Big Data for Open Innovation in SMEs and Large Corporations: Trends, Opportunities, and Challenges

Big Data for Open Innovation in SMEs and Large Corporations:

Trends, Opportunities, and Challenges

Abstract

The notion of "Big Data" has recently been attracting an increasing degree of attention from scholars and practitioners in an attempt to identify how it may be leveraged to create innovative solutions and business opportunities. Specifically, Big Data may come from a variety of sources, especially sources outside the usual boundaries of organizations, and it represents an interesting and emerging opportunity for sustaining and enhancing the effectiveness of the so-called open innovation paradigm. However, to the best of our knowledge, none of the prior works provided a broad overview of the use of Big Data for open innovation strategies. We aim to fill this gap. In particular, we have focused our investigation on two types of companies: SMEs and big corporations, reviewing the major academic works published so far and analyzing the main industrial applications on this topic. As a result, we provide a relevant list of the main trends, opportunities and challenges faced by SMEs and large corporations when dealing with Big Data for open innovation strategies.

1. Introduction

Changing times are often a sign of opportunities. The economist Joseph Schumpeter said, "*Innovations imply, by virtue of their nature, a big step and a big change ... and hardly any 'ways of doing things' which have been optimal before remain so afterward'*" (McCraw, 2007). If we read Schumpeter's words today, we find a hint of Big Data's potential disruptive effects on trade, services, production, and business models. Nowadays, the world is inundated with data generated every minute of every day, with the growth rate increasing approximately 10 times every five years (Hendrickson, 2010; Hilbert and Lopez, 2011). According to the Industrial Development Corporation (IDC) and EMC Corporation (IDC, 2014), the amount of data generated by 2020 will be 44 times greater [40 zettabytes (ZB)] than in 2009. By 2020, there will be 5,200 gigabytes of data for every person on earth, resulting in more than 40 zettabytes. To put it in perspective, 40 zettabytes is 40 trillion gigabytes, equal more or less to 57 times the number of grains of sand on all the beaches on earth[1]. Such a flow is not likely to slow down anytime soon, so this digital phenomenon represents a great opportunity for companies to obtain benefits and create value. For instance, value may consist in delivering new products and services, making faster and better decisions in real-time, and reducing cost or improving efficiency (Chen et al., 2012a).

In such complex global scenario, companies do not innovate in isolation, at least not in an effective way. