This correspondence means that Open Data as source for creating companies and jobs,

fostering innovation, and enhancing efficiency isn't only an abstract concept, but represents a reality.
Nevertheless, the expression "economic growth" should be used with caution when discussing the new digital markets; the economic value in terms of GDP is evident when new companies and new sectors arise, so that in the short term new jobs are created with benefits for the economy. Nonetheless, it should be considered that in the long term such new companies could replace old companies without a simple migration of the labour force to the new sector; in fact, it is possible that the new companies will not require the same amount of workers, leaving a gap in the market that could be not easy to fill. This issue could take place especially in the new markets, because of the high degree of machine-intensity that they are characterised of. While looking for the most efficient ways to organize the production process, computers represent the element that put in danger the human work force the most, as discussed in Chapter 3. Giving new resources as Open Data for computer to work can be both useful in the short term and dangerous in the long term, because the innovative activities created or helped by Open Data could represent the end for other, more human intensive but less efficient, activities. Therefore, employment and growth could be endangered by the availability of more easily accessible and reusable data. Nonetheless, as confirmed by the global research of Young and Verhulst, new economic opportunities are created thanks to Open Data activities, but in order to be able to speak of growth of a whole economy, these new activities should be balanced with the potential disappearance of activities that benefit from a context in which data is not freely usable, standardized and delivered.

Another correspondence between the classifications stands in the governmental benefits that come from Open Data: services and transparency are improved both in the theoretical perspective and from the evidence presented through the case studies. An additional element is introduced in the second taxonomy: corruption. Transparency and reduced corruption can actually be linked if certain data is published, making difficult for both public personnel and private individuals to engage in illegal activities. This type of impact can be most significant in countries and local contexts where public procedures are often subject to low control and disregard for the rules.

Finally, the last aspect that needs to be mentioned regards the impact on citizens. The two classifications don't collide in such matter too, and identify Open Data as tools for new forms of citizenship. Thanks to Open Data, citizens can be more active in the public debate, due to the more information they can access. Their opinions can be also enhanced in terms of quality, because they are based on data and facts instead of prejudice. The consequence of such scenario can be very useful for democratic processes, because more informed citizens can make better decisions, and consequently improve the system which they live in.

After identifying the impact areas deduced by the 19 case studies, Young and Verhulst analyse how impact is created through Open Data, and suggest that eight premises mostly determine the nature and the relevance of the impact:

- Premise 1

Intermediaries and data collaboratives allow for enhanced matching of supply and demand of data.

- Premise 2

Developing Open Data as a public infrastructure enables a broader impact across issues and sectors.

-Premise 3

Clear policies regarding Open Data, including those promoting regular assessments of open-data projects, provide the necessary conditions for success.

- Premise 4

Open Data initiatives that have a clear target or problem definition have more effect.

- Premise 5

The lack of readiness or capacity at both the supply and demand side of Open Data hampers its impact.

- Premise 6

Open Data could be significantly more impactful if its release was complemented with a responsiveness to act upon insights generated.

- Premise 7

Open Data does pose a certain set of risks, notably to privacy and security; a greater, more nuanced understanding of these risks will be necessary to address and mitigate them.

- Premise 8

Even though open-data projects can often be launched cheaply, those projects that receive generous, sustained and committed funding have a better chance of success over the medium and long term.

In order to understand where these results and insights derive from, it is necessary to look at the case studies taken into consideration by the research. In this work, 4 of them are presented, one for each impact's dimension. The case studies are also selected from 4 national contexts that are very different with each other in terms of social contexts.

1) OpenAid

Location: Sweden
Area of impact: Improving government – tackling corruption and transparency
Web: https://openaid.se/sv/aid/2016/

OpenAid is an initiative launched by the Swedish government in 2010, concerning transparency in the area of development cooperation. The initiative aims to make public all the information regarding public funding to development cooperation activities, in order to activate virtuous mechanisms of transparency and accountability, and to increase involvement of multiple actors in the national development policy process. OpenAid took form as a reform agenda, which included different elements: one is the Aid Transparency Guarantee, which consists in the obligation for

public actors to make all public documentation and information available; this action resulted in the creation of a website called *openaid.se,* where all the data about public transactions concerning development cooperation is published (*openaid.se).*
The website focuses on highlighting which activities are financed, the amount of the transaction, who is the recipient and why, and also the result obtained thanks to the operation. It represents a clear example of an Open Data initiative, because data is mostly published in an open format, which allows reusability by third parties and the government itself. In this policy transparency and technological innovation meet and build a really useful platform for multiple actors. It is a useful initiative especially for generous countries like Sweden, because it is estimated that nearly 1% of the GNI (Gross National Income) is allocated towards development assistance.

An action as OpenAid is not unexpected in the Swedish context. Sweden policies are historically advanced from the point of view of transparency a publicness of government activity. In fact, it was the first country in the world to approve in 1776 a bill which states that all authorities should public documents, if a specific law that restrict the access doesn't exist. Such action cleared the road for a system oriented towards a more transparent, participative and collaborative way of doing public business. Transparency is seen in this view as a powerful tool to reduce corruption and improve the process of decision making.

OpenAid follows this approach that started more than two centuries ago, focusing on the particular theme of development cooperation. With this tool, stakeholders are allowed to be a part of the funding chain, from the decision phase to the monitoring phase. The expectation is that data published on the website is taken by third parties to build applications and instruments to enforce the policy, in order to make information more understandable for the general public. The goals of such action promoted by the Swedish government are different:

- Promoting active transparency
- Providing knowledge for better planning
- Increasing participation in development cooperation
- Strengthening precondition for accountability
- Fighting corruption
- Promoting efficient allocation of resources

The website is designed in such a way that it is easy to browse and data is found through filtering tools that allow to select which particular aspect has to be highlighted in the visualisation of the information: the activity, the money spent, the actors involved, etc. It is built as an open source Wordpress site. One fundamental element of this policy is standardisation: data is published in the IATI format (International Aid Transparency Initiative), which is a standard that enables high levels of comparability and reuse. Data is downloadable in CSV formats, which is the standard format for publishing Open Data. It comes from the contributions made by the Ministry of Foreign Affairs,

The Swedish International Development Cooperation Agency (Sida) and other authorities and ministries that handle aid funds. 15 updates of the platform have been made since it was launched. Thanks to the website, new reporting mechanism have been developed, which have enabled government to save around USD 7 million annually. Also the participation of citizens in the development of tools for the site has been successfully used.

OpenAid wasn't an isolated initiative in the open government field, because it stimulated a continuous political commitment by the central government to promote the same values that are part of this action; two examples of this commitment are the signing in 2011 of Sweden's Open Government Partnership (OGP) and of the Busan Partnership Agreement for Effective Development Cooperation (*oecd.org, 2016*).

2) NOAA

Location: USA
Area of impact: Creating opportunity – economic growth
Web: http://www.noaa.gov/

One of the most known and successful examples of a public agency that decides to open its data is the U.S. National Oceanic and Atmospheric Administration (NOAA). In 1970, the United States chose to unify all the activities related to environmental monitoring in a central agency that refers directly to the central government, i.e. NOAA, thanks to the effort of President Nixon. The agency's mission is to "*understand and predict changes in Earth's environment and conserve/manage coastal and marine resources to meet the nation's economic, social and environmental needs*" (OECD & NAS, 2008). NOAA's activity consists mainly in collecting data on climate and the environment to run forecasts and analysis for policy planning and information services. The choice of leaving the management of environmental data to a public body is sustained by four reasons:

1. It is non-exclusive, because anyone can observe and record the environment
2. It is non-rival, because it isn't less usable after being used
3. Collecting it is highly expensive because of the infrastructure that is needed
4. It can be reproduced at near zero-marginal cost

The demand for environmental data is very strong in the country, especially from the sectors that are interested in using such data in their activities due to the influence that the weather has on them. This high demand of data was one of the keys that made the agency understand the importance of information and also of the way through which the information is published, which influences various stakeholders if such information has some value. The first Open Data portal of NOAA was published in the early 2000s, and saw a continuous evolution due to the technological development which pushed it towards more user-centred ways to publish data. Technological development is also

a phaenomenon that has blurred the distinction between the roles of government, academic and private sectors in the environment industry, as the National Research Council stated (NOAA Magazine, 2004); thanks to technology, participation of different actors can be fostered, if the right tools are built. In 2004 NOAA decided to reform its information policy and commit to Open Data (NOAA Magazine, 2004):

- NOAA will adhere to applicable law regarding government information, based on the premise that government information is a valuable national resource and the benefits to society are maximized when such information is available to all.
- NOAA will carry out activities that contribute to its mission and provide open and unrestricted access to publicly funded information at the lowest possible cost.
- NOAA will provide information in forms accessible to the public as well as the underlying data in forms convenient to additional processing.
- NOAA will promote the open and unrestricted exchange of environmental information worldwide.

The archive of NOAA's data is today part of the National Center for Enviromental Information (NCEI), which "*is committed to full and Open Data access in support of its community of Data Producers and Data Consumers*." The effort that NOAA constantly makes is aimed to publish precise and relevant information for the users. Challenges of this Open Data policy are mainly technical: NOAA produces every year a large amount of data (30 petabytes), but only 10% of it is available on the platform; still, NOAA's web pages are the most visited web addresses of the U.S. federal government. Another interesting feature of the whole policy that NOAA has put in action consists in collaboration and participation: users can develop application and services based on the environmental data that the agency publishes, and NOAA can implement such tools in its service to improve its quality. For example, a partnership with Google was established to make 1 petabyte of cloud space available for house satellite observations, digital elevation data and climate/weather model data sets and 50 million CPU hours of high-performance cloud computing resources on the Google Earth Engine geospatial analysis platform; this partnership is aimed to make environmental data more accessible to the public.

One of the most significant aspects of the NOAA initiative is that the agency has tried to assess the impact, and especially the economic benefit, of making its data available over the years. Measuring the impact of this type of policies is not an easy task; the methodology chosen is the Value Of Information (VOI), which assumes that more information improve the decision making process and that decisions have a tangible economic impact. These are some of the estimates that the agency developed (NOAA 2011, 2013):

- The United States' USD 8-10 billion and growing annual Weather Derivatives financial industry relies on NOAA's seasonal weather data and records.

- NOAA's forecasts and warnings and associated emergency responses result in USD 3 billion in a typical hurricane season.
- United States electricity generators save USD 166 million annually using 24-hour temperature forecasts.
- NOAA real-time data supplies a burgeoning private weather service industry with well over USD 700 million in value added annually.
- The farming industry has always been among the most important consumers of NOAA's products. By providing more accurate climate forecasts and more timely warnings of adverse weather, NOAA has helped the industry improve decision-making and crop yields. According to one estimate, data provided by the NWS Climate Prediction Center, a component of NOAA, has provided U.S. agriculture of over USD 460 million by helping guide planting decisions in El Niño, normal and La Niña years.
- According to a nationwide survey, 96 percent of the U.S. public obtains, either actively or passively, 301 billion forecasts each year. Based on an average annual household value of USD 286 placed on weather information, the American public collectively receives USD 31.5 billion in benefits from forecasts each year. These benefits far exceed the USD 5.1 billion spent annually by both private and public weather bureaus on generating forecasts.
- Some prominent examples of companies built around weather data include The Weather Channel, which uses NOAA data to reach some 97.3 million American households, and the Climate Corporation, which used weather data to provide "weather insurance" to businesses and was sold in 2013 to Monsanto for USD 930 million.
- Using the self-assessment methodology mentioned above to estimate VOI, a survey of households in states prone to hurricane damage was carried out to learn how much taxpayers would be willing to pay for enhanced hurricane forecasts. Researchers found that, on average, households in at-risk states were willing to pay an additional USD 14.34 per year.

Because of the strong commitment to Open Data that NOAA made during the years, it became one of the most successful examples of Open Government Data policies, so that the former US President Barack Obama mentioned it as a best practice when the national platform data.gov was launched.

3) "*Mejora tu escuela*"

Location: Mexico
Area of impact: Empowering citizens – informed decision making
Web: http://www.mejoratuescuela.org/

"*Mejora tu escuela*" is an Open Data initiative developed in Mexico which aims to deliver citizens information and data concerning the school system, in order to make them more informed and take better decisions as a consequence. The Mexican education system suffers from various issues; for

example, even though a relevant amount of the national budget is destined to sustaining schools, the performance of the students is very low: there is no OECD country that spends more on education, but fewer than half of students earn a high school diploma (Villagran, 2013). Furthermore, teachers and students perform bad in the international comparison tests. Another issue, probably connected to the afore mentioned, is corruption: it is always hard to estimate the magnitude of such phenomenon, but the insights available give a clear indication that the problem exists and persists in the Mexican contest. A billboard erected in the capital by activists in 2014, showed that more than USD 33 million were lost only in the first week of the school year (Cave, 2014). The 2009 report of Transparency International found that families pay an average additional fee of USD 30 for their children education, which should not happen given the fact that education is "constitutionally free" (Transparency International, 2009). Corruption in the education sector consists mainly in ghost teachers on payrolls, preselling of exam results, officials with lavish lifestyles, teachers demanding bribes to give good grades and candidate teachers bribing to pass training tests.

The first effort to make the education system more transparent was made in 2008. A law that required all the states to provide government the information about the conditions of schools, payrolls and other expenditures passed at the Mexican Congress; the response to such effort was not so encouraging, as many states handed over incomplete information (Zabludovsky, 2014). Even though also a general commitment to Open Data has been made by the central government, and a national Open Data portal has been built, the plans of such policy are poorly executed (Montiel, 2015).

"*Mejora tu escuela*" was launched in 2013 as a public, independent, non-profit initiative. The online platform, which is the main part of the program, has the goal of "*promote citizens participation to improve education in Mexico*", based on the assumption that "*education [in Mexico] will improve only with the active commitment of all members of the educational community, including parents*" (*mejoratuescuela.org).* The members of the team that developed the initiative are part of the Mexican Institute of Competitiveness (IMCO), with the support of Omidyar Network. The most relevant issue which the program focuses on is the engagement of parents with the education system; such phenomenon was observed through a survey in 2013 that concluded that 78% of Mexican parents were satisfied or very satisfied with their children education, even though the country performs badly in the international comparison. Lack of parents' engagement in the country is indicated as a cause for the influence of special interests in the delivery of education. So the evident problem in this case was that a relevant information gap of the families made them believe that the education level of their children was sufficient; the program tries to close this gap through a better delivery of information, which could make parents realize the real situation of schools and teachers. The platform consists in 4 features:

- Get to Know Your School: the central information provision aspect of the project focuses on giving parents and other stakeholders access to a diversity of information on schools.

- Compare Your School: it allows parents to compare their child's school to others in the same area, complete with an assigned school grade. The hope is that this piece of the platform will "get competition going between public schools" (Zapata Hojel, 2015).
- Grade Your School: asks parents and other members of the school community to rate different aspects of a school on a scale of one-to-ten, and leave comments. This information is then fed into the Get to Know Your School stream of information.
- Improve Your School: in its current form, Improve Your School provides "concrete tools and suggestions to parents on how to participate in addressing problems in their child's school" (Hasan, 2013). In the future, this section will also be able to suggest concrete actions to parents or teachers based on the hyperlocal needs identified (through crowdsourced reporting and/or existing data) for a given school (Zapata Hojel, 2015).

The best example of Open Data is the 2013 census database, which delivers information about the school infrastructure, the number of students in schools, the number of teachers, the names of principals, schools location and contact details.

The impact of the initiative can be considered for different aspects and beneficiaries. A broad view on the impact of the program can be made thanks to the traffic data: between 40,000 and 45,000 visits are registered every day; it's worth noticing that this result has been obtained without a relevant marketing action. Another area of impact regards the parent-teacher relation, which suffers from a substantial lack of contact due to the fear that parent feel in opposing a teacher. From the beginning of the program, stories about parents that were able to fight teachers' absenteeism, thanks to the platform data, have been collected. Therefore, Open Data represents a tool that gives new possibilities to parents and can improve their contact with the teachers and principals. Also schools officials and policy makers have gained a benefit from the data, and they have expressed satisfaction on the reliability of it. Having an impact not only on parents but also on officials is desirable, because in this way also education policies can improve thanks to the more information available. Teachers and principals too are stimulated towards better performance of their teaching and their schools. Also it has been reported that some of the schools' principals have found out to be enrolled in federal funding programs thanks only to the data published on the web platform. The last area of impact which was observed is corruption; in 2014 IMCO published a report, backed by the National Audit Agency, using the data available on the platform. Some of the findings are the following:

- 1,442 teachers on government payroll were between the ages of 100 and 105. Of these, 1,441 were registered in the state of Hidalgo, and all but one were born on December 12, 1912 (i.e.12/12/12).
- 70 teachers earned more than President Enrique Peña Nieto (who made around $15,000 per month). Of these, 19 worked at schools that had failed or barely passed the national standardized test (Zabludovsky, 2014).

- The report also found that the average monthly national salary for Mexican teachers was $1,954, and that the teacher with the highest salary lived in Oaxaca state and earned $46,849 a month. In the state of Guerrero, there was a school with a single student, but six employees whose payroll added up to $6,644 per month (Zabludovsky, 2014).
- The report uncovered 536 *telesecundarias*, or secondary and high school programs available in rural areas via satellite, that apparently operated without electricity. The report also showed that there existed a large number of "phantom schools" that were unregistered (and thus possibly non-existent), yet that had several teachers on payroll (Zabludovsky, 2014).

Despite the success of the program, various challenges remain to be dealt with. Transparency doesn't automatically results in substantial impact, which can be achieved only through a constant and collective effort for improving the education system. One of the obstacles that have been observed is represented by the so-called "special interest" that don't want to be put in discussion by such push towards transparency and accountability. For example, the national test for measuring students' performance was cancelled by the federal government after the first year of the program, under the pressure of teacher's union and other organisations. Also data quality and accessibility continue to be relevant issues.

4) Fighting Ebola

Location: Sierra Leone
Area of impact: Solving public problems – data driven engagement
Web: NERC (http://nerc.sl/); HDE (https://data.humdata.org/); Ebola Geonode (http://ebolageonode.org/)

In 2014, one of the main topics covered by international media was the Ebola outbreak in Sierra Leone and other West African countries. The fear of contamination spread also in Europe and USA, due to the magnitude of the epidemic: it was the largest outbreak of Ebola in history. Not only its magnitude, but also the difficulty to access to complete and reliable information was a factor that spread the fear among people, even in countries where the danger wasn't tangible. Misinformation and disinformation were issues for the operators who were dealing directly with the outbreak, causing major problems in managing the situation in an effective and efficient way, in order to contain the epidemic. Because of such issue, different actors developed initiatives aimed to collect relevant data, aggregate it in useful forms and platforms, and deliver it to subjects who could have had a significant impact on the resolution of the problem. Open Data became a crucial aspect in dealing with such emergency and represented an important resource for different agents operating on the field. The three initiatives here discussed are: Sierra Leone's National Ebola Response Centre (NERC), the United Nation's Humanitarian Data Exchange (HDX), and the Ebola GeoNode.

Sierra Leone is not one of the countries that have implemented relevant Open Data policies, as the Open Data Barometer put it at the 78th place of the international ranking (*opendatabarometer.org, 2015*). Nevertheless, in 2013 a Right to Access Information Law passed, which made information detained by public bodies easier to access. In the same year, the country joined the Open Government Partnership, a first step towards a national Open Data strategy. At this moment, one of the most important efforts in the Open Data field is an initiative called Open Data for Sierra Leone, developed by the African Development Bank Group's Open Data for Africa project aimed to publish information about economics, demographics, agriculture, energy, education, healthcare, food security and international trade (*sierraleone.opendataforafrica.org*). Furthermore, the Ebola emergency has significantly increased the national interest on Open Data, because of the impact that some initiatives have had in the management of the crisis.

According to the US Centers for Disease Control and Prevention (CDC), the 2014 Ebola emergency was the "*first Ebola epidemic the world has ever known*" (*cdc.gov*). The disease spread also in other nearby countries, like Guinea and Liberia; the reported cases of Ebola from March 2014 to September 2015 were 28,355. The UN Security Council declared that Ebola was a "*threat to international peace and security*" (United Nations, 2014) and some experts made the prediction that the disease could kill 100,000 people and would require $1 billion for a successful containment (Spooner, 2014). The country was able to contain and put an end to the outbreak thanks to the intervention of international organisations and to the collaboration, coordination, data sharing that occurred between the different actors involved in the emergency. One of the issues that emerged in the management of the crisis was the limited information sharing between national governments, organizations and operators on the field; simple data like the number of cases and dead were hard to obtain. The initiatives here described are aimed to answer the demand of such information and coordinate the management of the information held by different sources.

The National Ebola Response Centre (NERC) was created in October 2014 by Sierra Leone to deal with all the aspects of the on-the-ground Ebola efforts of the country, and to coordinate the activities of District Ebola Response Centres (DERCs) (Maxmen, 2015). NERC involved a significant number of important subjects: the Presidency of the Republic of Sierra Leone with the ministries of Health and Sanitation; Local Government and Rural Development; Social Welfare, Gender and Children's Affairs; Foreign Affairs; Finance and Economic Development; Defence; Internal Affairs; and Information and Communication. Other external subjects participated in the initiatives like the CDC, Red Cross, World Bank, UK Department for International Development (DFID), African Development Bank, and the U.S. Embassy. The goal of the program was to reduce the effects of the disease and control the outbreak. Along with the Situation Room, an initiative that helped to deliver relevant information to decision makers who could act based on it, NERC was an effort to create an infrastructure aimed to collect and disseminate information about the crisis. One of the main tasks of NERC was to collect daily reports of the 14 districts in Sierra Leone, trying to collect information such as the number of safe beds detained, the number of the dead and the number of the daily calls that they received to investigate a case. One part of the activity consists in

briefings, where subjects from different agencies and institutions were called to identify the action key points based on the information collected to that point. The program focused basically on communication: briefings with the public, newsletters and press conferences were held to deliver data and information about the emergency.

The development of the initiative called Humanitarian Data Exchange (HDE) was announced on July 15, 2014 by the United Nations Office for the Coordination of Humanitarian Affairs (OCHA). HDE was conceived as a data sharing platform to be used during emergencies of different contexts; the platform has been referred as a humanitarian data-centric Wikipedia. Through HDE users are allowed to track and follow specific datasets, create curated organizations data hubs, and share data across previously siloed organizations working to improve humanitarian efforts around the world (Griliopulos, 2014). The platform became an extremely useful tool in the Ebola crisis, because of the information issue previously discussed. At the end of 2014 it passed from being a simple information source to being an instrument with a tangible impact, when the UN appointed it to be the platform for Ebola data exchange. At the time the data published was mainly on population statistics and mortality rate of the regions struck by the emergency. An important collaboration with the World Health Organisation led to a substantial increase of the datasets, which became more specific about the Ebola issue: the number of Ebola cases and fatalities, the locations of cases, the amount of money being spent on the crisis, as well as information on Ebola Treatment Centres (ETCs —e.g., how many were open at the time, how many were eventually planned to be opened, and their locations) were put in the platform (Teran, 2015). All this information was published in an open and computable format for being easily used by analysts and developers. One of the most difficult tasks was to make all the data coming from different sources standardized and comparable. Dealing with other countries made such task even more difficult to perform. A resource that helped to improve the platform was the creation of tools and datasets by the users from the data published; such resources where implemented in the platform and made it more useful for various types of actors: citizens, intermediaries and policy makers where so able to use it for their scopes.

A useful visualisation tool was also build to fight the emergency that Sierra Leone was living. The Ebola Geonode was born thanks to the collaboration of Red Cross, World Bank, Global Facility for Disaster Reduction and Recovery (GFDRR), UNMEER and the US Humanitarian Information Unit (HIU). The Ebola Geonode is an open source geospatial platform that lets users build maps and conduct geospatial analysis on Ebola's impacts in West Africa (Dufour, 2015). The data on the platform concern aspects like administrative boundaries in affected countries, transportation and logistics information, and geo-tagged health crisis information; one of the most useful dataset contains the location of ETCs health facilities. The platform consists in three types of features: Layers and Maps allow users to manipulate directly the data on the website and to download it; documents includes mapped trend analyses and can be downloaded in different formats. The main goal of the Ebola Geonode was to alleviate information fragmentation that hampered the fight against Ebola because of the time that had to be spent in order to find and reassemble the datasets.

Like the other initiatives mentioned, the platform allowed collaboration between different institutions, organisations and countries.

As now should be clear, data has a variety of impacts that most of the times aren't even all identifiable before publishing it, because different types of subjects can be interested to it for gaining different benefits. One of these benefits is the improvement of decision-making processes; in the case of Ebola, data gave crucial information to policy makers who were able to respond to the crisis with more effective actions. Another example regards HDX, which was used by the UN World Food Programme (WFP) to understand how the emergency was affecting farmers and to forestall potential food shortages. The impact of maps was also important because the area of the outbreak is not well covered by tradition mapping tools like Google Maps: roads and villages are sometimes missing, so the Ebola Geonode was particularly useful to fill such gaps with more accurate information and geospatial data (Cassano, 2014). Furthermore, as more subjects become aware of the potential of data, and especially that published in open format, the Open Data commitment can spread throughout other organisations; for example, WFP has decided to publish information about food prices around the world after the Ebola crisis through which it became aware of the potential of data. This emergency's response was also a best practice that was taken as example for other emergencies in the following years, like the Nepal Earthquake of 2014. Some of the challenges of this experience were already discussed; here they are sum up:

- Data management: it was difficult to collect high quality data, because of the errors that were constantly encountered, like out-of-date data, missing data, spelling mistakes, different times at which the data arrived
- Collaborating with government: making humanitarian workers collaborate with governments wasn't easy, because they the formers had to deal with politics and spend a lot of time to discuss during and ongoing emergency; bureaucracy is one of the obstacles that can be in the way of a fast action
- Lack of existing structures and best practices: all the initiative was innovative and one of a kind, because of the magnitude of the emergency; relations, structures and platforms were established after the outbreak started, so that they were created in a rush which can reduce their quality

The four case studies presented confirm that various types of impacts and benefits can be achieved through Open Data and its different applications. While OpenAid is a more traditional action for monitoring public funds, the other three are quite innovative. The case of NOAA shows how much meteorological and environment information represent a resource for the economy, but also the societal benefits of having a more efficient forecasting system. "*Mejora tu escuela"* is a very interesting initiative because it addresses the issue of misinformation within families on schools' conditions and performance, and it allows decision on children's education to be more informed; citizens become interested and engaged when the initiative responds to a clear demand of the

population. Finally, the actions that were undertaken in Sierra Leone during the Ebola's emergency summarize the power of an efficient management and diffusion of information, which in this case contributed to save lives and to solve a humanitarian crisis.

4.5. Open Government Data challenges

The previous sections highlighted what value can be generated through the implementation of Open Government Data policies, both from a theoretical perspective and from the empirical evidence showed in some international case studies. Nonetheless, to achieve such value some challenges of different nature are to be overcome. Ubaldi (2013) summarizes them in 6 categories:

1. **Policy challenges**: disclosure policies can be in the way of a real Open Government Data policy, due for example to a lack of clarity in information like the data owner, which can restrict its use by third parties. Two principles that may have a conflict in this sense are the right to access information and the property right, i.e. copyright; such conflict could represent an issue for reusing data because the data owner can legitimately have some reasons to oppose the freedom of other subjects to use his data. Another issue, observed in particular in the national Open Data policies, is the lack of clear long-term strategies in this field that are necessary to ensure that opening data isn't only a mean towards certain goals, but a constant source for value to be created by third parties.

2. **Technical challenges**: because of the complexity of public agencies that many countries have, a major issue is represented by standardisation. Data is often published in different formats and standards, which makes its reuse more difficult. Other technical issues are represented by information technology infrastructure, privacy and information security, and OGD tools and applications. Sometimes the central government embrace the Open Data challenge through the creation of a national portal; although such choice satisfies the need for standardisation, it should be developed through a collaborative approach for creating ownership and sustainability. A particular aspect that poses a relevant issue for value creation out of Open Data is the so-called digital divide. The digital technologies are certainly growing among world's population, but contexts where the ICT sector has not effectively expanded remain. If access to the Internet is not fully guaranteed, less and slower benefits of Open Data are seen, which makes an Open Data policy less effective.

3. **Economic and financial challenges**: in the Open Data debate it is sometimes argued that opening data has almost zero cost for the public subjects who publish it; this assumption doesn't really find confirmation in reality. There are various costs and potential costs associated with opening data: production and presentation of Open Data, the acquirement of skills, employees, technology and the network infrastructure by public administration,

human-resource costs for organising and preparing information. Also converting data to reusable formats, liked linked Open Data and semantic web formats, can be costly especially if there is a high degree use of proprietary formats. These costs increase more if Big Datasets are to be published, because of the work of controlling and cleaning required. Creating the infrastructure for opening data represents also a cost, but this task can be left to third parties, focusing on creating a basic and accessible publishing tool for data which can be used by other subjects to develop more advanced websites and platforms.

4. **Organisational challenges**: organisational structures are needed to establish effective Open Data practices in the government, because of the complexity and cross-cutting nature of OGD. There are different action that can be undertaken to build such structures; for example the task of coordinate and provide leadership to Open Data initiatives can be given to a government body, which is able to bring different stakeholders on board. Also having a ministry who urges the release of data by agencies is a useful way to control the process effectively. Independent bodies have to be engaged to demand and publish data, in order to ensure transparency. In general a broad ecosystem of actors is needed for the Open Data strategy to be fully implemented, especially to create value across society. For becoming a common practice in the public sector, the Open Data commitment should be continuous in the governmental agencies, and not limiting to a short-term action.

5. **Cultural challenges**: a strong cultural change among all the stakeholders involved in an Open Data policy is the essential condition for obtaining the outcomes that such policy could cause. This cultural change should firstly consist in the citizens realizing and understanding their rights to information, so that they are not intimidated in demanding such information to public bodies. One of the main obstacles in publishing data could be the reluctance of public officials to disclosure of data, seen as a threat or a way to be disempowered as the activity of the public body is more vulnerable to monitoring and evaluation. This perception should be changed towards a proactive and smart disclosure of information, focusing on the possible benefits that the public sector itself could gain from this action. The same public officials are most likely to gain benefits if they understand how more data available can influence their decisions, improved in quality by the information obtained. The cultural change is also fundamental for the part concerning technology: it is always the case that some subjects, both among data supplier bodies and users, are not used to work with data, especially with through the raw formats in which are published from Open Data platforms. This issue has to be dealt with in order to build the right ecosystem and capacity to implement data in an increased number of aspects throughout society.

6. **Legal challenges**: one aspect that is problematic form the public bodies' perspective is the multitude and variety of rules that concern public information and its publication. A common issue is the lack of a consistent framework that allow Open Data to be delivered to third parties without wasting much time and breaking rules. Even in countries where

freedoms of information's laws exist, it is sometimes not clear if databases are part of such laws, resulting in an ambiguity that doesn't facilitate the publication. Fragmentation of the legal systems get in the way of giving a clear indication on how manage the disclosure of data held by public bodies. Obviously there should be some limitations to opening, basically due to privacy and security; for example data that concern the public defence should be treated in a certain way in order to avoid security gaps in the system. Another issue that is observed in some contexts concern ownership; even if public bodies produce and collect data for the public interest, there are cases where public bodies assert intellectual properties rights over such information and aim to sell it to gain an economic benefit from it. Within some countries, e.g. USA, different regulatory regimes exist, which combined create issues for public officials who wants to publish data, because they cannot do it without breaking some rules. In general, a constant updating of policies is necessary to meet public bodies' and citizens' needs.

CHAPTER 5

Open Data in the Italian context

As explained in chapter 4, currently different countries around the world are embracing today the Open Data challenge; policies and actions at various levels of government are been taken, testing different approaches to put to best use the information detained by the public sector. Also the private sector is playing an important role in such challenge, trying to generate value from data in ways that constantly change thanks to technological development.

This chapter analyses the Italian context: what is the Open Data strategy of the country, how data is being used, which areas have developed Open Data policies the most and how the economic and social context regarding technology is characterized are some of the questions put forward to deal with this analysis. This work tries to map the national situation regarding data usage of the private sector, focusing especially on some case studies of Italian companies that have used or are using Open Data to develop innovative business models. There could be a lot of insights concerning regional Open Government Data policies to put forward, but here the analysis will be mostly on the general Open Data commitment of the country. Because of such focus on companies, the value that is looked for is the economic one. In the first part of the chapter the national characteristics of the whole ICT sector are underlined, in order to contextualize the analysis.

5.1. ICT Italian sector

In the first chapter the development of ICT sector and the impact of Internet are discussed to show the importance of the new digital markets in the economic and social scenario. It is interesting now to see if Italy is a representative example of such change, comparing it also to other countries. The focus on ICT sector is brought because it is important that the conditions for value generation exist, for Open Data to be exploited in its full potential; these conditions exist when technology, competences and culture of using data have entered within the society and are spread throughout the population. The aspect here observed is the ICT technology, its availability and use in the Italian context.

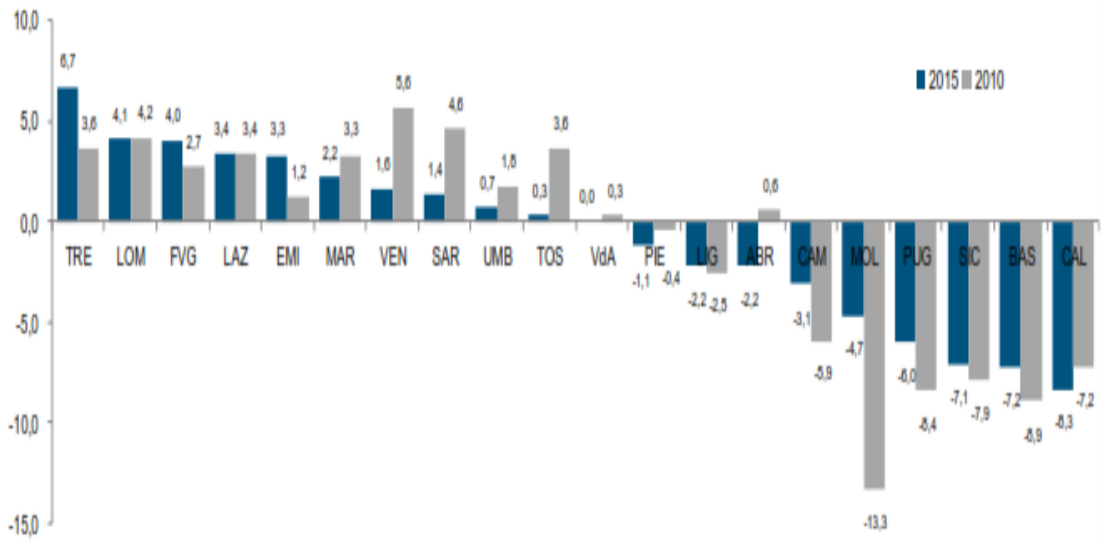
The national institute of statistics, i.e. Istat, has drawn a report named "*Cittadini, imprese e ICT*" in 2015 concerning the spread of ICT within citizens and companies (Istat, 2015); this report is based on a survey and it takes data from a sample of citizens and companies to understand how much Internet technology has penetrated the country from 2010 to 2015.

5.1.1. Families and ICT

It is observed that in 2015 the number of families that have been connected to the Internet with a broadband connection has grown from the previous year. Mobile broadband connections have also increased both for families (27.6% to 30.1%) and companies (49.8% in 2013 to 63.3% in 2015). Data shows that Internet users are very young, and from the age of six there is a 60% of probability that they have accessed to the Internet at least one time in a year. Compared to 2010, in 2015 more families possess a connection to the Web in their homes (52.4% to 66.2%); also a growth of 23.4 % is seen in the adoption of broadband connection, from 41.0% to 64.4%, due to the widespread of mobile broadband lines. Compared to the other European countries, Italy occupies the last positions regarding broadband penetration (74%), but the mean annual growth rate of this index has risen with values above the EU28 mean.

Regional gaps among families that can use broadband connections remain relevant issues in the peninsula, with this gap constantly in favour of the region in the centre and in the north of the country. Southern regions are all under the Italian mean in this analysis; this aspect represent a tangible obstacle for the development of such areas, and has to be addressed as a priority in both local and national policy.

Fig. 5.1 Families with access to broadband, regional comparison (2010-2015)

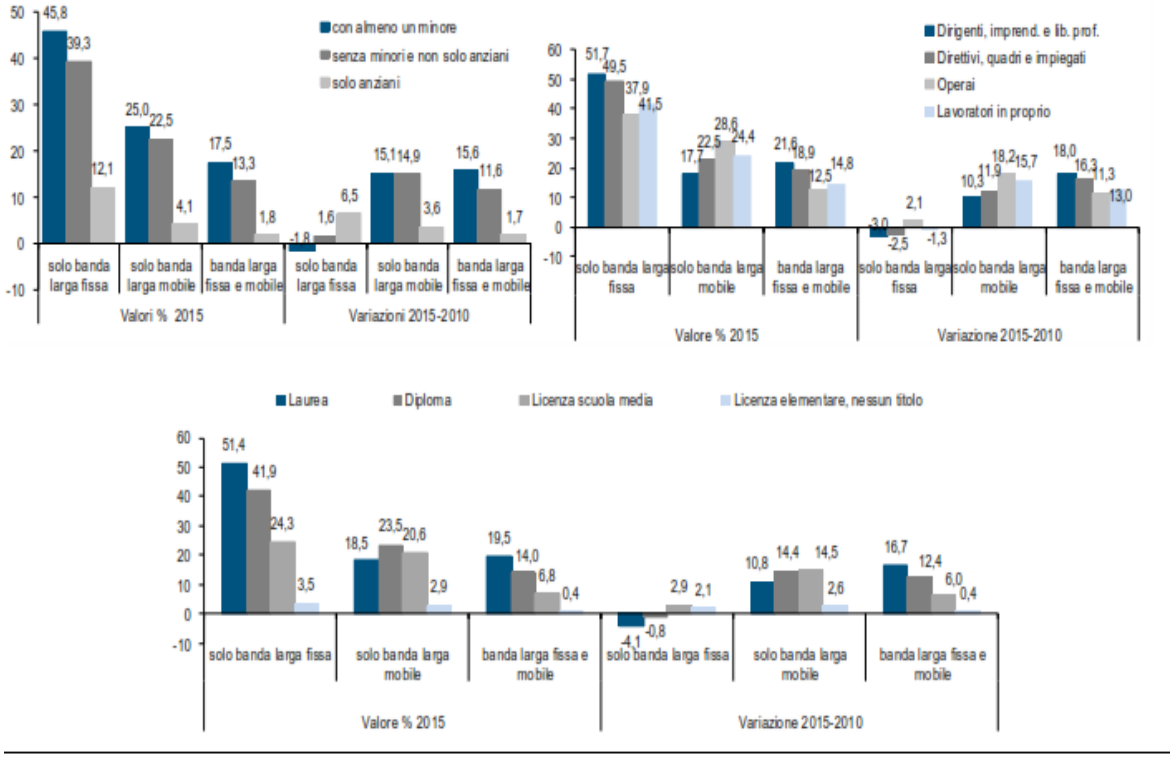


Source: Istat (2015)

Also among families relevant differences are present concerning broadband use; the three main factors observed as linked to such differences are the age of the family, especially the presence of a minor, the education level and the profession of the head of household. Result are presented in Fig.

5.2, showing a clear advantage in broadband use for families with a minor, high-level education families and families in which the head of the household is a manager, an entrepreneur, a sole practitioner, an executive, a board member and an employee instead of a manual worker or a private worker. Families composed only of people over 65 years old suffer the most from the technological gap. The main motivation of such gap is indicated by families in the lack of competences for using the Internet (56.3%), followed by the lack of interest for the Web (24.5%) and the cost of instrument necessary to access (14.4%). The aspects mentioned in the interviews are very interesting: the fact that more than half of Italian families don't have the competences required to access to the Internet represents an issue, caused obviously by the age. Elderly have difficulties in using new technologies because they require learning process that are easier for younger people; due to the aging of Italian society, it is possible that such gaps could remain in the years to come, but they will be eventually shortened by the growth of the so-called native digitals that will represent an ever increasing quote of the population.

Fig. 5.2 Families with broadband connection, analysed for education level, age and head of the household job (2010-2015)



Source: Istat (2015)

The same differences concerning broadband connection observed in Fig. 5.2 are seen in the data concerning Internet use. Age, job and territory are factors that are linked to the number of people

that have used Internet during the year (Fig. 5.3). From 2010 there has been a substantial growth of Internet use for all the age classes part of the analysis, but with a relevant growth for people between 55 and 59 years old (41 %to 64.4%), especially for women. There is also a gender gap which in 2015 is quantified in almost 10 percentage point: 55.8% of Italian women use Internet, against a 65.0% of men. Levels of Internet use are very low for housewives and retired from work, compared to the other categories. Southern regions and the islands suffer again in this statistic, with a major concentration of Internet use within northern and central regions. Categories that use Internet the most are students, managers, entrepreneurs, sole practitioners and office workers, which are all above 90%.

Fig.5.3 Internet use for age, geographic area and job (2010-2015)

SESSO, CLASSI DI ETÀ, RIPARTIZIONI GEOGRAFICHE E CONDIZIONE OCCUPAZIONALE	Uso di Internet (a)					
	2010	2011	2012	2013	2014	2015
Maschi	54.6	56.7	58.3	60.3	62.4	65.0
Femmine	43.6	46.7	47.1	49.8	52.8	55.8
Totale	**48.9**	**51.5**	**52.5**	**54.9**	**57.5**	**60.2**
6-10	36.7	38.3	40.8	45.1	44.6	43.8
11-14	75.7	78.1	76.5	80.8	80.9	80.4
15-17	87.2	89.1	88.5	89.7	91.2	92.0
18-19	90.4	88.8	88.8	90.0	93.9	92.0
20-24	82.1	85.8	86.0	85.7	89.3	90.7
25-34	73.3	77.5	79.2	80.3	83.9	85.1
35-44	64.6	69.7	69.1	73.5	76.1	80.1
45-54	53.0	56.2	58.7	61.7	66.1	70.0
55-59	41.0	42.2	45.3	48.5	52.0	60.4
60-64	25.2	28.6	31.0	36.0	41.1	45.9
65-74	12.1	13.8	16.4	19.0	21.2	25.6
75 e più	2.0	2.7	3.3	3.5	4.4	6.7
Totale	**48.9**	**51.5**	**52.5**	**54.9**	**57.5**	**60.2**
Nord-ovest	53.6	56.4	57.3	58.3	61.5	64.6
Nord-est	51.3	55.9	57.7	60.1	61.5	65.2
Centro	51.3	54.5	55.1	57.8	60.4	61.6
Sud	41.9	43.6	43.3	46.6	49.2	53.1
Isole	44.5	43.9	47.3	49.8	53.0	53.8
Italia	**48.9**	**51.5**	**52.5**	**54.9**	**57.5**	**60.2**
Occupati	68.7	71.9	73.2	75.9	79.1	81.1
Dirigenti, Imprenditori, Liberi professionisti	85.0	84.9	86.2	87.1	88.6	91.0
Direttivi, Quadri, Impiegati	85.1	88.3	88.9	89.9	91.5	91.8
Operai, Apprendisti	48.4	53.7	56.5	59.0	66.2	69.4
Lavoratori in proprio e Coadiuvanti	56.8	60.5	61.8	67.0	69.0	71.9
In cerca di nuova occupazione	54.8	59.0	56.4	61.1	65.1	68.5
In cerca di prima occupazione	59.7	69.1	66.6	68.1	75.7	77.9
Casalinghe	17.1	19.4	19.3	21.5	24.3	29.8
Studenti	91.8	92.4	93.3	92.2	93.4	94.1
Ritirati dal lavoro	13.3	14.7	16.2	18.2	19.1	23.2
Altra condizione	22.6	23.0	24.5	24.9	25.5	31.0
Totale	**48.4**	**51.1**	**52.1**	**54.3**	**57.1**	**60.3**

(a) Negli ultimi 12 mesi.

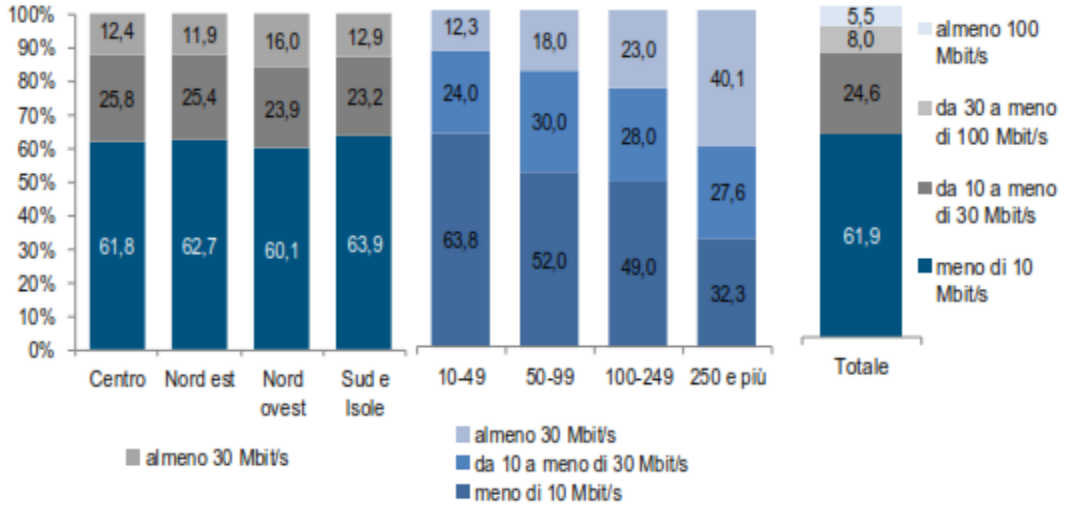
Source: Istat (2015)

5.1.2. Private sector and ICT: Italian companies in the digitalization era

One of the fundamental aspects to observe in order to understand the Italian ICT context is the penetration of ICT technologies in the private sector. Dealing with an ecosystem of companies that haven't still embraced the digitalization challenge could lead to relevant issues in the development of a competitive market, because of the trend showed in the first two chapters. Therefore, it is important to assess how much Italian companies use Internet technology, especially compared to the rest of European countries.

Istat takes part in this research, and shows interesting data concerning the degree of Internet usage and the possession of broadband connection by companies. The first fact worth mentioning is that in 2015 94.4% of companies with more than 10 workers dispose of broadband connections; although, the most part (61.9%) still has a low speed connection (under 10 Mbit/s). The speed of the broadband is highly link with the size of the enterprise, but not with the geographic area. The cluster of smallest companies (10 to 49 workers) shows high-speed connections only in the 12.3% of the cases, while 40.1% of the biggest companies (+250) have a 30 Mbit/s or more connection (Fig.5.4.)

Fig. 5.4.Companies with a fixed broadband connection, analysed for maximum speed of connection, geographic area and number of workers



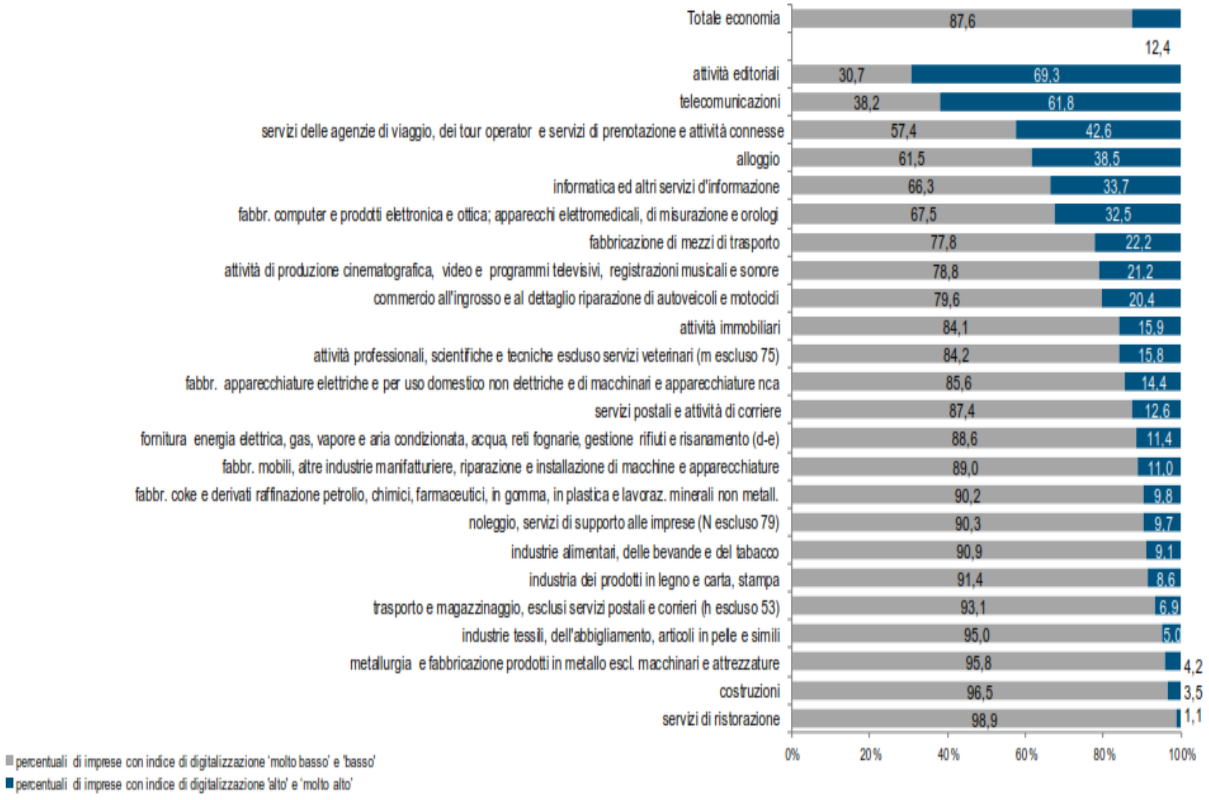
Source: Istat (2015)

Istat also build a composite indicator of digitalisation for companies, considering 12 specific activities. The results show that 87.6% of Italian enterprises have a "low" or "very low" level of digitalisation, because they don't incorporate more than 6 of the activities considered by the Istat indicator. There are huge differences among the various business sectors in this aspect, with the

editorial sector that leads with a 69.3% of activities that are in the "high" or "very high" levels of digitalisation (Fig.5.5). The explanation that Istat gives for the high majority of companies that don't embrace new technology is focused on Italian companies' size: due to the prevalence of small enterprises, most of them are provided of ICT services by external subjects; such services are not counted in the indicator, nonetheless they provide useful elements to be more competitive on the market. Therefore, the main factor that is connected with the level of digitalisation is companies' size. The territorial gap is present, but it is shortened, compared to the gaps of Internet use in the general population; the company's size is instead a relevant characteristic to identify companies with high or low degrees of digitalisation (fig. 5.6).

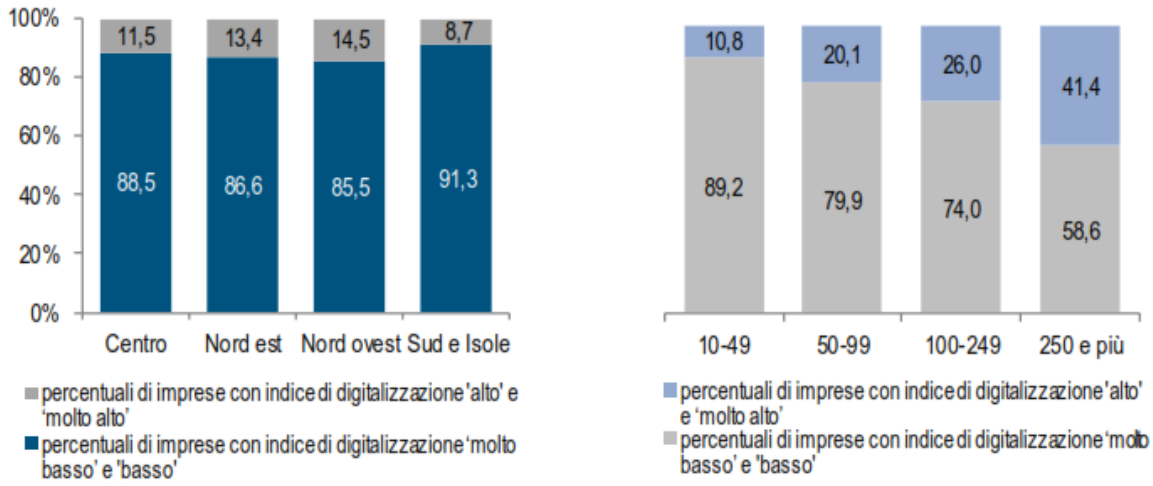
It is interesting also to see how much the online buying/selling channels are used by the private sector. Among the companies with more than 10 workers, in 2015 the 10% of them have sold their products online during the last year.

Fig.5.5 Digitalisation indicator for companies with more than 10 workers, analysed for business activity



Source: Istat (2015)

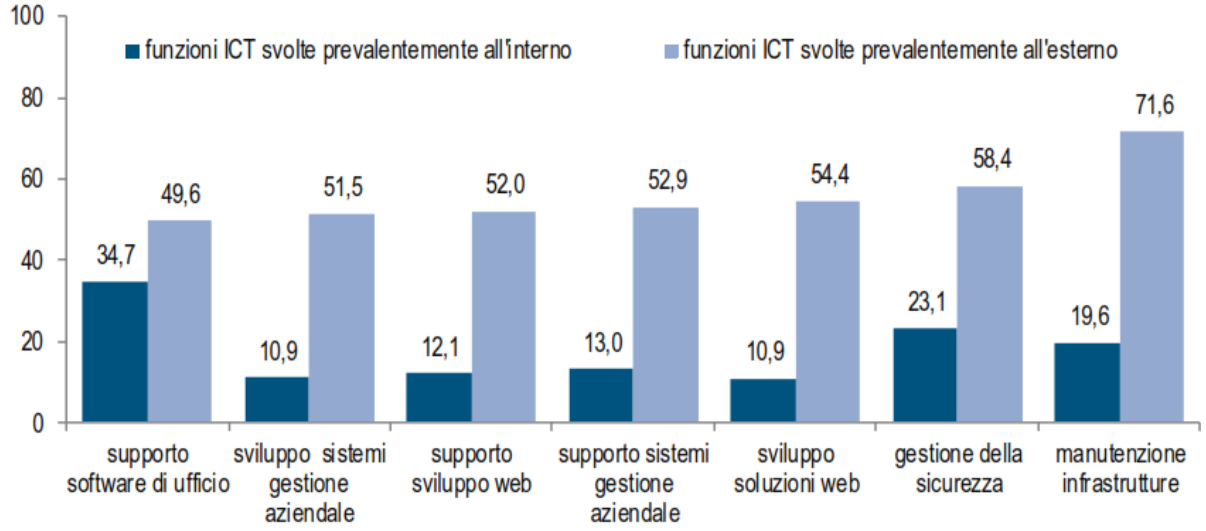
Fig.5.6 Digitalisation indicator for companies with more than 10 workers, analysed for geographic area and number of workers



Source: Istat (2015)

The general trend is an increase of companies that use such online channels (+1.8% from 2014); in the online selling business there is a predominance of companies with more than 250 workers (29.6%). Online revenues also represent a more relevant part of the total revenue for bigger companies (18%) than for small ones (2.6%); the most active sectors on the online market are information and accommodation services, the less active is the construction sector. Size is also linked with the presence of employees who are ICT specialists, and altogether 75% of them have at least one ICT specialist. Such technological competences can be used from the internal workforce, but can be found searched in external subject; the latter is the most common case for the Italian private sector: all the ICT functions that Istat identified are mostly managed through outsourcing systems by companies (Fig. 5.8).

Fig.5.8. Companies that perform ICT functions, analysed for function and majority of internal or external functions performed



Source: Istat (2015)

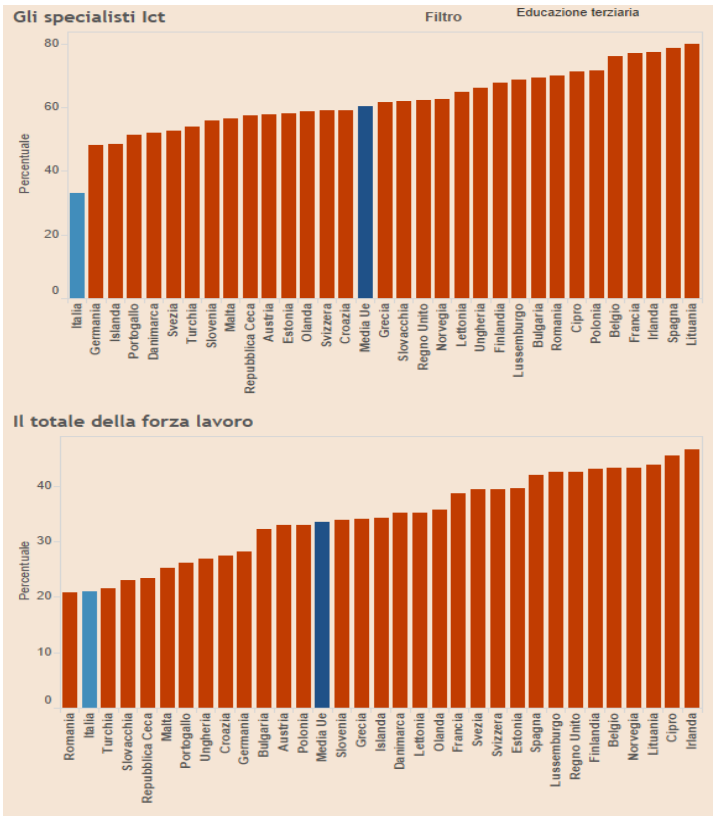
We have now seen that the main factor that is connected with ICT pervasiveness among Italian companies is the number of workers within enterprise; companies situated in the peninsula also present a low degree of digitalisation. These 2 aspects are connected to each other, because the Italian market is composed mostly of small-medium companies, therefore it is natural that the average degree of digitalisation is low, due to the first fact that we highlighted.

It is important to compare data between Italy and the rest of European countries, in order to understand if the scenario just analysed is common to other national contexts. To meet such need of comparison, in 2016 Eurostat has estimated the number ICT workers in the European Union between 2011 and 2015. The trend shows an increase of 1.5 million units between 2011 and 2015, an increase 0.5 percentage points in the quote of ICT workers to the total workforce (form 3.0 of 2011 to 3.5 of 2015). In Italy the trend is also an increase of ICT workers, but only of 0.2 % in four years (from 2.3 in 2011 to 2.5 in 2015). France and Germany have the highest growth rate in the time period (+1.1%); overall, the countries in which the ICT workforce represents the most relevant part of the total workforce in 2015 are Finland (6.5%), Sweden (6.1%), the Netherlands and United Kingdom (both 5.0%).

Eurostat shows also data regarding ICT workers in relation to the level of education, gender and age (Fig.5.9; 5.10; 5.11). European ICT workers are more educated, compared to the rest of the workers (60.5% of ICT workers with a degree compared to 33.40% for other sectors' workers). In ICT sector there is also a higher percentage of workers who are younger than 35 (36.40% against 30.9% for other sectors); furthermore, the concentration of men is higher in ICT sector (83.90% against 53.90%). Italy from this point of view occupies the last positions of the ranking both for age and level of education (25.40% and 33.10% of under 35 and graduate's workers). The gender gap is heightened in comparison with the European mean, with 86.20% of men that occupy ICT positions (Saporiti, 2016).

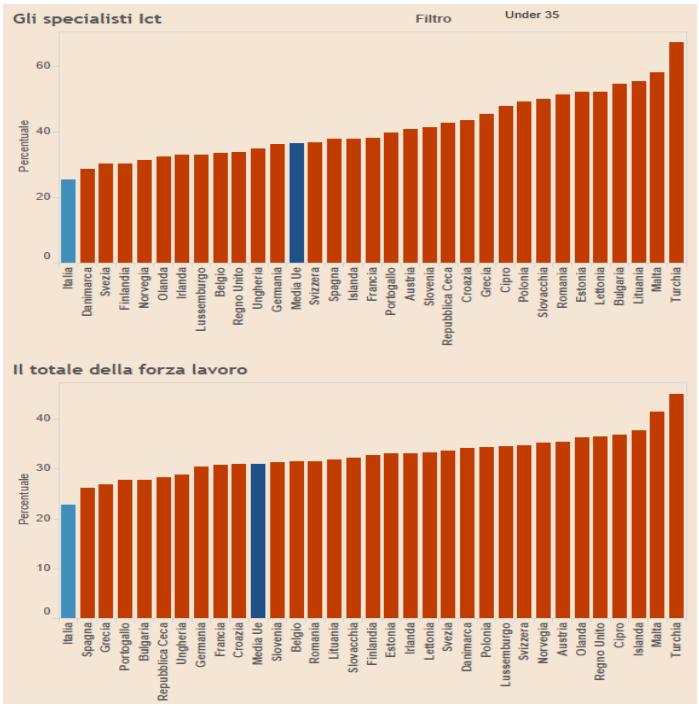
The situation underlined by the data showed isn't really encouraging for ICT sector in Italy. The fact that companies invest less for hiring ICT specialists compared to the rest of Europe is not a good sign for the economy, because of the value that such figures can bring to companies for growth and competitiveness, as showed in the first chapter. The issue should be promptly addressed, especially in the Italian context where youth unemployment has reached concerning values (Il Sole24Ore, 2017). Because of the speed of technological development, it is crucial to have an ICT workforce composed by young workers that can better use the technology available today.

Fig.5.9.Graduates percentage for ICT workers and total workforce



Source: Il Sole 24 Ore (2016); based on Eurostat (2016)

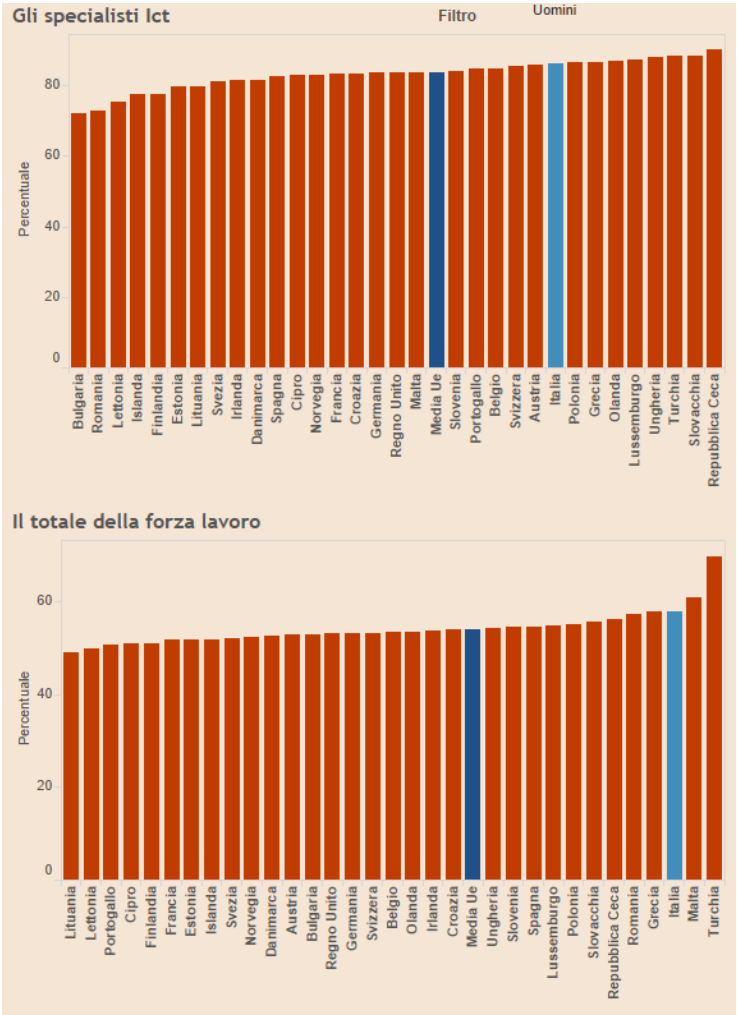
Fig. 5.10.Under 35 workers for ICT workers and total workforce



Source: Il Sole 24 Ore (2016); based on Eurostat (2016)

Some good news come from the Assinform observatory, that in the 2016 report ("Il Mercato ICT 2015-2016") analyses in detail the trend of Italian GDP generated through ICT technology, focusing on the various components of this sector and their use. Despite the decreasing trend for ICT market observed until 2014, in 2015 the revenue has grown of 1%, reaching 64.9 million euros; the forecast for 2016 is also positive, and estimated to be of 1.5%. Most of the components of such market, such as ICT services, Software and ICT solutions, digital content and advertise, have grown, except for telecommunication that suffers from a decrease of prices. An interesting aspect is that the use of computers and tablet has decreased (-13.7% and -15.1%) from 2014 in favour of smartphone (+9.9%), shifting the market towards the mobile application. Assinform highlights a growth also for the more innovative components of the sector: for example platforms for developed data management and web management have increased of 14.1%, followed by cloud computing (+28,7%) and IoT (+13.9%). Finally, the territorial gaps are showed not only by Istat, but by Assinform data too (Assinform, 2016).

Fig.5.11. % of men workers for ICT sector and total workforce



Source: Il Sole 24 Ore (2016); based on Eurostat (2016)

5.2. The Italian Open Data strategy

Since Open Data became a discussed matter around the world, policies and initiatives on such theme have been taken in Italy at various levels of governments and in different sectors. The country has seen a national commitment, some relevant initiatives at regional and municipalities' level, and also interest and debate brought by citizens. This section focuses on highlighting some of the main phases of the Italian commitment to Open Data, with particular attention to the actions taken by the central government.

In 2011 Italy joined the Open Government Partnership. The country also took an important step in opening data by launching a first version of the national Open Data catalogue dati.gov. The Open Government Partnership is an international initiative that works with a multilateral approach and aim to "*secure concrete commitments from governments to promote transparency, empower citizens, fight corruption, and harness new technologies to strengthen governance*" (*opengovenrmentpartnership.org).* It was launched in 2011 by eight countries (Brazil, Unite Kingdom, Indonesia, Mexico, Norway, Republic of Philippines, South Africa and United States of America) but now the members have increased to 75 countries. Being part of this international community implies a constant commitment to Open Data and in general to the value of transparency; such commitment has to become concrete through the following initiatives, which are parts of the Open Government Declaration approved by the member states:

1. The development of an action plan for the following years, with the involvement of civil society and the public administration, which collects commitments and projects concerning the themes on which Open Government Partnership has an interest on.
2. The production of evaluations and independent reports on the progress made
3. The dissemination of open government in other countries through the exchange of best practice, technical assistance, technology and resources

Source: www.open.gov.it; last visited on 29/03/2017

The production of the Action Plan is the most important task required from Open Government Partnership to define the Open Data policy that a country undertakes for the following two years. Italy has already produced and completed two different action plans since its participation to OGP started, and between June and July 2016 the Third Action Plan has been released. The first action plan was presented in 2012 thanks to the collaboration of various stakeholders, such as the Department of Public Administration, the Department of Digitisation and Technological Innovation of the PA, the Ministry of University and Research (MIUR), the Ministry for Territorial Cohesion, the Independent Commission for Evaluation, Transparency and Integrity of PA (CIVIT), as well as several universities and research centres. The Action Plan highlighted some initiatives in the field of open government which were already developed within the country; the sections of the document underlined the aspect of "simplification through participation in the PA", "Engaging citizens", "Open Data" and "Open Cooperation". The new commitments concerned Ethics, Transparency, Participation, Open Data and the promotion of social innovation and e-government. In 2012 other

two initiatives are worth mentioning: the release of the guidelines for semantic interoperability through Linked Open Data, which introduced the "open by default" principle (*agid.gov.it)*, and the third OGP European Regional Meeting hosted in Rome.
In 2014 an independent evaluation report on the first action plan was released, and the second action plan was presented in December by the Department of Public Administration together with the Agency for Digital Italy (AGID), the National Anti-Corruption Authority (ANAC) and representatives of civil society. Six actions were divided in three areas: "Participation", "Transparency, integrity and accountability", and "Technological Innovation". Furthermore, AGID released the Italian National Agenda and the Open Data Guideline.
In 2015 a new version of the national Open Data portal was launched. Also during the years several initiatives of Open Government have been developed, involving different levels of government, actors and sectors (OpenParlamento, OpenCoesione, OpenBilanci, OpenExpo, ItaliaSicura, SoldiPubblici, Confiscati bene etc.). Some Italian Regions createda regional Open Data portals, in the attempt to collect and publish data also at the local level (dati.piemonte, dati.trentino, dati.toscana, dati.emilia-romagna, dati.lazio, dati.lombardia etc.).
In 2016 the Italian version of the FOIA was approved by the government, a step further towards transparency and civic participation (*foia4italy.it)*. Through the establishment of a working group and the creation of the Open Government Forum, on the 20th September 2016 the Third Italian Action Plan was published; three different working groups participated: "Transparency", "Open Data, Accountability" and "Participation, Innovation and Digital Citizenship". The last action plan is a program for the 2016-2018 period, and includes 34 action divided in 4 areas: "Transparency and Open Data", "Participation and Accountability", "Digital Citizenship and Innovation" and "Digital Skills". The number of actions is much higher compared to the previous action plans because stakeholders of local level of government were involved in the process, so it was decided to make an action plan comprehensive of the whole Italian context, not only of the national policy perspective. This choice presents some limitations because of the difficulty that can emerge in monitoring and implementing the local initiatives. Involving different levels of government is certainly useful for collaboration, which can result in interesting exchanges of ideas and insights, but poses ultimately some issues in defining roles and competences in the policy cycle.

Monitoring the commitment taken in the action plans is one of the prerogatives for the Open Government Partnership. The Independent Reporting Mechanism (IRM) provides a report for each action plan; the last one was provided for the second action plan and analyses the progress made in the 2014-2015 period and give also an interesting look at the open government policy state in the country (IRM, 2015). The report states that only one ("SoldiPubblici") out of six actions has been fully completed; it is interesting also how it is argued that one action ("Digital Citizenship") cannot be evaluated because of its lack of clarity and specifics. This critique suggests that posing clear objectives is one of the fundamental aspects required to undertake an open government policy. The document also analyses the national context, looking at the themes that concern OGP the most, such as transparency and Open Data; it is underlined that, despite steps forward have been made in Italy since 2009, Open Data is still far to be a mean for tackling corruption, because the relevant datasets

aren't open, like the meeting's agenda of parliamentarians with important figures of the private sector. The Italian Open Data supply is defined as "primitive" compared to the British and the American ones: the data demand is still too low and it is guided by a small group of IT experts and journalist, instead of citizens. Such group is mainly composed of data re-users, which demand more data openness. Because of this "exclusivity" of the Open Data demand, it is hard to obtain interest on such theme within the common citizenship. Data confirms such tendency: only 7% of citizens use Open Databases, the 12% has used Open Data for monitoring scopes and 76% has never made a request to access public documents (Presidenza del Consiglio dei Ministri, 2015).

5.2.1. Open Data in the Italian context: an overview

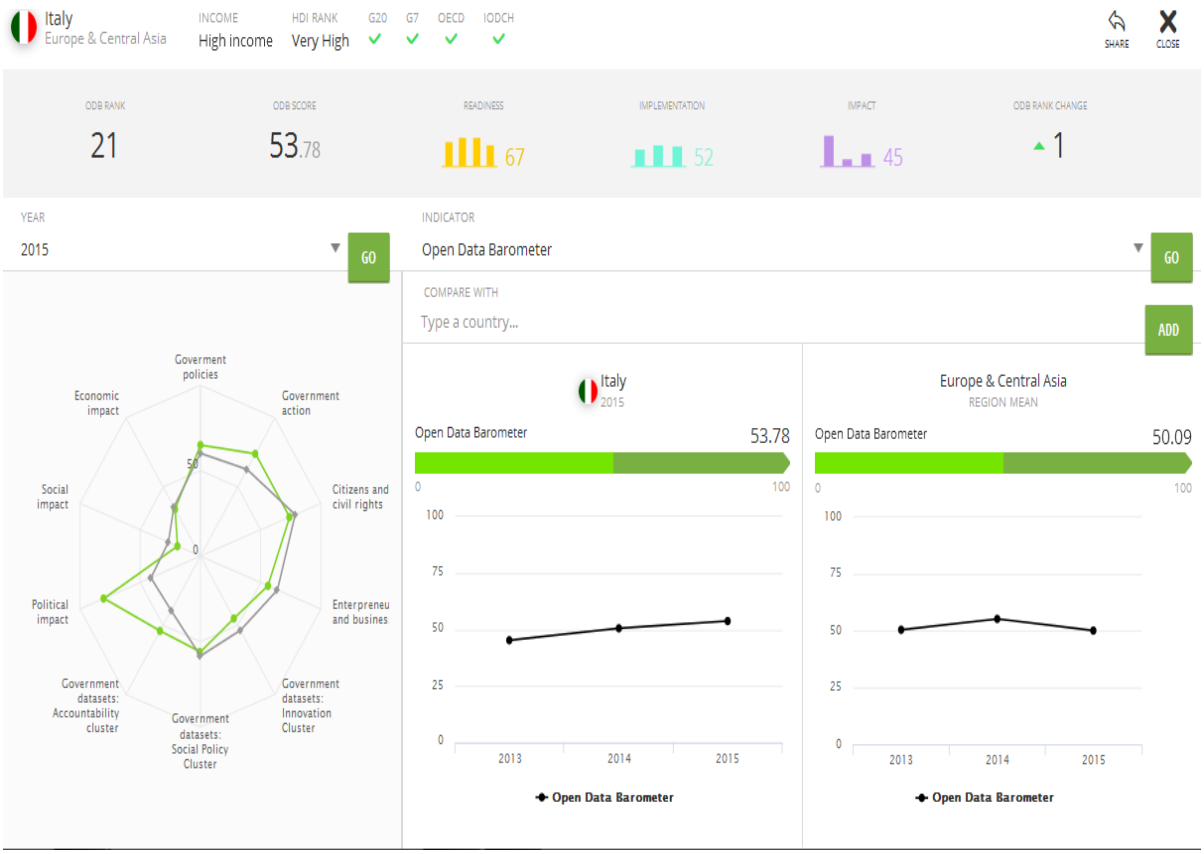
A point that needs to be stressed out, in order to understand how Italy is performing in the Open Data field, is the comparison with the rest of the countries, especially in Europe. The World Wide Web Foundation and the Open Knowledge Foundation try to meet this need and have developed two useful tools for international comparison. The World Wide Web Foundation has developed an initiative called "Open Data Barometer" (*opendatabarometer.org*): the Open Data Barometer uses data collected through surveys and other sources (World Economic Forum, World Bank, United Nations e-Government Survey and Freedom House) to elaborate Open Data statistics on the 92 countries that have an Open Data strategy. These statistics are gathered in three areas, "readiness", "implementation" and "impact", and from them a rank of the countries is made, based on the score achieved in the three areas. In the last edition of the Barometer (2015) Italy has scored 53.78 points, and occupies the 21st position (out of 92 countries) of the global rank (Fig.5.12). The higher parameters are measured for the "readiness" area, especially for the "government action" and the "citizens and civil rights" spheres. The lowest statistics are for "impact", especially the social and economic ones. The Open Data Barometer gives general scores of the crucial aspects that are required for a complete and successful Open Data strategy. But for more specific data on the degree of openness of national data ecosystems it is necessary to look at the Global Open Data index, developed by Open Knowledge Foundation (*index.okfn.org)*. This index analyses the degree of data openness for different sectors: 13 sectors are classified in 3 categories ("open", "non-open" and "unsure"). The approach is different from the Open Data Barometer and innovative because:

"The Global Open Data Index is not an official government representation of the Open Data offering in each country, but an independent assessment from a citizen's perspective. It is a civil society audit of Open Data, and it enables government progress on Open Data by giving them a measurement tool and a baseline for discussion and analysis of the Open Data ecosystem in their country and internationally from a key user's perspective."

(*index.okfn.org)*

The 2015 ranking sees Italy in the 17th position (out of 122 countries), with a score of 55% (Fig.5.13).

Fig.5.12. Open Data Barometer 2015 - Italy



Source: www.opendatabarometer.org (2015)

Fig.5.13. Global Open Data Index 2015 – Italy

Rank	Dataset	Breakdown	Location (URL)	Format	Info	Prev. (2014)	Score
1	National Statistics		http://dati.istat.it/	XLS, ...		#1 100%	100%
1	Election Results		http://elezioni.interno.it/ope...	CSV		#1 100%	100%
1	Government Budget		http://www.rgs.mef.gov.it/VERS...	CSV, Excel		#1 100%	100%
7	Legislation		http://www.normattiva.it/	XML		#6 90%	90%
8	Government Spending		n/a	n/a		#15 10%	10%
12	Procurement tenders		http://portaletrasparenza.avcp...	XML		n/a	90%
19	Pollutant Emissions		http://www.brace.sinanet.apat....	XLS, CSV		#23 60%	60%
21	National Map		http://www.igmi.org/download.p...	shp, ...		#64 15%	60%
22	Land Ownership		n/a	cxf		n/a	35%
30	Location datasets		n/a	mdb, shp		#34 25%	35%
58	Company Register		n/a	n/a		#26 45%	15%
69	Weather forecast		n/a	n/a		n/a	20%
74	Water Quality		http://www.sintai.sinanet.apat...	XML		n/a	0%

Source: index.okfn.org (2015)

Another interesting initiative launched in 2012 by Open Knowledge Foundation is the "Italia Open Data Census": with a similar approach to the Global Open Data Index the Census aims to collect information from citizens on the openness of the municipal datasets in different sectors (*it-city.census.okfn.org/, 2016*). Despite most of the data is still not available, in the future it could become a useful instrument to assess the Open Data state of Italian cities.

As Italy has taken steps forward in the development of an Open Data ecosystem, it is useful to identify the aspects which still represent an issue in the Italian context, as well as the ones that represent positive aspects. Most of the critiques from experts are moved to the national policy; Iacono (2016) argues that even though there are good national guidelines, there is a lack of tools for monitoring and coordinating, which stems in a frozen national agenda. Also the national portal hasn't been updated since October 2015, but the last version offer more functionalities. Furthermore, the data culture remains limited to few and specialized sectors and most of the Open Data initiatives aren't developed with a clear address for reusing the data. Positive aspects are for sure the definition of a common standard for the Open Data catalogues and the creation of an online forum (*network.ot11ot2.it*) where some of the European Digital Agenda goals are discussed. The number of datasets available has increased, but this aspect doesn't necessarily mean that transparency and Open Data are becoming more effective. More standards for structuring data are needed in order to create better comparative tools (Napolitano, 2016). The private sector isn't still much active on the Open Data front, because of the lack of high-quality data and high-valuable data that the country offers. The Open Data Maturity in Europe report (European Union, 2015b) identifies the projects of ISTAT and "A Scuola di OpenCoesione" as good practices, but also argues that a multiannual national strategy doesn't exist; the Italian Open Data scenario is described by the report as being able to generate interesting initiatives, but with a very low impact on society; in fact the social impact is evaluated as nearly zero and the broad impact is at the 20th position in the European ranking.

What is to be addressed by future actions and how the process of Open Data dissemination should continue are themes that are being discussed from experts of the field; here are some of the recommendations provided (Iacono, 2016; Napolitano, 2016; Mochi Sismondi, 2016):

- The national strategy should focus on the adoption of standards and rules that define the data structure, because this aspect, along with the quality of data, is needed to make the data's reuse an easier task.
- An ecosystem has to be created through the continuous involvement of all actors who are interested in reusing data.
- The necessary skill for reusing data are needed, so education should go in this direction and allow the development of digital and data skills throughout the population.
- Policies and practices shouldn't start from zero, but must take the already existing experiences as examples and materials to develop more effective actions.

- The commitment to Open Data should comprehend more broad commitment to open government in order to achieve participation, collaboration, transparency, social cohesion and development
- In order to pursue a more effective Open Data strategy investments and resources are needed. Opening the data in the public sector could seem a practice that doesn't require so much resources, but in reality a strong financial commitment is needed to develop effective projects, especially if connected with a more general open government strategy

The Open Data approach within the Italian public administration could also present one issue: in order to guarantee truthfulness and formalisation of public data and documents, they are published in non-modifiable formats. If this approach would be changed and, in order to make data reusable, open formats will be used, other forms of guarantee should be created. The public administration should first assure true data, and then find ways to make it accessible and reusable; it is not easy to make such different requirements to meet.

CHAPTER 6

Open Data for business

6.1. Economic impact - literary review

In section 4.3 of the fourth chapter benefits that can derive from using Open Data are listed, and the generation of economic value is mentioned as one of them. Such value can be generated by the creation of businesses, efficiency and decision making, all elements that contributes to the economic growth and development of a system. It is important to see if this link between Open Data and economic value can be actually observed through empiric evidence, looking for business models and ways of using Open Data that allow profit making. As it is described in chapter 4, a first global qualitative research confirms that economic value can be effectively generated by the opening of data (Young & Verhulst, 2015).

It has to be noted that in most of the analysis conducted the term Open Data doesn't only refers to data shared by public sector's subjects. Despite the fact that this theme has been discussed referring mostly to Open Government Data, because of the focus that the debate has had on transparency and accountability, Open Data per se can refer also to information shared by companies and privates. GovLab defines such element as "shared corporate data", information held by companies and made available for external subjects; such information can be also not strictly speaking "open", thus made available and reusable for everyone, but identifies exclusively the information that comes from companies and not from public bodies (Verhulst and Caplan, 2015). Shared corporate data represents a relevant element because companies are also important data producer, especially of information regarding their own sector of business. Open Government Data and shared corporate data can be merged to create even more consistent value for economy and society.

An attempt to quantify the economic value that can derive from Open Data is made by McKinsey Global Institute (2013), which analyses seven sector of the US economy on which Open Data has an impact. Estimates are made on how much this value is in US dollars; the forecast on global impact on these sectors is estimated to be between USD 3.2 trillion and USD 5.4 trillion, with the highest estimates made for education (USD 890-1,190 billion), transportation (USD 720-920 billion) and consumer products (USD 520-1,470 billion). Economic value emerges also from sectorial studies, concerning especially geo-spatial data (Oxera, 2013; ACIL Tasman, 2011)

Focus on the European Union is brought by the European Commission, with a study that aims to highlight the economic value of data among EU 28+ countries (European Union, 2015a). The challenge of this study is also to predict the value trend to 2020. Value is estimated for every European country, but also distinguished in direct impact and indirect impact. Direct market size for 2016 is expected to be EUR 55.3 billion, with an increase of 36.95% to 2020 leading to EUR 75.7

billion. Furthermore, total market value is identified in a number between EUR 193 billion and EUR 209 billion with a projected increase to EUR 265-286 billion for 2020. Most of the potential is identified in the Public administration sector (EUR 22 billion in 2020). Total number of direct Open Data jobs is estimated to be of 75,000 in 2016, while the forecast for 2020 is 100,000 jobs.

Looking only at Public Sector Information, Vickery (2011) estimates that the broad economic impacts that can be achieved through the use of the whole information detained by public bodies is between EUR 70 billion and EUR 140 billion. A comparison between the Open Data approach and the data pricing approach in the public sector can be useful to understand if the economy can benefit from a system where data is more accessible and reusable. The research of Koski (2011) finds that:

"Firmsfunctioning in the countries in which public sector agencies provide fundamental geographical information either freely or at maximum marginal costs have grown, on average, 15 percent more per annum than the firms in the countries in which public sector GI is priced according to the cost-recovery principles."

(Koski, 2011)

In the UK only, an independent study estimated that the direct benefit of Public Sector information is around EN 1.8 bn per year (Shakespeare review, 2013). Useful insights are provided by Deloitte (2012) on how economic value is created from Open Data through the creation of new businesses. Five business model archetypes are deducted from evidence, after an observation on 240 companies and non-profit organisations:

1. Suppliers: these organisations provide data in open format, making other subjects able to use it and reuse it for their scope. Because of the free availability of data, it is impossible to generate direct revenues from it; nonetheless, indirect revenues can be achieved thanks to increased customer engagement, loyalty and reputational enhancement.

2. Aggregators: these subject aggregate data from different sources, most of the times both from proprietary sources and open sources. They also look for correlations between data, collecting them in the most useful way possible for reuse. From the data collected organisations can provide insights for businesses, consumers and also governments. The revenue mechanism is based on charging recipients for these insights; also a freemium mechanism can be used, with a free access to basic data and a paid access to premium data and other functions.

3. Developers: organisations that take data from suppliers and aggregators, and build functions and applications to sell to consumers. The Open Data used is usually very dynamic and in

the need of constant updating; this is the case of transport applications developed on transport data made available by public transportation companies.

4. Enrichers: organisations with an already established business that use Open Data to foster and enhance their products and services. Businesses can access to more efficiency thanks to data; for example, an insurance company can use demographic data for a better understanding of its clients.

5. Enablers: organisation that are not direct re-users of Open Data, but facilitate the supply of it through cost-effective solution for business.

Despite the evidence provided by such studies, giving precise estimates on value of information remains a difficult task to perform; in order to observe such value, the GovLab (Verhulst & Caplan, 2015) supports the idea of focusing on Small and Medium Enterprises (SME), because they are the subjects that could take most advantages from an Open Data ecosystem; also they are likely to be the main drivers for innovation through Open Data, because big companies see this new resource more as a threat than an opportunity.

Some reasons of the difficulty for value quantification are highlighted by World Bank (2014):

- The most relevant part of Open Data impact is indirect, because most benefits are taken by users of such applications, products and services based on data. Being most of the time online products, the marginal costs are very low and so the revenues for the provider of these services are also low. This aspect generates an imbalance between direct and indirect impact, in favour of the latter; indirect impact is hard to quantify, therefore it gets in the way of Open Data's total impact quantification.

- Data itself doesn't have a certain value, because the value comes from the combination of data with other factors, like innovation, ideas, technology.

- Benefits are hard to attribute exclusively to Open Data, when a certain product or service is based on data; quantifying the impact of every input that gives value to the output is almost impossible if, as it is in most of the cases, many inputs are used. Naturally a dataset can be generated by the integration of both open and non-Open Data, so assessing the impact of Open Data only seems quite difficult

- In many cases the benefits of Open Data takes time to emerge, precise estimates require so time to include all the activities generated by Open Data

6.2. Open Data 500

In this sort of research, which attempt to assess the economic impact of Open Data across society, the Governance Lab of New York, i.e. GovLab, launched an initiative that today involves six different projects of six different countries. This initiative is called the "Open Data 500 Global Network" and takes its first step in a research developed in the US in 2013, i.e. Open Data 500, which aims to identify the companies that use Open Data across the country. Other similar projects are following such approach: Australia, Mexico, Canada, Italy and South Korea have embraced the challenge and are now analysing their national context to look for companies that use Open Data for business.

According to its website, the Governance Lab's goal is to "*strengthen the ability of institutions – including but not limited to governments – and people to work more openly, collaboratively, effectively and legitimately to make better decisions and solve public problems".* The focus is on Open Data because they *"believe that increased availability and use of data, new ways to leverage the capacity, intelligence, and expertise of people in the problem-solving process, combined with new advances in technology and science can transform governance".*

(thegovlab.org)

The Italian research "Open Data 200" will be the centre of this chapter; first, it is useful to look at the US project to understand how this approach is built and its goals. Also the US research is completed, and the data collected is available, so that it is easier to analyse its results; the Italian research is instead at its first steps, and most part of the data has still to be collected; therefore, complete results cannot be presented and useful insights cannot be made because of the incompleteness of data.

Goals and research questions

Because nearly zero studies that identify companies which use Open Data exist, the GovLab tries to fill this gap by developing a research that aims to map the American ecosystem of businesses in such area. Open Data 500 is the first national scan of companies that use Open Data as a tool for their business, but it doesn't look for the identification of a significant sample for statistical analysis; instead, it is an attempt to obtain a broad understanding on which types of company use Open Data, how they can use data to develop a business, and the nature of the Open Data demand. The Open Data 500 goals are so defined in (*opendata500.com)*:

- Providing a basis for assessing the economic value of government Open Data
- Encouraging the development of new Open Data companies

- Fostering a dialogue between government and business on how government data can be made more useful

The GovLab in 2015 has provided a report based on the first results of the research; through the observation of 354 case studies of SME that use Open Data, some insights on the link between Open Data and business have been collected. There is a clear assumption that has oriented the research on focusing on SME: SME are considered the subjects that gain the most advantage and benefit from the availability of data because of the reasons highlighted in the previous section; in the US they are in fact estimated to be the source for 60% of new jobs. Although, it should be said that the research was still ongoing when the results of this report were presented, and at this time the dataset of companies collected by GovLab has expanded to 529 enterprises. Because of the partiality of the data in the report, the final analysis that will complete the project is expected to give some final conclusions.

The questions that GovLab tries to answer in the report are the following:

1. What is Open Data, and how is it related to shared data?
2. Why focus on small and medium- sized enterprises (SMEs)?
3. What types of open and shared data are being used by SMEs and start- ups?
4. What sectors are using Open Data the most?
5. In what market segments is Open Data being used?
6. What are four important steps that SMEs can take to add value to open and shared data?
7. What types of products are being created out of Open Data?
8. How are SMEs monetizing their products and services?
9. How can we evaluate the impact of open and shared data on SMEs and the wider economy?
10. What challenges do SMEs need to overcome to leverage Open Data for value creation?

These questions cannot be answered only through the data collected specifically, but a wider look to the whole research is needed: this is the case of questions number 9, 10 and 6, which focus respectively on the methodological issue of evaluating Open Data impact on the economy (question 9) and the challenges that SME find in the creation of value from Open Data (question 6 and 10). Also the first two question concern the ontological issue of defining Open Data and the explanation for the focus on SME, so they are not answered by the direct results of the research too.

Methodology

Finding the most proper methodology required to assess the economic impact of Open Data across society represent a particularly difficult task. It is also difficult to understand which is the best way to observe such value, because research in this field is still at its first steps; an analysis that tries to look for a precise estimate of the value generated by companies that use Open Data in a national context hasn't been concluded yet, so other approaches are needed to set a better basis for future research. Open Data 500 is one of these first steps in the field; the goal is not to provide a

comprehensive map of all the US companies that use Open Data, but to identify a group of companies and use them as case studies to understand how value is generated through data in the country. These case studies are analysed with a look to their business model, the market segment in which they operate, their size expressed by sales and employees data and their products.

The method chosen for obtaining such data is the survey, presented in appendix 1. The companies object of this research are not properly selected: a list of companies has been compiled through outreach campaigns, advice from experts and professional organisations, and additional research (Fig. 6.1); the survey is sent to this group of enterprises; nonetheless, companies which are not in such list can also participate spontaneously by taking the survey online. This method is representative of the attempt to obtain the more data possible and accordingly take more useful insights from it. It should be noted that the method just presented has been developed to reach only the first research goal, i.e. "providing a basis for assessing the economic value of government Open Data"; the other two goals, which focus instead on developing a better Open Data ecosystem of initiatives, has been addressed through the organisation of roundtables to which both government officials and companies have participated. Participants are the specific data providers from government and data users from profit and non-profit organisations; they have been brought together to discuss how improve the data providing mechanism for each agency and to exchange feedbacks on such matter.

Fig.6.1. Open Data 500 company identification

Outreach Campaign

- Mass email to over 3,000 contacts in the GovLab network
- Mass email to over 2,000 contacts OpenDataNow.com
- Blog posts on TheGovLab.org and OpenDataNow.com
- Social media recommendations
- Media coverage of the Open Data 500
- Attending presentations and conferences

Expert Advice

- Recommendations from government and non-governmental organizations
- Guidance and feedback from Open Data 500 advisors

Research

- Companies identified for the book, Open Data Now
- Companies using datasets from Data.gov
- Directory of Open Data companies developed by Deloitte
- Online Open Data Userbase created by Socrata
- General research from publicly available sources

Source: www.opendata500.com; last visited on 31/03/2017

Data and findings

- Which types of companies use Open Data?

The first aspect worth mentioning is the definition of which types companies use Open Data, in terms of their size and the number of employees. Most of the activities analysed are private businesses (397), 93 are public companies, 16 are non-profit and 6 are partnerships. Although the private realities are the most relevant users, it is interesting to find that also other types of activities are using Open Data as a source for innovation. This aspect underlines the fact that the availability of Open Data doesn't only concern the private sector, but also the public sector itself in its complexity of agencies, companies and departments is interest in having access to more data.

Regarding the size of the companies that use Open Data, it has already been highlighted how SME represent in theory the most interested subjects in using such instrument. Data collected by OD500 research in this particular aspect is shown in fig. 6.2. Although, turnover data of companies has not been collected or only partially collected: nonetheless, it can be said that the majority of them meet the number of employees' criteria necessary to be defined as a small-medium enterprise. Also in the GovLab report of 2015 a number of 354 SME is mentioned, so it can be assumed that most part of the companies applied to such category. An interesting fact is that 143 businesses employ less than 10 people. Nevertheless, big companies are not excluded from the research: 93 businesses have more than 10,000 employees, which mean that also big companies are interest in Open Data; this fact merits further consideration, because the assumption the mostly SME can generate value from Open Data could be lacking in empirical evidence. GovLab argues that big companies are not likely to be a part of the Open Data movement because they see Open Data more as a threat than an opportunity; with future research it will be clearer if such statement represents the truth.

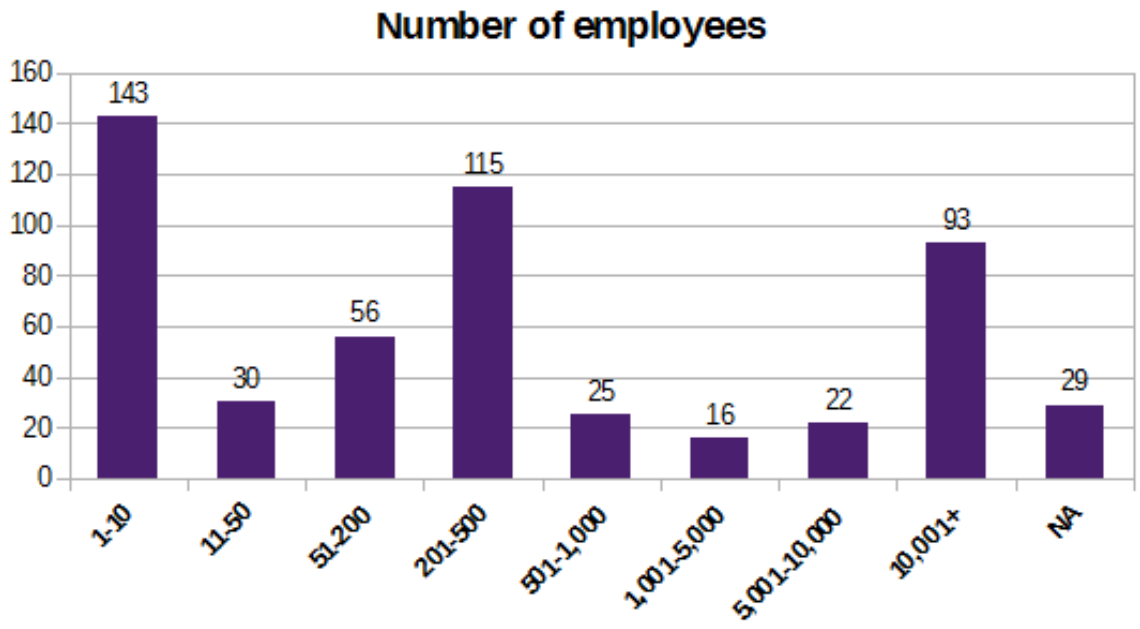
Unsurprisingly, the companies that are included in this research are mostly funded in the third millennium, with a peak in the 2007-2013 period; an explanation of such fact can be found from the data on companies sectors: being the technology sector the most interested in the Open Data market, as showed in the next section, and because of the rise of such sector in contemporary economy, it is clear that new sources of innovation are mostly exploited by these type of companies, which possess more technological tools and expertise compared to older businesses.

- Which sectors of the economy use Open Data the most?

Different studies have tried to identify the sectors of the economy that benefit the most from Open Data. For example, McKinsey (2013) finds that the highest impact of Open Data in terms of value in the USA can be achieved in the health care and finance sectors; from a study conducted in England by Deloitte (2012) the four sectors the emerged as the most likely to benefit from Open Data are Information and communication, finance and insurance, real estate and health care. The companies analysed by the Open Data 500 research show similar results, concerning the economic sectors in which the companies operate; although, it cannot be made a direct connection between

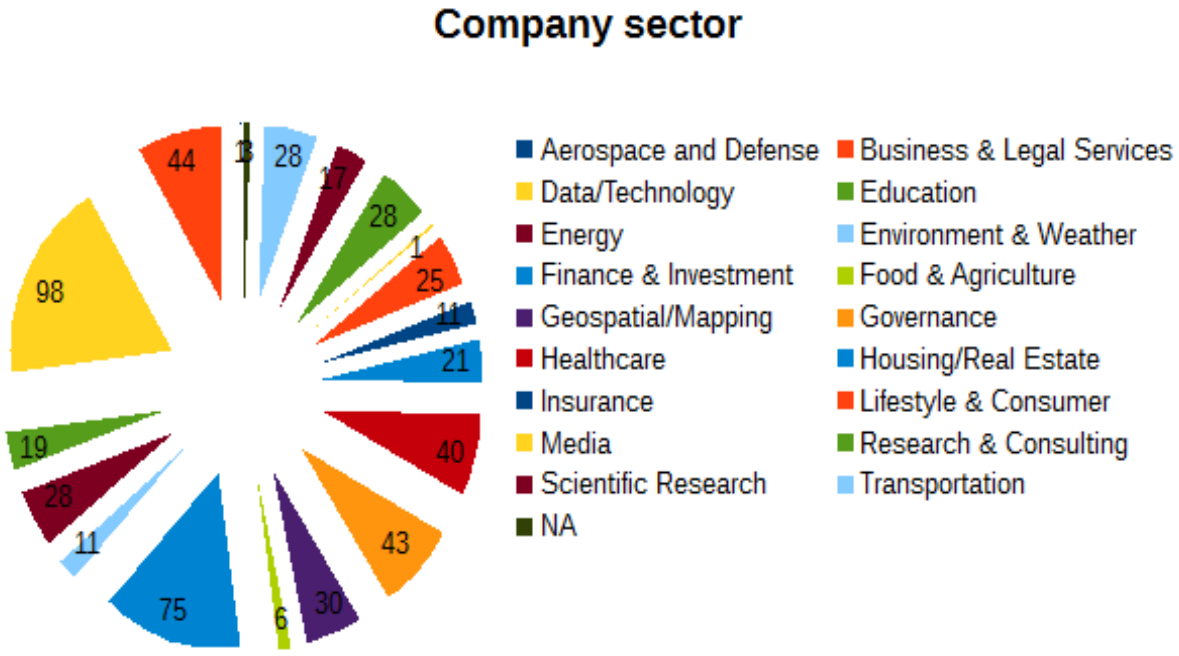
the number of companies that use Open Data and the value generated, and it should also be remembered that OD500 companies aren't a representative sample of the US economy.

Fig.6.2. Companies classified for number of employees



Source: personal elaboration from OD500 data (2017)

Fig.6.3. Distribution of companies for sector



Source: personal elaboration from OD500 data (2017)

As shown in Fig. 6.3, the main sectors of this group of companies are data & technology, finance & investment, business & legal services, governance and healthcare. The data & technology sector is obviously the sector that can benefit the most from the availability of more data; data indicates that these companies use mostly Open Government Data to develop innovative services and products and sell them to other companies. Finance & investment also work in the business to business market segment; they provide financial information and services to other companies that need them to be more competitive and to foster better decision making processes. Similar outcomes are provided by the business & legal services companies, which develop business software and analytics, consulting and strategic services and information for businesses and consumers. The health care sector benefit from data through the provision of patient-centred hospital and research services, personal health data management, outcomes research, and referral management solutions. This sector has recently shown a particular interest in the use of technology, and especially data, to deliver more efficient services and also to better manage the resources that are needed; it is known that health care presents relevant costs, so it is important to find innovative solutions to reduce resources' wastes to a minimum.

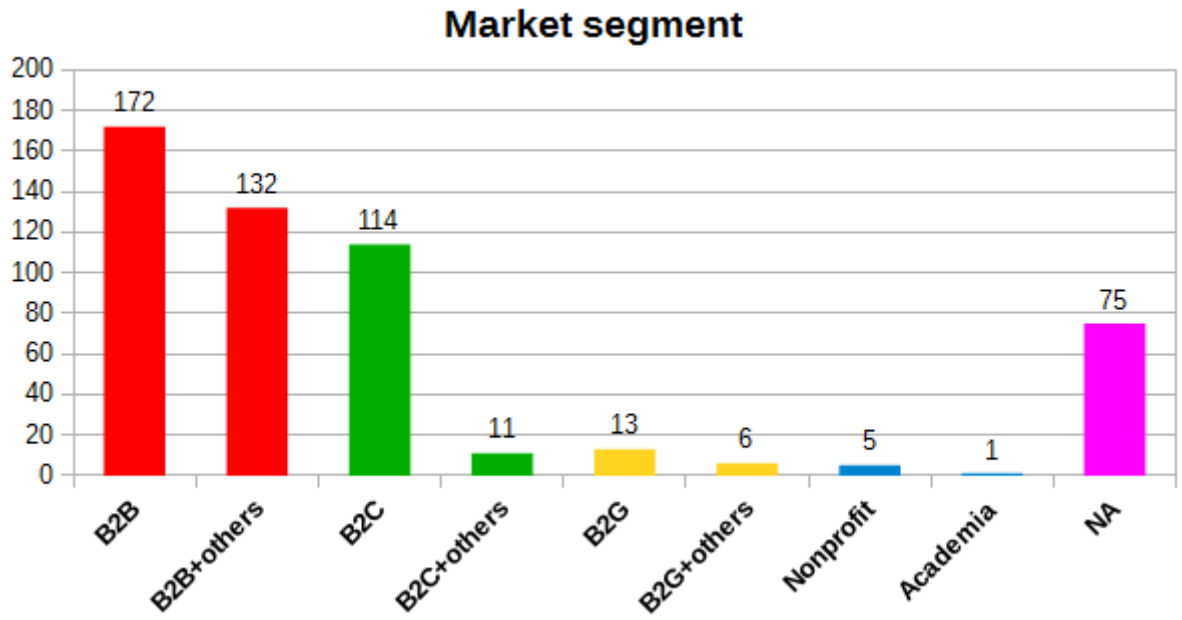
-In which market segments is Open Data being used?

In order to elaborate the business models through which companies generate value from Open Data, it is important to look at the market segments in which such companies operate; in other words, the subjects to which Open Data services and products are addressed need identification. It is interesting to identify these subjects because they are also beneficiaries of Open Data, in the sense that they receive a certain product or service, delivered by a company, that earlier wasn't available or useful as it can be now, thanks to the more data available. Therefore, they benefit of an indirect impact from Open Data: companies gain an economic benefit from the selling of the product, and the subject who acquire such product gain also a benefit from its use.

The OD 500 research classifies companies in five different market segments: "business to consumer", "business to business", "business to government", "non-profit" and "academia" (Fig. 6.4).

The market segment indicated the most by companies is the business to business, both in the preliminary results showed by GovLab in 2015 and in the updated data available at this moment on the research website.

Fig. 6.4.OD500 companies market segment



Source: personal elaboration from OD500 data (2017)

172 companies operate exclusively in the b2b market segment, and 132 jointly with other market segments. Two reasons in particular are presented by GovLab to explain such majority of b2b companies. The first concerns the characteristics of business in the digital era: business has become more and more data-centric as the digital revolution has advanced during the years. Economy and society have embraced a more intensive use of data, and companies are clearly not excluded in this process, in fact they are probably the most interested subjects of it, due to the benefits that data can bring to them (see chapter 2). For this reason they are highly interested in acquiring information and services to improve their business. Another explanation for such statistic can be found in the more interest that company have, compared to customers, in the so called "data as a service"; data itself can generate value in its raw form, but being able to use it through innovative tools that process it in useful ways is even more interesting for the private sector. The analytical tools can bring the value of data to its maximum level.

The b2c market segment is the second most chosen option by OD500 companies: 114 of them indicate it as their exclusive market segment and for other 100 companies is only a part of their business. Deloitte argues that the sectors that are more costumer-centred are the ones that will gain most benefits from Open Data; this fact represents an encouraging perspective, because one of the goals of Open Data is to reach people and to improve their quality of life.

Even though only a minority of companies of this group indicates that they do business for government, such companies are not to be ignored. As specified in chapter 4, benefits from Open Data can be received also by the government itself; companies can develop tools and services from the data and sell them back to the government. In this way the public sector is able to gain access to

resources that it can't develop on its own, due to the lack of competences and resources. Finally, for 75 companies data on market segment is still not available.

The data collected by the Open Data 500 research offers useful insights to develop other and more specific researches on the relation between Open Data and business in the US context. An issue encountered during the research activity concerns the survey: a relevant part of the data has not been collected due to the lack of response by the companies, which have not completed some of the important part of the survey. For example, one of the goals is to identify which types of data are used by US companies, but from the data available on OD500 website results that 387 businesses have not answered to the question. This fact poses a methodological challenge that should be addressed in future projects.

6.3. Open Data 200

Open Data 200 is the Italian node of the Open Data 500 network. It is a research that, as Open Data 500, aims to identify, map and analyse the activities that use Open Data for creating business or improve an existing one. The research is the first systematic study on Italian companies and organisations that look for the economic impact generated by the private and public sector thanks to Open Data. It consists in the identification of approximately 200 companies which operate in the national territory in order to offer an exhaustive picture of the Italian context. Because it is the first scan of companies in this field, it is not yet possible to determine the exact landscape of Open Data companies.

The project is designed by the Digital Common Lab of "Fondazione Bruno Kessler" located in Trento, with the support of the GovLab of New York that provides help for the identification of activities that use Open Data as a source for their business. The partners of the project are Atoka, an online platform developed by the company "SpazioDati" that aggregates and links data on enterprises, and Unioncamere, the Italian Chambers of Commerce. The Digital Commons Lab has decided to involve also experts in this field to work for this project. The results of the research activity can be useful for taking some insights and conclusion on the Open Data offer and demand in Italy; such conclusion could be used for developing and designing better Open Data policies. Furthermore, being a part of the Open Data 500 Network means that the study can benefit from the collaboration with other similar projects around the world, and can also compare the results.

Research goals and questions

The goals of the project are the same of its US equivalent:

- Providing a basis for assessing the economic value of government Open Data
- Encouraging the development of new Open Data companies
- Fostering a dialogue between government and business on how government data can be made more useful

(opendata500.com)

The lack of investigation in the reuse of Open Data in Italy has brought the necessity to fill this void by developing a research that look for economic and impact obtained thanks to Open Data. The open policies and programs that have been developed in Italy during the last years need now to be analysed and evaluated with the support of evidence from the users and re-users, in order to understand how the Open Data strategy should be addressed in the future. To find such evidence, the OD200 project has defined some research questions that aim to establish a picture of the overall Open Data users' ecosystem. Being the goals of the research the same of the US project, also the main research question are very similar (De Chiara, 2016).

1. What types of Open Data are used by Italian companies?

2. What sectors use Open Data the most?

3. In what market segment is Open Data used?

4. What types of products are created out of Open Data?

Methodology

One of the most challenging aspects of designing a research with such goals regards the company identification task. At this moment there is no effective method to collect a comprehensive database of all the activities that use Open Data, because of two different reasons:

- Today data is used in a large number of businesses that operate in many different sectors of the economy. Using data is no longer an exclusive operation made by a handful of companies, because the current tools that allow data to be used have reached a certain level of dissemination among society, thanks to the fast technological development in this field. Because of such widespread use of data, companies don't necessarily report in their website or communications that they use data, nor they say which types and formats of data are used. So it is hard to identify a sample of companies that reuse Open Data without having to contact them directly or to take information from experts of the field.

- Another issue concerns the companies themselves; Open Data is a relatively recent theme in the Italian context, which is still not fully understood and discussed by the general public. This fact is due to the gap that Italy suffers of in comparison with other countries like USA and United Kingdom, which have developed earlier and broader Open Data policies during the years. Because of such issue, it is clear that a large number of people, and businesses as a consequence, can't realize if they are using data in open format or not. Also because of the various formats in which data is used, distinguishing which format are used and take record of it represent a laborious task, considering the complexity of activities that a company

needs to face on a daily basis. Therefore, company identification results very difficult if neither the company themselves can identify the data they use.

For such reasons, the task of identifying the more activities possible that use Open Data has been followed through the consultation of some experts, the knowledge of the partners involved in such activities, and an outreach campaign. Also Big Data has been used, thanks to the huge database that contains information on all the Italian companies, i.e. "Atoka", made available by the company "SpazioDati". The companies were identified based on the following criteria: be Italy-based; earn revenue from its products and services; use Open Government Data as a key resource for its business. The final list includes 260 businesses of different size.
The methodology chosen for collecting data from companies is the survey. It is a more extended survey compared to the OD500 survey, with four section and 25 questions; it is presented in Appendix 2. The survey focuses on:

1. collecting fundamental information about companies, their business and their activity
2. understanding how companies use Open Data in their business

The survey is sent by e-mail, but also companies that are willing to take part of the project are allowed to submit their survey on the OD200 website (opendata500.com/it/). As for OD500, the goal of this project is not to create a rank of companies; it also doesn't provide a sample of companies for definitive statistical analysis.

Some issues emerged during the research activity. One of the major ones concerns the identification of company members able to respond to the questionnaire; CEO and CTO were chosen for this part of the research because they seemed the subjects that know better the company characteristics. As it has been easy to deal with small enterprises, contacting big companies has represented an issue. Also sometimes companies don't compile the full questionnaire, so some data is hard to collect.

Early findings

As above mentioned, the project is only at its early phases, and most of the data has still to be collected. For now, questionnaires have been sent to 55 of the companies selected and 140 survey have been collected through the online version of it. Data has been cleaned and presented at the 2016 Open Data Research Symposium in Madrid, after selecting 50 companies to deduct some early findings. Because these are preliminary results, this work focus instead on presenting some case studies, which can be a useful resource to understand in depth the role of Open Data in business; the case studies were selected for different geographical, business type and data sources characteristics. Nonetheless, some data on this first sample of 50 companies is presented to see where the OD200 project seems to be going.

- What sectors are using Open Data the most?

The technology & data cluster of companies is for now the primary sector of activity with 16 companies that belong to it. These companies are of different size and are engaged in different types of services, like software development, tourism and culture, scientific research, geospatial technologies, legal and business services. They all operate in the business to business market segment and the role of Open Data is very important for most of them. Other relevant sector in which Open Data is used are software and geospatial/mapping (Fig.6.5).

- In what market segments is Open Data being used?

Identifying the market segments in which companies operate is essential for understanding how business is created in this field. As fig.6.6 shows, the 50 companies selected are mostly in the b2b market segment (47), but many of them indicate more than one market segment; in fact, also a significant number (35) sell its products and services to government (business to government). Finally, 18 companies are part of the business to consumer market segment.

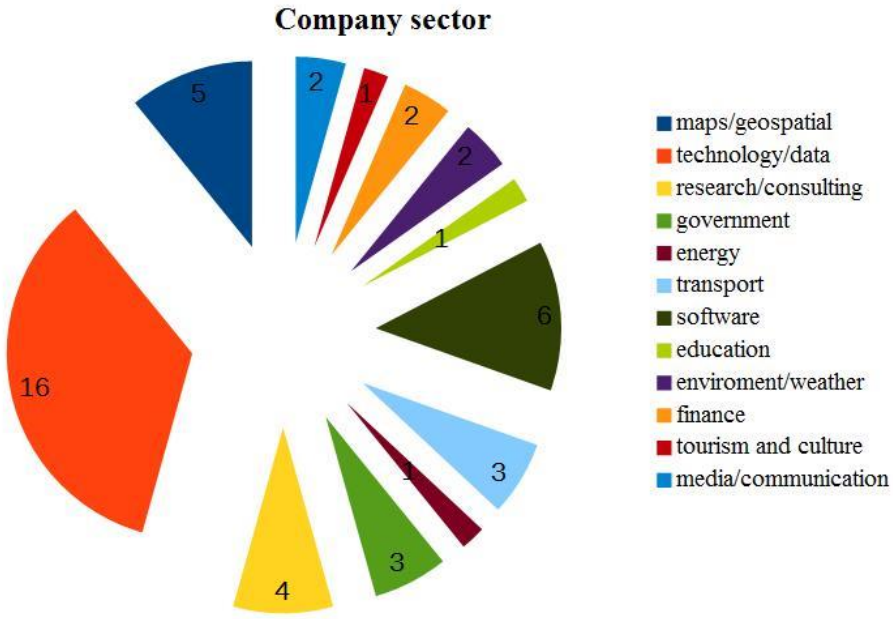
- Which types and source of data are used by companies?

The national data providers are the most used data sources by this group of companies. The National Institute of Statistics, i.e. Istat, is above all other sources in this rank, followed by the Ministry of Economic Development and the portal OpenCoesione. Some sectors like technology, media and finance use a wide variety of sources, and only the transportation sector uses data from only one source. The explanation is that transportation data are all harvested in a unique portal, the national data catalogue dati.gov.it. The most used data type is the geospatial data, used by 20 of the 50 companies (Fig. 6.7). This derives from the fact that more than one sector use such type of data, ranging from tourism and culture to agriculture. It should be noted that the health sector is disadvantaged by the lack of Open Data in this field.

- How companies use Open Data?

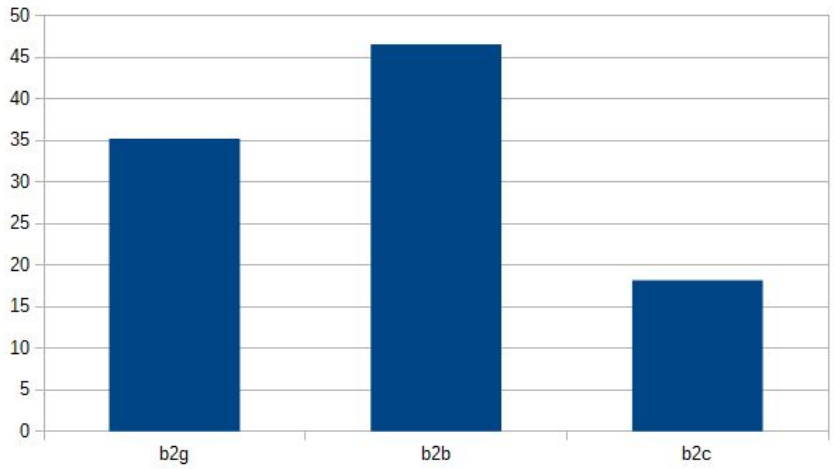
It is interesting to look at the role that companies take in the Open Data value chain: from these early results it seems that companies tend to occupy mostly the last phases of such chain. Only a small part of them are data publisher or data portals developers; instead most of them process data, provide analytics and develop API and apps (Fig.6.8). The main products and services created out of Open Data are: web and mobile applications, data visualizations tools, decision support and business intelligence systems.

Fig.6.5.Open Data 200 companies sectors



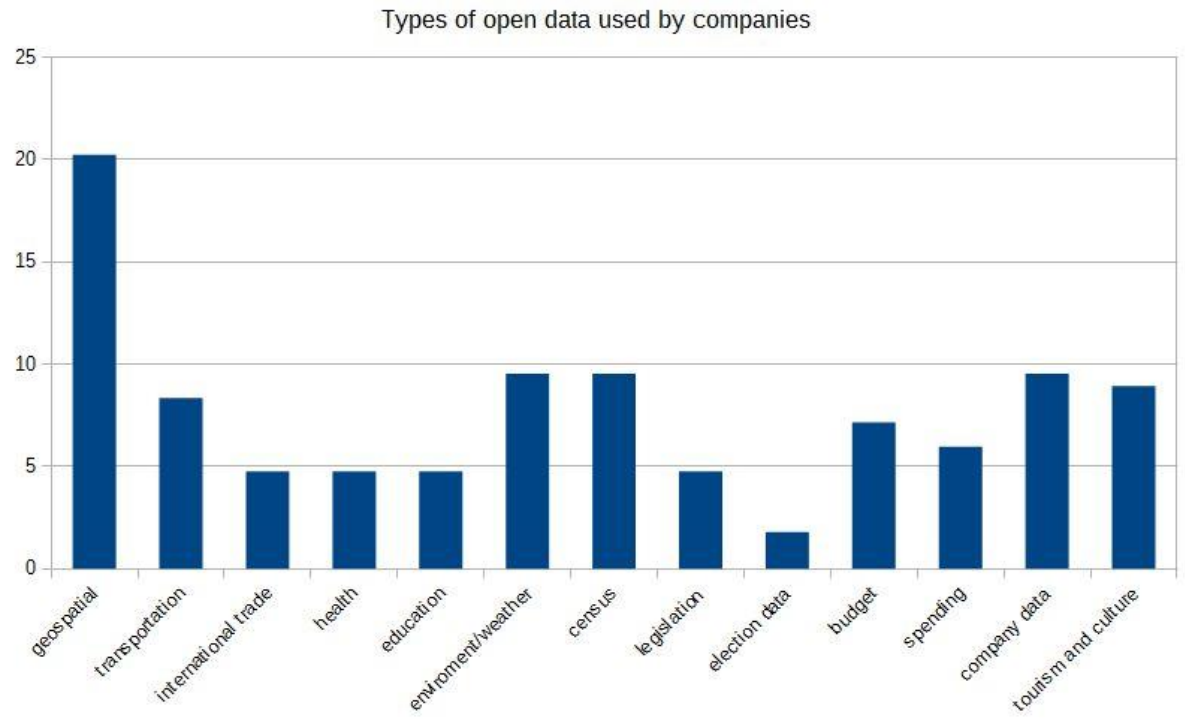
Source: personal elaboration from De Chiara (2016)

Fig.6.6. Open Data 200 companies market segment



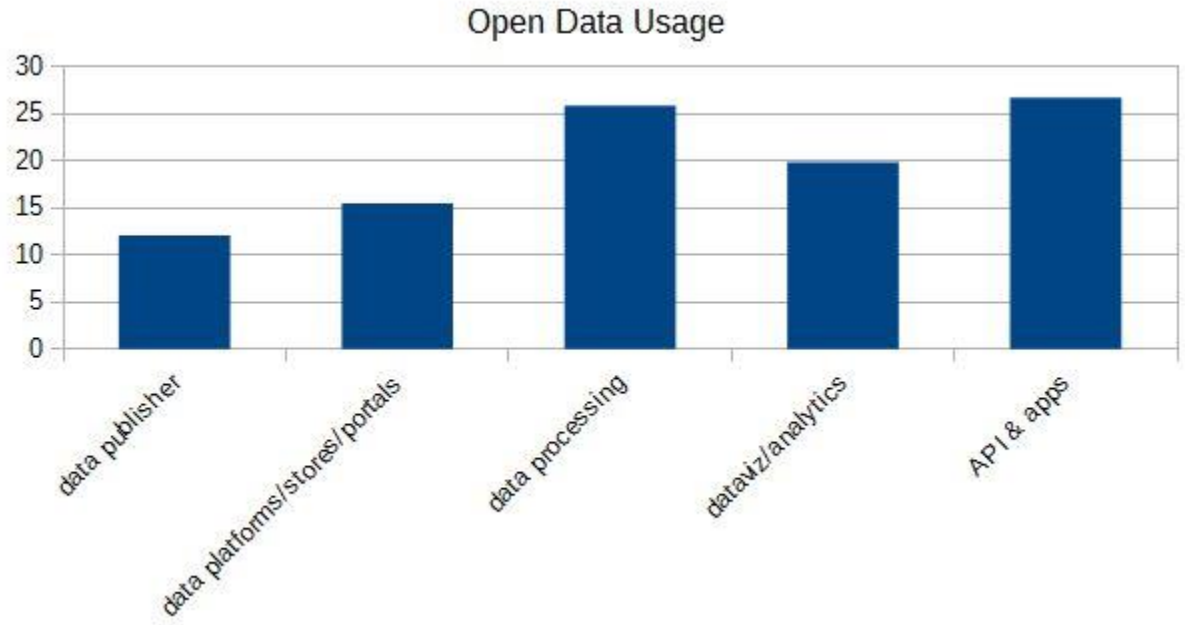
Source: De Chiara (2016)

Fig.6.7. Types of data used by Open Data 200 companies



Source: De Chiara (2016)

Fig.6.8. Open Data usage of Open Data 200 companies



Source: De Chiara (2016)

6.4. Open Data in the Italian private sector: four case studies

1) SpazioDati

Sector: Technology & data
Location: Trento (TN)
Year of foundation: 2012
Number of employees: 27
Web: http://www.spaziodati.eu/it/

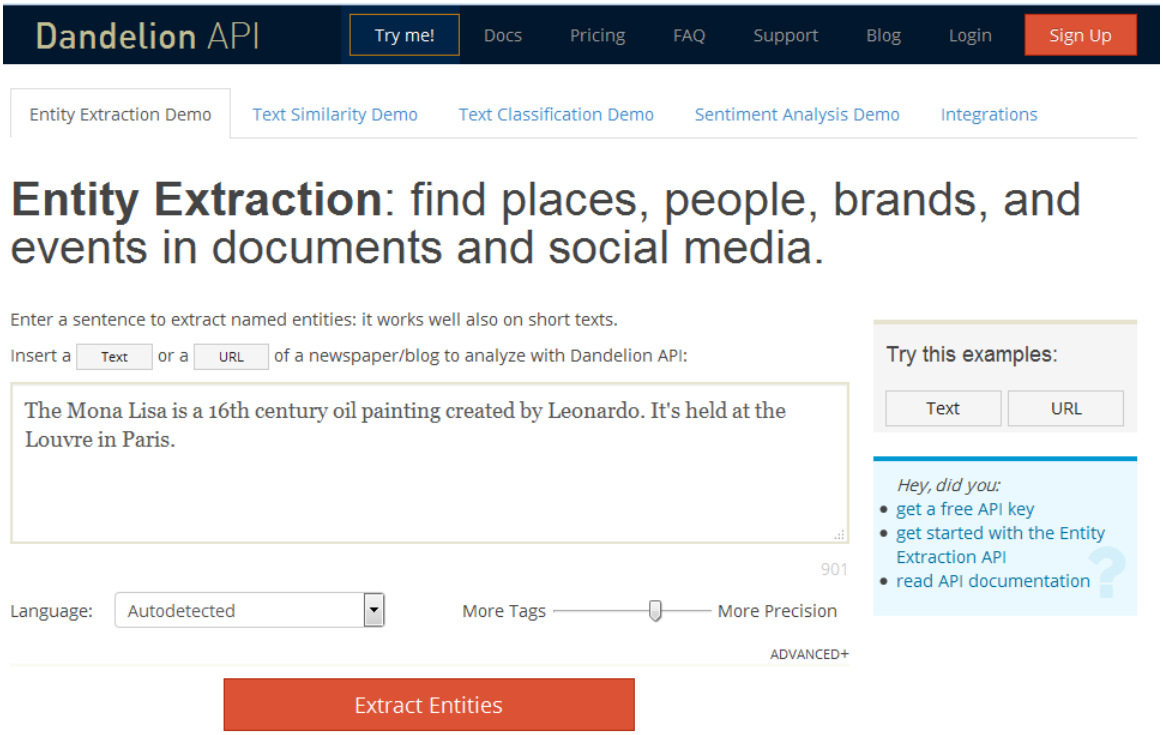
SpazioDati is a company located in Trento that operates in the Big Data and semantic web fields since 2012. Its activity is completely focused on data, for both the collection phase and the selling/publishing phase. The goal of the team that funded the company was to create a data market: "*SpazioDati was born on the idea of creating a data-market, what we liked to call a supermarket of data. We wanted to build the shelfs, taking the Open Data that existed on the web and put it on the shelfs after having rationalised, cleaned and linked it, but also taking data from companies (…) which can be reused in a completely different context*" (Barbera, 2016). The main product that the company offers is a semantic engine, a software that can extract data and information from a text. Such product was seen as a key resource by Cerved group, which through the analysis of economic and financial documents rates companies. The rate of a company is used by banks to assess the risk of loaning it a certain amount of money, so it is fundamental that the rating operation is very precise, in order to avoiding misevaluation by financial institutions. Because of such need for precision in the analysis of a company, in 2013 Cerved saw SpazioDati's product as a tool to improve the service that it provided, and due to the usefulness of the project decided to invest in SpazioDati. This operation contributed to a substantial growth of the company, which was able to expand its offer. Today the company counts 27 employees.

The activity of the company officially started with the development of an interesting project called DbPedia, an international project which Spaziodati is responsible for in its Italian node. DbPedia extracts the information contained in the known platform Wikipedia to make it reusable; in other words, the knowledge that Wikipedia offers is used to build a so-called knowledge graph, which links all the data coming from different sources in a single resource. The operation of linking data and the use of this knowledge graph are possible thanks to semantic tools, which through algorithms can identify entities in texts and connect them. From this process, accessing and browsing information becomes easier, due to the quantity of information potentially available in one single text. Wikipedia is now probably the most powerful tool to build such knowledge graph, because it is the most used online encyclopaedia throughout the world. Spaziodati doesn't only contribute to DbPedia in the linking of information, but also enriches the information contained in

the platform with other data taken from other sources. This constant work of data aggregation and enrichment is at the core of the company activity.

The two main products of SpazioDati are the semantic engine Dandelion and the database Atoka. Dandelion was earlier described for its function that had in the collaboration with Cerved Group, which helped the improvement of the company rating service offered by Cerved. The semantic engine basically takes data from texts and makes it analysable by computers. From this process multiple operations become possible due to the connection between the entities that the algorithm creates. Some features of Dandelion can be used online by anyone, and allow multiple operations to be performed, thanks to the knowledge graph that sustains the platform and the semantic engine that makes the knowledge useful and more efficiently usable. These operations are converted in Application Programming Interfaces (API) available for free in the demo version of the product; these API allow operations like entity extraction from a text, comparing different texts looking for similarities, text classification and sentiment analysis (Fig.6.9).

Fig.6.9. Dandelion demo API



Source: dandelion.eu

The product is sold through a freemium scheme, which allows free use to 1000 text per day, and requires to be bought in its full version after reaching such limit.

The other main product of SpazioDati is Atoka. It is the result of the ongoing collaboration with Cerved, which has provided the company its database containing information on all the Italian companies. "*As a part of the investment, in addition to an injection of capital in the form of cash, one of the most interesting part is that a deal for data exchanging has took place. Therefore, the*

database of Cerved has been given to SpazioDati (…); and Spaziodati, having access to this database, uses it do two things: it enriches it with unofficial data on the companies took from the web, social media, etc. (…); on the other hand we use the database for building our product, Atoka, which is a sales intelligence tool (..) that is built to make market analysis and targeting" (Barbera, 2016). The data that the company implements in the database comes not only from social networks, but also from useful platforms like OpenCoesione that takes record of all the funds received from companies through European structural funds; this information is very useful for understanding if an enterprise is able to develop successful projects or not. The database of Cerved has also been linked to the Wikipedia one, to build an even more complete source of information. Like Dandelion, Atoka is sold with a freemium scheme, which gives a free access to the database for a limited period of time. This tool can be very useful for different types of actors and institutions, giving information on all the Italian companies' framework.

The two products described are unique due to the multiple uses and scopes that can be used for; this diversified outcome that can be obtained from the products results eventually in a high diversification of clients. The nature of data lies at the base of such diversification, because as it has been said earlier in this work it is a raw matter that can be developed in various ways, depending on the needs. The market segments in which SpazioDati operates are b2b and b2g. Different clients in different sectors: obviously the main client remains Cerved, which now also owns a relevant quote of company; the data offered by the company was used by different actors, for example the "Commissione Stragi" in Italy and an Australian news company; the services of sales intelligence are sold to private companies that want to collect information about their market to make more informed decisions; the English Ministry of development commissioned a sector analysis on the digital health cluster to collect information on companies, and gave also in return this data that was added in the database of SpazioDati. This diversity shows how data services are highly requested both from business and government. The Italian market in which SpazioDati operates is relatively small, because less than 10 company have developed a semantic engine; the most known is IBM with Watson (*ibm.com).* For the part of data collection and the construction of a so detailed knowledge graph, SpazioDati is almost unique in the national context.

The company isn't focused only on the sales intelligent sector, but is one of the more active entities in the Open Data community, being both a suppliers and a re-user of data. Actually, SpazioDati can be identified with 4 Open Data business models: it is at the same time aggregator, developer, enricher and enabler of Open Data. This high commitment to Open Data comes from the belief of the value of reusing, which in the data sector becomes very tangible from experiences like SpazioDati. Not only Open Government Data, but also shared corporate data represent a resource for the company, which is able generate value for SpazioDati but also for the supplier of data itself. In this way the value generated can be distributed among both the data offer side and the data demand side. There is still data that SpazioDati aims to collect, but costs for acquiring Big Databases represent a relevant barrier for a small company; for example the data contained in the "Albo Pretorio" of public administrations could be very useful, but "*taking 8.100 Albi Pretori made in 8,100 different ways is a monstrous investment, impossible*" (Barbera, 2016). The opening of data

by the public sector should have precisely such goal: eliminate the existing barriers for using data.

2) DEPP - Openpolis

Sector: Government
Location: Rome (RM)
Year of foundation: 2003
Number of employees: 8
Web: http://www.depp.it/

DEPP is the acronym of Electronic Democracy and Public Participation. It is a company that in Italy is known for the development of web platforms that aggregate, process and shows public data regarding politics, both at the local and at the central level. The company has expanded its range of activities during the years, developing platforms of information for the monitoring of public funds and providing service directly to the public administration. The platforms are open source projects in which citizens can participate: they can enrich the information available by adding more information, respecting rules of accuracy; in this way the citizen is engaged and plays a very important role in verifying and monitoring the official information published by public agencies and distributed by DEPP. Another interesting characteristic of the platforms is that they are no-profit initiatives, and can be used freely by anyone. The company was founded in 2003 and the first platform, called Openpolis, was developed in 2007; the first experience saw a relevant success that led to the creation of the association Openpolis, which now incorporates and promotes the following non-profit projects: OpenParlamento, Open action, Open bilanci, Open municipio, Open politici and Voi siete qui. Openpolis is the primary Parliamentary Monitoring Organisation (PMO) in the Italian context; as the website describes it, it is an association that "*works with Open Data, makes open source projects, promotes open government, builds polis on the Internet (...); the volume of information is available to anyone for free, and from the aggregation of data it becomes data-driven journalism, a profound and constant investigation on Open Data which reveals facts. This work has transformed us in little time in a civic observatory of politics that analyses on a daily basis the complex mechanism that move Italy"*(*openpolis.it*). DEPP and Openpolis today counts jointly 8 employees and 4-5 consultants.

The main activity of the company is the development of online platforms, so it is interesting to look at some of them. The first platform developed by DEPP is Openpolis, which later took the name of "Open politici" and "OpenParlamento". Open politici allows the monitoring of all the elected officials of the country through simple tools that show the relevant information about them and especially their political careers (Fig.6.10).

Fig.6.10. Open politici website

Source: politici.openpolis.it

The data source on which the platform is built on is the national archive of the Ministry of the Interior, which offers the data but in a less useful form, being sometimes hardly accessible and incomplete. The data on politicians can be found by searching the institutions in which they are elected, the territory from which they come and obviously their name. OpenParlamento is a similar platform, but focused only on the representatives at the central level of government. All the data about their activity, such as votes, laws' proposal and signatures, number of presences and matters of discussions are aggregated and easy to browse thanks to the platform and its visualisation tools. The information of the Ministry of the Interior has become partially open since 2011, but some restrictions on the activity of parliamentarians remain. Useful tools are the index of productivity, which allows ranking parliamentarians and making comparisons between them. It is also possible to monitor a specific representative and receiving constant information about him. The same scheme has been applied to another platform, Open municipio, which uses the official political and administrative data and publishes it in open format. For now only two municipalities, Udine and Senigallia, have joined the project, so it is still work in progress. From the information available during the electoral campaigns, coming especially from official statements, a tool for helping the voting decision of citizens has been developed: "Voi siete qui". It is the first experience with such goal on the country; after making 25 questions to the user, the web platform gives the positioning of the user's ideas on a graph that shows the position of different candidates, based on the answers given in the questionnaire. From the success obtained with such platforms, especially with OpenParlamento, the company obtained also some official assignments from the public administration; from the collaboration with the Ministry of Cohesion the platform OpenCoesione was developed. OpenCoesione is a tool that allows browsing all the destinations of European structural funds in Italy; data is grouped by territory, organisation and project, and the precise amount of money received is showed. Data is also published in open format, which makes it

reusable from anyone. Another interesting project is Open Ricostruzione, commissioned by the Region Emilia-Romagna to monitor the donations and the projects of the reconstruction process that took place after the earthquake of 2012.

The activity of Openpolis is totally no-profit, so during the years it became necessary to find a source of income; such source was found in the service of information that the company started to provide, such as data feed, data extraction and political analysis based on the data collected with the platforms. Some relevant clients of DEPP are now editorial groups like "L'Espresso" and "RCS", or newspapers like "*Il Fatto Quotidiano*" and "*Repubblica*". The major clients of the company are today public administrations, regarding the open government projects, such as "*Ministero per la Coesione Territoriale*", "*Regione Lazio*" and "*Senato della Repubblica*", but also communication and information entities that demand services based on the data collected by DEPP. The business is basically divided in the b2b and b2g market segments. A relevant choice made by DEPP is to not include commercials on the websites, which makes the projects totally independent but also hard to maintain. One source of income is represented by the donations and contributions made by users to help Openpolis in its sustenance. Although, the main source remains the already mentioned activity of DEPP in terms of data services, but the company aims to separate the Openpolis association from DEPP in the fastest time possible. The future prospects of the company include the development of a web platform aimed to sustain the social initiatives, such as petitions and class actions. Moreover, the platform Open Bilanci is being developed in order to "*make Italian municipalities' balances transparent, understandable, trackable and sharable by citizens"* (*openpolis.it).* The broad goal is to extend the action of transparency and data publishing to administrative processes and in general to all the sectors that are made obscure by bureaucracy.

DEPP is clearly a company that is interested in Open Data. Like SpazioDati, it embraces different Open Data business models, because it aggregates, enriches, develops, and enables the use of Open Data at the same time. The activity of Openpolis is fundamental in the Open Data framework in Italy, because it makes data that is available in open and semi-open format more usable by citizens; also the service of information that makes through data is fundamental to enforce democracy in Italy, because lack of clear and reliable information are evident issues in the country. This push towards transparency has been followed in some cases by important institutions like the Senate, which has made the data on the votes available after the development of OpenParlamento. The platforms developed by DEPP are the true tools necessary for mechanism of open government to take place across society, because they can establish communication and collaboration practices between public institutions, companies and citizens. In spite of the huge amount of data already available, barriers for further data use remain; Vittorio Alvino, chief executive of DEPP, says that the company "*is taking forward a campaign for the disclosure of data and information concerning the parliament's commissions"* (Alvino, 2017), because no data concerning the activity of the commissions is available today. The issue of data availability comes out often in the activity of the company. Another barrier is represented by the format of data, which sometimes cannot be easily reusable, for example "*the database of municipalities' balances is in html format charts, not*

reusable and intellegible". Costs barrier are encountered when dealing with the Chambers of commerce, which can be used only if bought. It should be noted, even though the company is the most experienced business in the field of PMOs, it counts only 8 employees; this fact can be seen as an example of the opportunities that such market has in terms of occupation: even though DEPP is the most relevant company in this sector, it can't really offer important job opportunities in terms of numbers, so probably the market dimension cannot really know a substantial growth in this sense. Quality of services and value from information can still be generated from an increased availability of data, but from the point of view of business the opportunities don't seem so clear. Another type of value added that these open projects can give concerns the competences; "(*in Italy) the competences (within public administrations) for developing and managing projects are very low (…); in the case of OpenCoesione the competences were consolidated in the public administration thanks to the project"*.

3) Planetek

Sector: Geo-spatial/Mapping
Location: Bari (BA)
Year of foundation: 1994
Number of employees: 45
Web: http://www.planetek.it/

One of the most growing sectors of the data economy is the geospatial field, which in Italy has seen a constant growth with double digits even during the years after the 2008 economic crisis (Zotti, 2017). One interesting and almost unique company of such sector is Planetek. Planetek operates since 1994 in the Geographical Information Systems (GIS) field, using geospatial information to create services that make this information reusable for a particular scope and for a particular client. The focus is on satellite data, pictures and information of earth's observation. The mission of the company is to "*simplify the complexity of space*" (Zotti, 2017); basically the activity always starts from the collection of data that can be georeferenced. Planetek is on the international market with different lines of business, so its peculiarity, compared to other companies, is that it works simultaneously in multiple sectors with different types of clients. A constant collaboration with important public spatial agencies has been developed, in order to collect data produced by them and also to sell services that are demanded by the agencies themselves. The solutions developed by the company are "*aimed to storing, updating and sharing geographical data for planning, projecting, managing and monitoring the territory*" (*planetek.it).* The context in which the company works is the Tecnopolis technological park incubator; thanks to the network of Tecnopolis, the company was able to exploit the competences and technologies of the GIS and Remote Sensing Laboratory technicians. The employees are today 45.

In the opinion of Giovanni Sylos Labini (2009), the company administrator, when Planetek was founded its degree of innovation was maybe too high for that time, so that it wasn't recognised as a potential successful business by the institution that provided funds for new entrepreneurship. The first line of business that the enterprise developed was focused on the provision of information services based on commercialized data, such as satellite pictures. An example of this type of business is the service that Planetek provided to Aeroporti di Roma (ADR), the company that manages airports in Rome; Planetek, through a contract established from tender, gives satellite images on the airports and services aimed to understand two things: how the territory evolves in time, and the state of construction works in the airports' fields. In this sense both data and a service are sold to the client. This product was also provided for the universal exposition in Milan (EXPO), which was given high-resolutions images every two months used to understand the evolution of the local territory due to the public works that were changing it. Other services and products are provided in the fields of environment, mobility, tourism, energy and utilities, security and defence, planetary exploration. Another type of data that is used to create services is transport data; thanks to the GPS data provided every 5 seconds by buses in the city of Bari, real time information on the vehicles is published, in addition to estimates of the time of arrival in the bus stops. In less than 5 week other API that exploited such data were developed by common users. The infrastructure and a service were built by Planetek, but also the community participated through the development of other services. The institution responsible for public transport in Bari was also provided with the statistical information on traffic, based on the steaming times of buses. Therefore, the transport company was given the instruments for a more efficient planning of the paths, and the municipality had new information to develop better traffic services throughout the city. These type of products represent the traditional business model of Planetek; "*these are things that we will continue to do for the next 10 years for sure, they are activities that are based on job orders, without a continuity"* (Zotti, 2017).

The first example of Open Data concerning the space sector has been seen nearly 10 years ago: a satellite mission called "Landsat" started in 1972, launched by the United States Geological Survey (USGS), which is the U.S. geological agency. The data provided by the satellite was initially commercialized, but during the years different studies showed that the costs of maintaining the mission infrastructure were not covered by the revenues coming from data selling. So the agency decided release data in open formats, at first only in part and then completely. This is the first historical example of Open Data in the satellite field, and opened relevant opportunities for business. An opportunity for a more sustainable business model for Planetek presented itself when the European program "Copernicus" started. Copernicus has the goals of monitoring Europe to have a better and fastest response to environmental emergencies. To achieve such goals, different satellites were launched, and the European Commission stated that the data produced were to be published with an open and free license. This service will be available for the next 34 years at least, which represent an interesting prospect for business. Planetek is now able to have a free source of satellite images for the next 34 years, which is a huge breakthrough in terms of sustainability of the business. Thanks to this data, a service for monitoring the territory is now sold on annual subscriptions. A

client of such service is "Metropolitana Milanese", which manages the water service of Milan; thanks to the service, the company is able to save resources in monitoring the soil's stability, intervening only in the critical areas showed by the satellite images.

Examples of clients have already been presented, so it is clear that the core is represented by public agencies and administrations: b2g represents the main market segment. In addition to the ones already mentioned, other public clients are the spatial agencies, both Italian and international, and the defence agencies. Looking at revenues, the government represents directly a quote of 45%, but in general the public demand in the spatial sector is the engine that allows these products to be sold, because important private companies are often financed by public entities. Altogether the defence sector finances the 80% of the company activities, through direct and indirect ways. For the b2b market segments the clients are "*big engineering companies and energy companies both in the oil and gas and the renewable energy sector*" (Zotti, 2017). The markets in which Planetek operates are different from one another: there are many Italian companies in the GIS sector, but less in the earth observation sector. The uniqueness of Planetek is the simultaneous work in different fields, which gives a competitive advantage to the company.

Planetek is the example of a company with a consolidated business that thanks to Open Data was able to improve it and make it more sustainable. The sustainability is possible for two reasons: the commitment of the European Commission to Open Data that made the Copernicus satellites information available for everyone; the long-term production of data that characterizes the satellite mission. Italy too has its satellite missions; the most relevant one has put in orbit 4 satellites, which can collect images at a time distance of only 6 hours one from another. Through the pressure made AIPASS, i.e. the association of Italian small aerospace companies, to the Italian Satellite Agency (ASI), a procedure for obtaining the data produced by these satellites with an open licence is now available. Planetek helps public administrations to follow this procedure for obtaining the data and at the same time sells its service, making the public administration interested pay only for the added service and not for the data. Massimo Zotti of Planetek notes that "*we are a clear demonstration of how Open Data can empower business models the essential condition for a commercial business are met: a reliable data source with a long-term prospect and with the same way of delivering the data over the years, so that it is machine-readable".* Planetek embodies mainly 2 Open Data business models: it is an aggregator and a developer.

An interesting fact is that when public data is not standardized across different sources, for example in the case of Italian regions, two things can happen; a service or a product that has been developed in a certain context cannot be replicable because of the cost barrier of re-formatting all the product on a different way of publishing and structuring data; this is the case of the traffic service developed by Planetek in Bari that it wasn't replicable for other municipalities. Nonetheless, from the differentiation of data structures a company can create a business, because it can be a collector of data coming from different subjects and sources and sell a standardisation service. Although, it should be said that the value created in this last case is only for the company, while standardized data can have an impact on multiple subjects; in the case of Bari public transport, the whole city

benefits from the service provided, and such benefit cannot be gained by other municipalities because of the standardization problem.

4) Mobygis

Sector: Environment/Weather
Location: Trento (TN)
Year of foundation: 2014
Number of employees: 5
Web: http://www.mobygis.com/wordpress/

Geospatial data is often indicated as an information type that can spur innovation throughout society, if freed of accessibility barriers and used for developing products and services (Napolitano, 2013). One tangible application of this premise is the innovative start-up "Mobygis" founded in 2014 by Matteo Dall'Amico, Fabrizio Zanotti e Silvia Simoni. Mobygis works with geospatial data to offer mobile solutions, implementing such data on a platform that, in addition to information, delivers geo-spatial tools and services. The main product of the company is called "Mysnowmaps", a mobile platform that deliver information on the quantity of snow present in the whole Alpine area. Because Mobygis started recently its activity, the focus of the company is mostly on this product, which has produced interest especially among other companies but also among public sector bodies and citizens. Thanks to the data collected Mobygis is also able to sell services that inform on the volume of water in rivers and forecast its changes. Maps with forecasts of the soil moisture are also provided for the agriculture sector. Today Mobygis gives work to 5 people, remaining a little but very innovative reality in the geo-spatial context. In 2015 the company won the Edison pulse competition for innovative projects in the environment field (StartupItalia!, 2016).

The most interesting activity of Mobygis concerns Mysnowmaps (Fig.6.11; Fig.6.12). The company has developed an algorithm that takes different types of data coming from different sources to deliver data on snow's height together with forecasts for both snow's volume and weather. 3 types of data are used:

- Data that concerns the territory, such as terrain elevation and various types of maps that show the territory's configuration. Data is in the form of spatial raster, and is basically static because it concerns aspects of the territory that don't change easily. This data is available at a regional level and also at the European level, and it is delivered through licences.

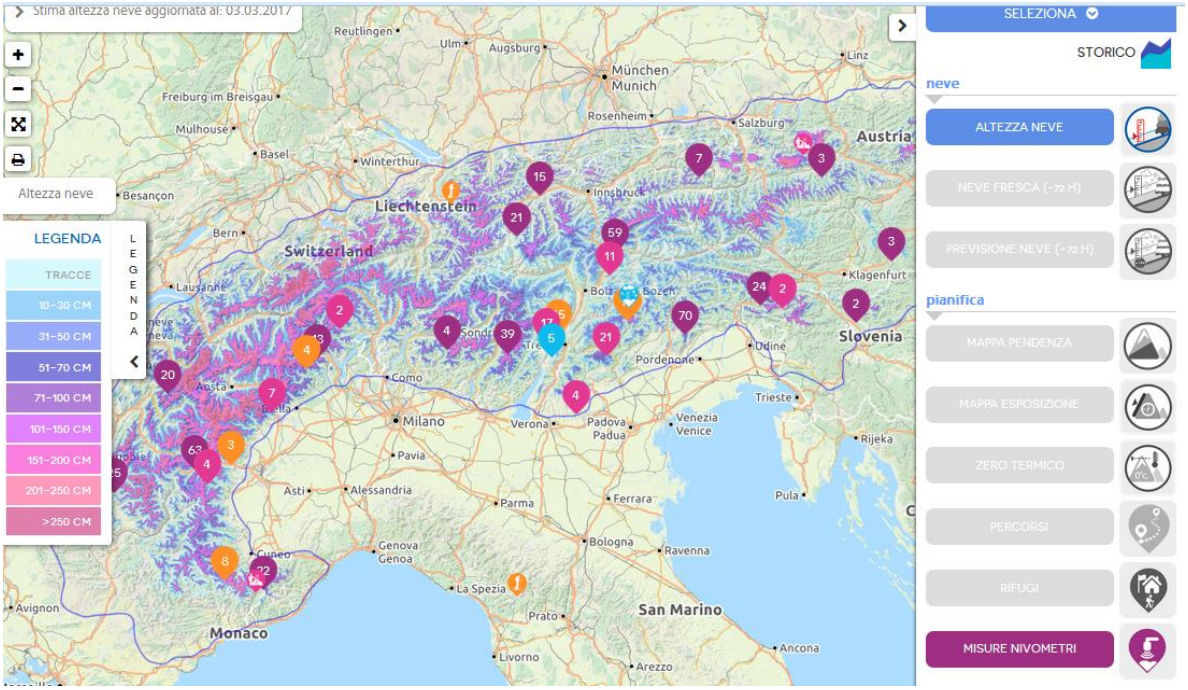
- Meteorological data, which is dynamic data that shows the evolution of the snow level. The meteorological forecasts are provided by external services and providers in map formats; every map has certain informatics content; for example, air temperature and relative

humidity correspond to 2 different maps. The data is fundamental for Mysnowmaps to deliver its product to tourists that can plan more efficiently their excursions.

Fig.6.11.Mysnowmaps: snow level and forecasts

Source: mysnowmaps.com

Fig.6.12.Mysnowmaps: map



Source: www.mysnowmaps.com

- Data provided by meteorological units, spread throughout the territory, which collect information directly on the field to identify the meteorological conditions. The meteorological units are placed by the public administration to deliver public services, such

as avalanches' reports or air quality reports; data is open, so it can be used for free by third parties. These units are very useful, but limited because data collected concerns only a particular portion of the territory. It becomes more useful if merged with the meteorological forecasts provided by other subjects, which is the operation that the algorithm of Mysnowmaps does.

From these 3 types of data the app is able to deliver a complete service on the conditions of the snow. Another source of data comes from users that measure directly on the field the snow's height and give the data to the application that registers and embodies it.

Another particular character of Mysnowmaps is that it is operative in all the three market segments: b2g, b2b and b2c. Simple consumers who are interested in snow or in planning their trips can use the platform to obtain information and forecasts. Companies that sell outdoor and sport equipment, like Montura, are interested in the platform users to be redirected to their products, so they put their logo on Mysnowmaps platform to gain some potential clients; thanks to such interest, a commercial relation is established with Mobygis. Public agencies are interested in different types of services that the company offers. Local tourism agencies want to promote their territory and use the same scheme as private companies: they commercialize their offer through the platform. Other public bodies, like "Protezione civile" are interested in monitoring the territory, so they demand information and forecasts that they not possess due to the lack of data they produce; the company is able to meet such demand thanks to the complete dataset and the competences that it has. The hydroelectric market needs also information about water volume that is forecasted to flow in water courses, information that is directly connected with snow level and for this reason can be delivered by Mobygis.

The activity of the company is almost unique in its sector. The development of a so rich and complete dataset of information on snow levels in the whole Alpine region is an operation that neither public nor private subjects have yet undertaken, except for Mobygis. Some research teams in Italy work in this field, but only with a scientific interest and not with the goal of developing an actual product. There are some initiatives concerning snow monitoring in Europe, but again, not with Mobygis' degree of coverage and completeness. The company could expand in other mountain regions, but the CEO Matteo Dall'Amico argues that certain contexts lack the basic conditions for a business to be sustainable. Referring to the possibility of expanding the business towards the "Appennini", he says "*My great worry is data collection, data accessibility, the creation of connections and relations with the bodies that own data in different regions like Molise, Marche, Abruzzo, ecc. Every contact means time and money, so I have to give priority to business. For Mobygis Appennini are interesting as the Pirenei, Scndinavia, or the Rocky Mountains, from an economic point of view"* (Dall'Amico, 2017). Such issue for the expansion of the business introduces a relevant theme for this type of data.

It is clear that a small company like Mobygis benefits greatly from the publication of data in open formats, because it allow it to save important resources that could endanger the whole business if

dedicated to the process of collecting data. Dall'Amico notes that in some cases the data owners say publicly that data is open, but finally information can be only visualized on their website and not reused without establishing contacts and commercial negotiations. This is a common confusion that exists because of a misinterpretation of the word "open". The company is forced to overcome or elude these barriers by offering a counterpart in exchange for data; such exchange is necessary not only with private companies, but also with public bodies. The counterpart consists in a simple money transfer, or in services of consultancy. The issue of collecting data from different sources is evident in the case of dealing with different countries, as Mobygis constantly does in its activity. Dall'Amico notes that every country has a different policy on meteorological data; as France and Switzerland have a central government body that collects all the data, Austria and Italy have sort of federalism, so every region collects data in its own way and with its own standards, which makes difficult for a company to deal with. Not only in terms of formats and data characteristics, but also licenses are different within the same country, so some agencies publish it in open formats and others not. Because Mobygis need to collect data from 14 different Alpine regions, this is very crucial issue for the company. Thanks to the know-how that Mobygis has obtained in scraping and homogenizing data, its work has been useful also for public administrations, both for regional and central ones. Homogenization has been important for aspects like the maps' legenda, which is often different depending on the source. In this sense the work of the company can really generate some value for the whole community, due to the work of standardization. Looking at a recent catastrophic event happened in central Italy, the value is even clearer: during the snow emergency of January 2017, 29 people died after a hotel in Rigopiano collapsed, due to the huge snowfall. The event was not promptly forecasted; consequently, the rescue services were not able to respond just in time to save the situation. The lack of an efficient forecast system can produce some serious consequences in these cases; a service like Mysnowmaps could really help the fragmentation of data collection to answer to emergencies in situations similar to Rigopiano, where a fast intervention is needed.

Concerning Open Data, Mobygis is obviously interested in the implementation of Open Data policies by countries. The company embodies 2 Open Data business models: the "aggregator" and the "developer". The lack of culture and competences for opening data within public administration is seen as obstacles in this sense, which has to be overcome to be able to respond to the private sector demand for information.

Although these are only four examples of Open Data reused for business, they give some interesting insights on how value is generated from information in Italy.

1. One aspect worth mentioning is that all the companies analysed embrace more than one Open Data business model: SpazioDati and Depp embrace four of them, MobyGis and Planetek two. This aspect shows that companies don't simply represent the information beneficiaries at the end of the data life cycle, but can have a role in different phases of the data value chain, and develop themselves useful publishing tools. The collaboration between the public sector and the private sector represents in this sense a value for both the actors.

2. Another interesting aspect is that the products developed by companies are in three cases Web platforms (Atoka, OpenCoesione, Mysnowmaps, ecc.), but the modalities of earning profits from them are different:
    - Atoka uses the freemium scheme that gives advanced features for the premium members
    - DEPP develops its platforms for no-profit, but won some tenders after the know.how that was able to gain before
    - Mysnowmaps uses private and public sponsors to make the platform economically sustainable
3. Value is often generated not only for the company, but also for other actors:
    - Citizens can express more informed political preferences, thanks to product like OpenParlamento and Open politici
    - Companies can more efficiently target their clients with the services offered by SpazioDati
    - The territory can be better managed and monitored thanks to the tools and services sold by Planetek, with benefits for all the community
    - Tourist have more information to plan their trips if they use the Mysnowmaps platform
4. Data accessibility and quality is mentioned by all companies as a concrete obstacle for their activity; this poses an important challenge for the public sector, that need to build a complete and valid Open Data framework for value to be generated.

Conclusions

This research has the ultimate goal of finding some relevant elements to enforce the idea that Open Data is a source for business. The Italian case studies presented show clearly how accessibility of data detained by public bodies is a relevant resource for the development of their products and services; although, future research should more deeply analyse which types of data can generate the most value, in order to priorities such types in the Open Data policies implemented by public bodies. Looking at the whole spectrum of the public data production, there is certainly some data that is not as useful for companies as others; nonetheless, it could represent an important resource for other subjects, such as citizens or the public sector itself, so it should not be left aside only because the private sector doesn't actually demand it.

One of the reasons that are cited to oppose the opening of data is that a free access cannot generate the revenues necessary for covering data production and collection costs. Such idea has been overcome in the earlier cited case of spatial data: the satellite missions "Landsat" experienced a change in data policy during the years, passing from a commercialization of data to an Open Data policy, due to the lack of revenues that were generated through the data selling channel; such change is taken as a relevant example because satellite data is probably the most expensive data to collect, which makes it difficult to make it sustainable from an economic point of view. This conclusion could be a really useful insight for policy makers.

Although, the creation of new opportunities for business should not give the idea that an entire new economy could be generated by the availability of public data; further analysis should focus on assessing the economic impact that Open Data policies are able to generate, but through significant quantitative estimates, which are necessary for ultimate cost/benefit analyses. Without clear evidence, an Open Data practice could be seen as a risky investment of resources by a public administration, especially in the Italian context where financial resources currently lack in the public sector. The creation of a relevant amount of jobs related to public data in the national economy has been estimated by the studies previously cited (MGI, 2013; European Union, 2015a), but such impact should be balanced with the aspect mentioned along different parts of this work: the digital economy that has started to emerge in the third millennium suffers of a high level of automatization, therefore a low level of human capital necessary in the production processes (Rifkin, 2015). Moreover, concerning the specific aspect of data standardization, some actors gain today an economic benefit from the lack of a unique standard for data collection, because of the work and activity that has to be dedicated to make data reusable by companies and the public sector itself. A context in which a certain data type is completely standardized at a national or supra-national level, due to a legal framework that makes it so, causes all the work spent for standardizing data completely unnecessary, with a subsequent loss of economic opportunities for such activities. It has still not be assessed if the resulting balance from these aspects could be an increased jobs demand or the contrary; in other words, the impact of Open Data in terms of jobs demand's growth in the medium-long term is not clear.

The focus of this work is mainly on the Italian context, which is described through data and insights on the ICT sector and the Open Data policy status at this time. Two key points are worth highlighting:

- The delay that data shows in terms of Internet spread, usage, an competences across Italian society, in comparison with the other European countries, could represents an obstacle for the creation of new opportunities derived from Open Data; for this reasons, and for other reasons linked to the social and economic opportunities that Internet is able to create, this is an issue that policy makers should address.

- Although some steps further have been taken by public administrations in the last years towards Open Data practices, the lack of clear and strong national Open Data commitment, policy and governance is blocking the development that could result from such actions. Both interviewed companies and experts of the field notice that without the coordination of the central government it is difficult to undertake innovative economic and social initiatives based on data, due to the fragmentation of data policies of local governments.

Sahunget and Sangokoya (2015) argue that the problem of assessing Open Data's impact can be described as the chicken and egg problem: "*National governments may have little incentive to release data until they see the economic benefits, but companies cannot demonstrate those benefits until they have access to Open Data*" (Gurin, 2014).

Bibliography

ACIL Tasman (2011-10-10), The Value of Spatial Information for Tasmania, http://www.acilallen.com.au/cms_files/ACIL_spatial%20information_tasmania.pdf

Assinform (2016), "Ict 2015: + 1% a 64,9 mld di euro Santoni: bene il segno positivo. Ora dobbiamo accelerare"

Alvino, V. (2017-02-15), *Interview with Vittorio Alvino*

Ayres, R.U. & Ayres, E.H. (2010), *Crossing the Energy Divide: Moving from Fossil Fuel Dependance to a Clean-Energy Future*, Upper Saddle River (NJ), Wharton School Publishing, p.14

Aware, U. B., Obama, M. (2008) 'Perils of the Internet', article available on peelschools.org

Alexander J. A. M. van Deursen, Ellen J. Helsper. (2016-02-13) The Third-Level Digital Divide: Who Benefits Most from Being Online? Emerald Insight. DOI: 10.1108/S2050-206020150000010002

Ban, K. (2016-09-19), "Press Release: Urgent Need to Address the Digital Divide if we are to achieve the Global Development..." http://www.itu.int/en/mediacentre/Pages/2016-PR37.aspx

Barbera, M. (2016-04-15), *Interview with Michele Barbera*

Barca, F. (2015) (*III incontro)FABRIZIO BARCA – Un nuovo management pubblico come leva per lo sviluppo*, https://www.youtube.com/watch?v=akqZysMR9U0

Baum, C. (2013-10-14), *So Who's Stealing China's Manifacturing Jobs?*, in "Bloomberg"

Berners-Lee (2013) "Open Data and a change of mindset is the next step in the Internet revolution", http://www.thedrum.com/news/2013/11/03/open-data-and-change-mindset-next-step-internet-revolution-says-web-founder-sir-tim

Beyer, M. A. & Laney, D. (2012) "The importance of Big Data: A definition". *Stamford, CT: Gartner*

Bizer, C., Heathm T. & Berners-Lee, T. (2009) "Linked Data - The Story So Far". International Journal on Semantic Web and Information Systems, Vol. 5(3), Pages 1-22. DOI: 10.4018/jswis.2009081901

Bresnahan, T. F. & M. Trajtenberg (1995), General purpose technologies 'Engines of growth'? Journal of Econometrics 65 (1), 83106

Brynjolfsson, £. & Saunders, A. (2010), "How Information Technology Is Reshaping the Economy", *Wired for Innovation*

Buchanan J.M. (1965): The Demand and Supply of Public Goods, Rand McNally, Chicago

Capgemini Consulting (2013), *The Open Data Economy. Unlocking Economic Value by Opening Government and Public Data*

Cassano, J. (2014-10-22) "Inside the Crowdsourced Map Project That Is Helping Contain the Ebola Epidemic." Fast Company. https://www.fastcompany.com/3037350/inside-the-crowdsourced-map-project-that-is-helping-contain-the-ebola-epidemic

Cave, D. (2014-09-01) "Billboard Drives Home Extent of Corruption as Schools Suffer." The New York Times. https://www.nytimes.com/2014/09/02/world/americas/billboard-drives-home-extent-of-corruption-as-schools-suffer.html?_r=0

CDC, "Ebola (Ebola Virus Disease)−Outbreaks." Centers for Disease Control and Prevention. http://www.cdc.gov/vhf/ebola/outbreaks/

CHANGE.GOV (2012a), "Agenda: Ethics", http://change.gov/agenda/ethics_agenda (last visited June 8, 2012)

CHANGE.GOV (2012b), See Policy Working Groups. http://change.gov/learn/policy_working_groups (last visited June 8, 2012).

Chatain, O. (2011) "Value Creation, Competition, and Performance in Buyer-supplier Relationships," Strategic Management Journal, vol. 32(1), p. 76-102

Dall'Amico, M. (2017-02-27), *Interview with Matteo Dell'Amico*

Davies, T. (2010-08), *Open Data, democracy and public sector reform*

De Chiara, F. (2016), *Mapping the reusers of Open Data in Italy- Early findings of Open Data 200 Italy study*

Deligianni, H., Ahmed, S. & Romankiw, L. (2011), *The next frontier: Electrodeposition for Solar Cell Fabrication*, in "Electrochemical Society"

Deloitte (2012), *Open growth; Stimulating demand for Open Data in the UK*

Department of State (1976), "Remote sensing : Acquisition of satellite data : #b memorandum of understanding between the United States of America and Iran signed at Washington and Tehran July 25 and October 29, 1974", Volume 8203 di Treaties and other international acts series

Di Girolamo, R. (2014-01-16), *Gli oceani di dati e il sesso degli adolescenti*, http://voices.telecomitalia.com/2014/01/gli-oceani-di-dati-e-il-sesso-degli-adolescenti/

Douglas, L. (2001) "3d data management: Controlling data volume, velocity and variety". *Gartner*.

Dufour, P. (2015-12-17), GovLab interview with Patrick Dufour, Former Senior Web GIS Developer, U.S. Department of State, Humanitarian Information Unit

Dutton, William H. (2013-06-20), "The Internet of Things". Available at SSRN: https://ssrn.com/abstract=2324902 or http://dx.doi.org/10.2139/ssrn.2324902

European Union (2015a), *Creating Value through Open Data. Study on the impact of Re-use of Public Data Resources*

European Union (2015b), *Open Data Maturity in Europe 2015*

European Union (2016), *Open Data Maturity in Europe 2016 https://www.capgemini-consulting.com/resource-file-access/resource/pdf/open_data_maturity_in_europe_2016_final_v1.0_0.pdf*

European Court of Human Rights (1950-10-04), *European Convention of Human Rights and Fundamental Freedoms – article 8*, Rome

Executive Office of the President of the United States (2009-12-08), *Memorandum for the heads of the Executive Departments and Agencies – Open Government Directive*

Frey, C.B. & Osborne, M.A. (2013-09-17), *The future of employment: how susceptible are jobs to computerisation?*
http://www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf

Gandhi, Mahatma (1936-05-16), *Mahatma Gandhi on mass production (interview)*

Gartner (2014-03-19). Gartner says the Internet of Things will transform the data center.

Géczy, P. (2015), "Data Economy Dimensions*"*, *Global Journal of Business Research*, Vol. 9, No. 4, 2015, pp. 77-89, ISSN: 2157-0191

Gurin, J. (2014), *Open Data Now*, McGraw-Hill Education

Habermas, J. (1962), *The Structural Transformation of the Public Sphere*

Hasan, M. (2013-10) "Mexico: Encouraging Parents' Engagement in Public Education Reform." Open Government Partnership. http://www.opengovpartnership.org/sites/default/files/Mexico_0.pdf

Heinrich, E. (2014-06-23) *Telecom companies count $386 billion in lost revenue to Skype, WhatsApp, others*, http://fortune.com/2014/06/23/telecom-companies-count-386-billion-in-lost-revenue-to-skype-whatsapp-others/

Hoda,, S. (1997), *Gandhi and the contemporary world*, Indo-British Historical Society

Iacono, N. (2016-02-12), *Open Data, la strategia Paese per un mercato da 8 miliardi di euro, http://www.forumpa.it/pa-digitale/open-data-linee-di-azione-per-la-crescita?utm_source=newsletter&utm_medium=DATA&utm_campaign=MAILUP*

IDC (2012-03), *Worldwide Big Data Technology and Services 2012– 2015 Forecast*

Iemma, R. (2013), *L'informazione del settore pubblico: elementi per un'analisi economica*

IFLA (2013), *2013 IFLA Trend Report*, http://trends.ifla.org/expert-meeting-summary/the-future-of-online-privacy-and-the-security-of-personal-data

Il Sole14Ore (2017-01-31), *Istat: la disoccupazione dei giovani risale al 40,1% http://www.ilsole24ore.com/art/notizie/2017-01-31/istat-disoccupazione-giovani-risale-401percento-100915.shtml?uuid=AEVOdGL*

Internet Society (2015), Global Internet Report 2015,
http://www.internetsociety.org/globalinternetreport/2015/

Ireg (2010-10), *Global Higher Education Rankings 2010*, http://www.ireg-observatory.org/pdf/HESA_Global_Higher_EducationRankings2010.pdf

IRM (2015), *Italia – Rapporto sui progressi 2014-2015*

ISTAT (2015-12-12), *Cittadini, imprese e ICT*,

Joseph, P., Meadows, R. & Fresco, J. (2009-02), *The Zeitgeis Movement: Observations and Responses*, in "Zeitgeist Movement"

Kallas, P. (2016-10-14), *Top 10 Social Networking Sites by Market Share of Visits [August 2016]*, http://www.dreamgrow.com/top-10-social-networking-sites-by-market-share-of-visits-august-2016/

Kaplan, A. M. & Haenlein, M. (2016). "Higher education and the digital revolution: About MOOCs, SPOCs, social media, and the Cookie Monster". Business Horizons. 59 (4): 441–50. doi:10.1016/j.bushor.2016.03.008

Kosky, H. (2011-09-28), *Does Marginal Cost Pricing of Public Sector Information Spur Firm Growth?*, Etla Discussion Papers n.1260, https://www.etla.fi/wp-content/uploads/2012/09/dp1260.pdf

Lanier, J. (2013), *Who owns the future?*, UK, Simon and Schuster

Lee, I. & Lee, K. (2015), "The Internet of Things (IoT): Applications, investments, and challenges for enterprises", *Business Horizons*, vol. 58, issue 4, pages 431-440. Available online at www.sciencedirect.com

Maxmen, A. (2015-01-27) "How Ebola Found Fertile Ground in Sierra Leone's Chaotic Capital." National Geographic. http://news.nationalgeographic.com/news/2015/01/150127-ebola-virus-outbreak-epidemic-sierra-leone-freetown-photos-pictures/

MGI (2013-10), *Open Data: Unlocking innovation and performance with liquid information*, http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information

MGI (2011-06), "Big Data: The next frontier for innovation, competition and productivity", McKinsey & Company www.mckinsey.com/~/media/McKinsey/dotcom/Insights%20and%20pubs/MGI/Research/Technology%20and%20Innovation/Big%20Data/MGI_big_data_full_report.ashx.

Minister of State for the Cabinet Office (2012-06), *Open Data White paper*

Ministro per la Semplificazione e la Pubblica Amministrazione (2016), *Open Government in Italy – Third Action Plan 2016-2018*

Mochi Sismondi, C. (2016-10-13), *Il decalogo dell'innovazione che viene dalle città e dalle comunità locali*, http://www.forumpa.it/pa-digitale/linnovazione-che-viene-dalle-citta-e-dalle-comunita-locali

Montiel (2015-08-19), GovLab Interview with Oscar Montiel, Codeando México..

Moore, Gordon E. (1965-04-19). "Cramming more components onto integrated circuits". Electronics.

Naam, R. (2011-03-16), "Smaller, Cheaper, Faster: Does Moore's Law Apply to Solar Cells?", in "Scientific America"

Napolitano, M (2013-12-16), *OpenStreetMap - il potere delle mappe condivise: Maurizio Napolitano at TEDxTrento*, https://www.youtube.com/watch?v=9eYVH1tZxjo

Napolitano, M. (2016-12-23), *Open Data: senza trasparenza e sviluppo economico aprire i dati serve a poco*, http://www.forumpa.it/pa-digitale/open-data-senza-trasparenza-e-sviluppo-economico-aprire-i-dati-serve-a-poco

Nest Labs (2015), "Energy Savings from the Nest Learning Thermostat: Energy Bill Analysis Results

NOAA (2011-10-13). "Value of a Weather Ready Nation." http://www.ppi.noaa.gov/ wp-content/uploads/PPI-Weather-Econ-Stats-10-13-11.pdf

NOAA (2013-03-18). "The Economic Value of Resilient Coastal Communities." http://www.ppi.noaa.gov/wp-content/uploads/EconomicValueofResilientCoastalCommunities.pdf

NOAA Magazine (2004-12-01), "NOAA Issues New Partnership Policy: Response to 'Fair Weather' Recommendations Strengthens Relationships Among Government, Universities and the Private Sector." NOAA Magazine. http://www.noaanews.noaa.gov/stories2004/s2348.htm

NOAA Magazine (2004-12-01), "NOAA Issues New Partnership Policy: Response to 'Fair Weather' Recommendations Strengthens Relationships Among Government, Universities and the Private Sector." NOAA Magazine. http://www.noaanews.noaa.gov/stories2004/s2348.htm

Nonnecke, Brandie M; Bruch, Mia; & Crittenden, Camille (2016). IoT & Sustainability: Practice, Policy and Promise. CITRIS. UC Berkeley: Center for Information Technology Research in the Interest of Society (CITRIS). Retrieved from: http://escholarship.org/uc/item/7dp1t4p8"

Obama, B.H. (2009-01-21), *Memorandum on Transparency and Open Government*, https://www.archives.gov/files/cui/documents/2009-WH-memo-on-transparency-and-open-government.pdf

OECD (2008), "The Seoul Declaration for the Future of the Internet Economy, Ministerial session", 18 June 2008, Available at: http://www.oecd.org/internet/consumerpolicy/40839436.pdf

OECD & NAS (2008), "Assessing the Economic and Social Benefits of NOAA data", https://www.oecd.org/sti/ieconomy/40066192.pdf

OECD (2013a), "Measuring the Internet Economy: A Contribution to the Research Agenda", OECD Digital Economy Papers, No. 226, OECD Publishing. http://dx.doi.org/10.1787/5k43gjg6r8jf-en

OECD (2013b), "Exploring Data-Driven Innovation as a New Source of Growth: Mapping the Policy Issues Raised by "Big Data"", OECD Digital Economy Papers, No. 222, OECD Publishing. http://dx.doi.org/10.1787/5k47zw3fcp43-en

Oxera (2013-01), *What is the economic impact of Geo-services?*, Summary report prepared fo Google, http://www.oxera.com/Oxera/media/Oxera/downloads/reports/What-is-the-economic-impact-of-Geo-services---summary_2.pdf

Oxfam (2016-09-09), *A Europe for the many, not the few*, 206 Oxfam briefing paper https://www.oxfam.org/sites/www.oxfam.org/files/file_attachments/bp206-europe-for-many-not-few-090915-en.pdf

Perry, M.J. (2012-11-06), "The US Economy is now Producing 2.2% More Output than before the Recession, but with 3.84 Million Fewer Workers", in "American Enterprise Institute"

Pîrjan, A. & Petroşanu, D.M. (2013), "The Impact Of 3d Printing Technology On The Society And Economy," Romanian Economic Business Review, Romanian-American University, vol. 7(2), pages 360-370, December.

Piirsalu, K.-L. (2012), *Privacy issues for social networks – introduction*, http://social-networks-privacy.wikidot.com/

Presidenza del Consiglio dei Ministri (2014), Piano crescita digitale 2014-2020, Roma, pg. 70

Press, G. (2014-09-03), *12 Big Data Definitions: What's Yours?*, Forbes, https://www.forbes.com/sites/gilpress/2014/09/03/12-big-data-definitions-whats-yours/#1e54d6db13ae

Rayna, T. (2008), "Understanding the Challenges of the Digital Economy: The Nature of Digital Goods," Communications & Strategies, IDATE, Com&Strat dept., vol. 1(71), pages 13-36, 3rd quart

REN21 (2016), *Renewable 2016 Global Status Report*

Rifkin, J. (2011), *The Third Industrial Revolution*, Griffin

Rifkin, J. (2014), *La Società a Costo Marginale Zero*, Ristampa 7, Milano, Arnoldo Mondadori Editore

Robinson, D. at al. (2009), "Government Data and the Invisible Hand", 11 J.L. & Tech.160, available at: http://ssrn.com/abstract=1138083

Rodriguez_Pose, A. & Tselios, V. (2008), *Education and Income Inequality in the Regions of the European Union*, SERC discussion paper 11

Sahuguet A. & Sangokoya D. (2015), *A "calculus" for Open Data*, available at https://medium.com/@sahuguet/p-b-d-c-1218ee894400

Samuelson, P. (1954). "The Pure Theory of Public Expenditure". The Review of Economics and Statistics, 36(4), 387-389. doi:10.2307/1925895

Saporiti, R. (2016-10-31), *Cresce l'Europa delle tecnologie. Italia in ultima posizione*, http://www.infodata.ilsole24ore.com/2016/10/31/leuropa-dellict/?refresh_ce=1

Shakespeare, S. (2013-05), *Shakespeare Review; An Independent Review of Public Sector Information*, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/198752/13-744-shakespeare-review-of-public-sector-information.pdf

Slater, G. (1973), *The English Peasantry and the Enclosure of the Commons*, New York, A.M. Kelley

Simonite, T. (2016-05-13), *Moore's Law Is Dead. Now What?*, https://www.technologyreview.com/s/601441/moores-law-is-dead-now-what/

Sniderman, D. (2012-05), *3D Print Takes Off*, https://www.asme.org/engineering-topics/articles/design/3d-printing-takes-off

Solow, R.M. (1963), *Capital Theory and the Rate of Return*, Amsterdam, North-Holland Pub.

Spooner, S. (2014-09-21), "Ebola could kill 100,000 and needs $1 billion to defeat, but Nigeria shows the world how it's done." Mail & Guardian Africa. http://mgafrica.com/article/2014-09-19-as-the-war-against-ebola-rages-nigeria-showsthe-world-how-its-done

StartupItalia! (2016-06-10), *Le 3 startup che hanno vinto Edison Pulse e portano a casa 65K ciascuna*, http://startupitalia.eu/58194-20160610-startup-vincitrici-edison-pulse

Strauss, S. (2012-01-02), *The Connection between Education, Income, Inequality and Unemployment* http://www.huffingtonpost.com/steven-strauss/the-connection-between-ed_b_1066401.html

Sunshot Vison Study (2012-02), in "US Department of Energy"

Teran, J. (2015-09-18), GovLab interview with Javier Teran, Statistician, United Nations Office for the Coordination of Humanitarian Affairs, Humanitarian Data Exchange

Theel, S. (2013-01-24), *Myths and Facts about Solar Energy*, https://mediamatters.org/research/2013/01/24/myths-and-facts-about-solar-energy/192364#cost

The Economist (2013-05-18) *Open Data. A new goldmine*, http://www.economist.com/news/business/21578084-making-official-data-public-could-spur-lots-innovation-new-goldmine

The Open Data Institute of North Carolina (2014-09-19), *Tim Berners-Lee's 5-Star Open Data Scheme*, https://www.slideshare.net/TheODINC/tim-bernerslees-5star-open-data-scheme

Transparency International (2009), 'Corruption in the Education Sector', Working Paper no. 04/2009 (Berlin: TI, 2009), p. 6

Ubaldi, B. (2013), "Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives", OECD Working Papers on Public Governance, No. 22, OECD Publishing. http://dx.doi.org/10.1787/5k46bj4f03s7-en

UNEP & Frankfurt School of Finance & Management (2016), *Global Trends in Renewable Energy Investment 2016*, *http://fs-unep-centre.org/sites/default/files/publications/globaltrendsinrenewableenergyinvestment2016lowres_0.pdf*

United Nations (2014), United Nations Security Council Resolution 2177 (2014). http://www.ifrc.org/docs/IDRL/UN%20SC%20Res.pdf

United Nations Framework Convention on Climate Change (UNFCCC) (2015-12-12), *2015 United Nations Climate Change Conference, COP 21 (Paris Agreement),* Paris.

Vale, S. (2013-01-27), *Classification of types of Big Data*, http://www1.unece.org/stat/platform/display/bigdata/Classification+of+Types+of+Big+Data

Verhulst, S. & Caplan, R. (2015-04), *Open Data – a twenty-first-century asset for small and medium-sized enterprises*, GovLab

Vickery, G. (2011-09-16), *Review of recent studies on PSI re-use and related market developments*. Editor Connect, https://ec.europa.eu/digital-single-market/en/news/review-recent-studies-psi-reuse-and-related-market-developments

Villagran, L. (2013-03-01) "Test the teacher? Educators balk at Mexico's reforms." The Christian Science Monitor. http://www.csmonitor.com/World/Americas/2013/0301/Test-the-teacher-Educators-balk-at-Mexico-s-reforms

Warren, J.(2001-05-15) http://www.svipx.com/pcc/PCCminipages/z2854bc4b.html

W3C (2002-2004), *Architecture of the World Wide Web, Volume One*, https://www.w3.org/TR/webarch/#acks

W3C (2004), "Lead Your Organization to its Full Potential. Join the World Wide Web Consortium", https://www.w3.org/2004/04/w3c-flier-v1.6.3A4.pdf

Wikibon (2016), *2016 – 2026 Worldwide Big Data Market Forecast*, accessed from http://siliconangle.com/blog/2016/03/30/wikibon-forecasts-big-data-market-to-hit-92-2bn-by-2026/

Wood, A. (31/03/2015) "The internet of things is revolutionising our lives, but standards are a must", *The Guardian*, https://www.theguardian.com/media-network/2015/mar/31/the-internet-of-things-is-revolutionising-our-lives-but-standards-are-a-must

World Bank (2014), *Open Data for Economic Growth*, Washington, DC https://openknowledge.worldbank.org/handle/10986/19997

Worthy, B. (2014-09-17), *Making Transparency Stick: The Complex Dynamics of Open Data*, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2497659

Wu, T. and Yoo, C. (2007) "Keeping the Internet Neutral?: Tim Wu and Christopher Yoo Debate," Federal Communications Law Journal: Vol. 59 : Iss. 3, Article 6. Available at: http://www.repository.law.indiana.edu/fclj/vol59/iss3/64

Young, A. & Verhulst, S. (2016-09-13), *The Global Impact of Open Data*, O'Reilly Media, Sebastopol

Yu, H. & Robinson, D.G. (2012), "The New Ambiguity of Open Government", UCLA L. REV. DISC. 178

Zabludovsky, K. (2014-05-30) "Data Dive Finds 100-Year-Old Teachers and Phantom Schools in Mexico." Newsweek, http://www.newsweek.com/data-dive-finds-hundredyear-old-teachers-and-phantom-schools-mexico-252871

Zapata Hojel, A. (2015-08-07) GovLab Interview with Alexandra Zapata Hojel, Coordinator of Education Projects, IMCO

Zappa, F. (1979) "Packard Goose", Joe's Garage Acts II & III

Zotti, M. (2017-02-23), *Interview with Massimo Zotti*

Web references

http://data.worldbank.org/indicator/IT.NET.USER.P2

http://index.okfn.org/

http://it-city.census.okfn.org/

http://sierraleone.opendataforafrica.org/

http://livingeconomics.org/article.asp?docId=239

http://network.ot11ot2.it/

http://opendatabarometer.org/

http://openpolis.it/

http://open.gov.it/

http://politici.openpolis.it/

http://www.foia4italy.it/

http://www.ibm.com/us-en/

http://www.infogineering.net/data-information-knowledge.html

http://www.internetlivestats.com/internet-users

http://www.mejoratuescuela.org/quienes-somos

http://www.mysnowmaps.com/it/

http://www.netmarketshare.com/

http://www.oecd.org/development/effectiveness/busanpartnership.htm

http://www.opendata500.com/

http://www.opendata500.com/it/

http://www.thegovlab.org/

http://www.tradingeconomics.com/united-states/gdp-growth-annual

https://creativecommons.org/licenses/by/3.0/it/legalcode

https://dandelion.eu/

https://en.oxforddictionaries.com

https://openaid.se/sv/aid/2016/

https://www.foia.gov/index.html

https://www.merriam-webster.com

https://www.opengovpartnership.org/about/open-government-declaration

https://www.statista.com/

https://www.techopedia.com/definition/23371/digital-revolution

Appendix 1 Open Data 500 survey

*Fields marked with * are required*

1. Contact Information

First Name: *

Last Name: *

Title: *

Email: *

Phone:

2. Company Information

Company Name: *

Company URL: *

In which city is your company's headquarters located?

State:*

Country: United Stat

Zip Code: *

Year Founded: *

3. Type of Company: *

- Public
- Private
- Nonprofit
- Other

4. Number of full-time employees: *

- 1-10
- 11-50
- 51-200

- ○ 201-500
- ○ 501-1,000
- ○ 1,001-5,000
- ○ 5,001-10,000
- ○ 10,001+

5. Business Model
(check all that apply) *

- ☐ Business to Business
- ☐ Business to Consumer
- ☐ Business to Government
- ☐ Other

6. Which of the following are significant sources of revenue for your company?
(check all that apply) *

- ☐ Advertising
- ☐ Consulting
- ☐ Contributions/Donations
- ☐ Data analysis for clients
- ☐ Database licensing
- ☐ Government contract
- ☐ Lead generation to other businesses
- ☐ Membership fees
- ☐ Philanthropic grants
- ☐ Software licensing
- ☐ Subscriptions
- ☐ User fees for web or mobile access
- ☐ Other

7. Does your company's mission have a direct, beneficial social impact in any of the following areas?
(check all that apply)

- [ ] Citizen engagement and participation
- [ ] Consumer empowerment
- [ ] Educational opportunity
- [ ] Environment and climate change
- [ ] Financial access
- [ ] Food access and supply
- [ ] Good governance
- [ ] Healthcare access
- [ ] Housing access
- [ ] Public safety
- [ ] Other

8. What category best describes your company? *

- ○ Business & Legal Services
- ○ Data/Technology
- ○ Education
- ○ Energy
- ○ Environment & Weather
- ○ Finance & Investment
- ○ Food & Agriculture
- ○ Geospatial/Mapping
- ○ Governance
- ○ Healthcare
- ○ Housing/Real Estate
- ○ Insurance
- ○ Lifestyle & Consumer
- ○ Media
- ○ Research & Consulting
- ○ Scientific Research

- Transportation
- Other

9. Please give us a short public statement describing your company's mission and work. You can take this material from your website or other publications if you choose to.
***[200 words or less]* ***

10. As a summary, please provide a one sentence description of your company.
***[25 words or less]* ***

11. Please include any financial or operational information that will help us understand your company. We are interested in specific information like past and projected annual revenues, total outside investment dollars to date, and significant investors or partners.
[100 words or less*]*

12. Provide any specific examples of how your company uses Open Data.
[200 words or less]

13. From approximately how many sources (i.e. agencies, subagencies, local and state governments) does your company use data?

- *1-10*
- *11-50*
- *51-100*
- *101+*

14. What types of data does your company use?
(check all that apply)

- Agriculture & Food
- Business
- Consumer
- Demographics & Social
- Economics
- Education
- Energy
- Environment
- Finance
- Geospatial/Mapping

- ☐ Government Operations
- ☐ Health/Healthcare
- ☐ Housing
- ☐ International/Global Development
- ☐ Legal
- ☐ Manufacturing
- ☐ Science and Research
- ☐ Public Safety
- ☐ Tourism
- ☐ Transportation
- ☐ Weather
- ☐ Other

15. What impact does Open Data use have on your company? *(check all that apply)

- ☐ Cost efficiency
- ☐ New or improved product/service
- ☐ Job growth
- ☐ Revenue growth
- ☐ Identify new opportunities
- ☐ New/improved research
- ☐ Other

16. Please tell us more about the Open Data sources your company uses.

Use the search bar to find and select government agencies, subagencies, city and state governments, or government-funded research programs from the list provided. If you can't find your data provider on the list, please fill out the form below to suggest an agency, subagency, organization, etc**.**

[Optional]: Please tell us which datasets your company uses from the agencies, city, or state governments you select, and please score each dataset on a scale of 1 to 4 (i.e. 1 = Poor; 4 = Excellent

17. Please give your comments about the usefulness of all these data sources and datasets. *[250 words or less]*

Appendix 2 Open Data 200 survey

Informazioni sull'azienda

1 []
Nome dell'azienda che rappresenti:

Scrivere la propria risposta qui :

2 []
Anno di costituzione:
*
La risposta deve essere compresa tra 1000 e 2050
Solo un valore intero può essere ammesso i n questo campo.
Scrivere l a propri a risposta qui :

scrivere l 'anno intero. Es. 1999

3 []Tipo di azienda/organizzazione: *
Scegli solo una delle seguenti :
Pubblica (comprese società a capitale pubblico/misto)
Privata
No profit
Altro

4 []Quanti dipendenti ha la tua azienda? *
Scegli solo una delle seguenti :
meno di 10 dipendenti
10 - 50 dipendenti
51 - 250 dipendenti
251 - 1000 dipendenti
Più di 1000 dipendenti

5 []
A quanto ammonta l'ultimo fatturato annuo della tua azienda?
*
Scegli solo una delle seguenti :
Meno di 100mila euro
Da 100mila a 250mila euro
Da 251mila a 500mila euro
Da 501mila a 1 milione di euro
Da 1 milione a 3 milioni di euro
Oltre 3 milioni di euro

Maggiori dettagli sull'azienda

6 []
Il modello di business della tua azienda è rivolto a:
***
Scegliere tutte le corrispondenti :
pubblica amministrazione (business to government)
aziende (business to business)
utenti finali (business to consumer)
altro:
Seleziona uno o più model l i .

7 []Quale di questi settori descrive meglio la tua azienda?
Tutte le risposte devono essere diverse e ordinate in classifica.

Selezionare da 1 a 2 risposte
Numerare ciascun campo in ordine di preferenza da 1 a 19
Scegliere almeno 1 voci .
Non selezionare più di 2 voci .

Agricoltura e settore alimentare

Ambiente e clima

Assicurazioni

Energia

Finanza e Investimenti

Governo e pubblica amministrazione

Immobiliare

Istruzione

Mappe / Geospaziale

Media e comunicazione

Ricerca e consulenza

Ricerca scientifica

Sanità

Servizi legali e aziendali

Software

Tecnologia e dati

Trasporti

Turismo e cultura

ALTRO (non presente in questa lista)

8 []
Nella precedente risposta hai scelto "ALTRO (non presente in questa lista)", t i
chiediamo di specificare qui la voce.

*
Rispondere solo se le seguenti condizioni sono rispettate:
-------- Scenari o 0 --------
La risposta era 'ALTRO (non presente in questa lista)' Alla domanda '7 [G1Q00007]' (Quale di questi settori descrive meglio l a tua azienda? (CLASSIFICA 1))
-------- o Scenari o 1 --------
La risposta era 'ALTRO (non presente in questa lista)' Alla domanda '7 [G1Q00007]' (Quale di questi settori descrive meglio l a tua azienda? (CLASSIFICA 2))

Scrivere l a propri a risposta qui :

9 []
La tua azienda conduce attività di business anche all'estero?
*
Scegli solo una delle seguenti :
Sì

No

10 []
La tua azienda è quotata in borsa?
*
Scegli solo una delle seguenti :
Sì
No

11 []
Potresti descrivere brevemente la missione dell'azienda?
Scrivere l a propri a risposta qui :

Uso di Open Data

12 []
Scegli la definizione di Open Data che ritieni più corretta
***
Scegli solo una delle seguenti :
dati pubblicati sul web
dati a cui ognuno può accedere, liberamente scaricabili e riusabili
dati che non possono essere riutilizzati a fini commerciali

13 []
Quanto essenziali sono gli Open Data nella vostra attività di business?
***
Scegli solo una delle seguenti :
stiamo esplorando
gli Open Data non hanno un ruolo importante nel nostro business
gli Open Data sono importanti per il nostro business
gli Open Data sono il nostro modello di business

14 []
In che modo l'azienda usa gli Open Data?
***
Scegliere tutte le corrispondenti :
La mia azienda pubblica Open Data
La mia azienda fornisce ad altri l'infrastruttura per pubblicare Open Data (piattaforme, portali, data stores)
La mia azienda elabora Open Data (aggregazione, classificazione, anonimizzazione, pulizia, integrazione,arricchimento)
La mia azienda produce analitiche e visualizzazioni sulla base di Open Data
La mia azienda sviluppa prodotti sulla base di Open Data (APIs, apps)
Altro:

15 []
Quali prodotti/servizi, proposti dalla tua azienda, fanno uso di Open Data? *

Il commento è permesso solo quando l'opzione relativa è stata scelta.
Scegliere tutte quel l e che corrispondono e inserire un commento:

Analisi e supporto alle decisioni strategiche

Piattaforme e strumenti per la visualizzazione dei dati

Servizi di consulenza

Servizi di rating e analisi predittive

Software e sviluppo di applicazioni web e mobile

Servizi per la formazione

Altro:

Aggi ungi dettagli accanto all ‘opzione o più opzioni prescelte.

16 []
Che tipologia di dati utilizzate?
*

Scegliere tutte le corrispondenti :
dati cartografici/mappe (es. dati per la produzione di mappe)
dati di trasporto (es. orari degli autobus)
dati sul commercio internazionale (es. dati su import_export)
dati su sanità e salute (es. apparecchiature, utilizzo dei farmaci, statistiche sanitarie)
dati su istruzione (es. scuola in chiaro)
dati ambientali (es. qualità dell' aria, qualità dell' acqua)
dati del censimento (es. statistiche nazionali censimento Istat)
dati legislativi e sulle attività parlamentari (es. Normattiva, dati camerali)
dati elettorali (es. risultati elettorali, tasso di partecipazione)
dati sui bilanci pubblici
dati sulla spesa pubblica
dati sulle imprese
dati su turismo e cultura (es.strutture alberghiere, eventi)
Altro:

Maggiori dettagli su uso di Open Data

17 []
Da quali portali, cataloghi di dati aperti hai scaricato Open Data? *

Scegliere tutte le corrispondenti :
Nazionali
Regionali
Municipali
Cataloghi di dati non italiani (ad esempio EU Unione Europea o di altri paesi)

18 []Di quale di questi cataloghi nazionali fate uso in azienda? *

Rispondere solo se le seguenti condizioni sono rispettate:
La risposta era 'Nazionali ' Alla domanda '17 [G2Q00006]' (Da qual
i portali , cataloghi di dati aperti hai scaricato open

data?)
Scegliere tutte le corrispondenti :
Portale nazionale (dati.gov.it)
Camera dei Deputati
Senato
Consiglio Nazionale delle Ricerche
Funzione Pubblica
INAIL
INPS
ISTAT
Ministero Affari Esteri
Ministero Beni e Attività culturali e Turismo
Ministero Economia e Finanze
Ministero Salute
Ministero Sviluppo Economico
OpenCoesione
Soldi Pubblici
Italia Sicura
OpenEXPO
Altro:

19 []Di quale di questi cataloghi regionali/provinciali fate uso in azienda? *

Rispondere solo se le seguenti condizioni sono rispettate:
La risposta era Alla domanda '17 [G2Q00006]' (Da qual
i portali , cataloghi di dati aperti hai scaricato Open Data?)
Scegliere tutte le corrispondenti :

Provincia Autonoma di Bolzano
Provincia Autonoma di Trento
Regione Abruzzo
Regione Basilicata
Regione Calabria
Regione Campania
Regione Emilia-Romagna
Regione Autonoma Friuli-Venezia-Giulia
Regione Lazio
Regione Liguria
Regione Lombardia
Regione Marche
Regione Molise
Regione Piemonte
Regione Puglia
Regione Autonoma Sardegna
Regione Autonoma Sicilia
Regione Toscana
Regione Umbria
Regione Autonoma Val d' Aosta
Altro:

20 []Da quali di questi cataloghi Open Data municipali fate uso in azienda? *

Rispondere solo se le seguenti condizioni sono rispettate:
La risposta era Al a domanda '17 [G2Q00006]' (Da qual
i portali , cataloghi di dati aperti hai scaricato Open Data?)
Scegli ere tutte le corrispondenti :
Alessandria
Bari

Bergamo
Bologna
Bolzano
Cagliari
Carbonia-Iglesias
Catania
Catanzaro
Como
Cosenza
Ferrara
Firenze

Foggia

Forlì-Cesena

Genova
Lecce
Matera
Milano
Modena
Napoli
Palermo
Perugia
Ravenna
Reggio Emilia
Rimini
Roma
Siena
Torino
Trento
Trieste
Udine
Venezia
Verona
Vicenza
Altro:

21 []
In che misura questi fattori hanno influenzato la decisione di usare Open Data nella tua attività?

1 (nessuna influenza 5 (massima influenza)
*
Scegliere la risposta appropriata per ciascun elemento:

	1	2	3	4	5
Provenienza del dato					
Licenza					
Accuratezza					
Tempestività					
Formato					
Documentazione di accompagnamento					
Gratuità					

Questa lista si basa e integra la checklist sviluppata da Open Data Institute per gli Open Data certificates.

22 []

Puoi indicarci quali ulteriori fonti di Open Data utilizzi? Se stai usando Open Data,per esempio dati di comunità (es. Wikipedia, OpenStreetMap, DBpedia, etc.):

Scrivere la propri a risposta qui :

Specifica qui la fonte:

23 []

La tua azienda investirà in Open Data nei prossimi 12 mesi?

*****

Scegli solo una delle seguenti :

Sì

No

24 []

Quali sono i fattori critici per abilitare un riutilizzo più esteso degli Open Data in Italia?

*****

Scegliere tutte l e corrispondenti :

Legislazione, licenze

Supporto politico

Chiarezza nel processo di pubblicazione dei dati

Gestione e supporto nel processo di pubblicazione dei dati nelle organizzazioni

Qualità dei dati

Metriche e indicatori come parte integrante delle iniziative Open Data

Sostenibilità

Identificazione della domanda reale di dati

Mantenimento dei dati pubblicati

Accessibilità, interoperabilità e standard

Altro:

25 []

Suggerisci in breve quali dati vorresti vedere aperti e cosa fare per migliorare l'attuale situazione dell'Open Data in Italia.

Scrivere l a propri a risposta qui :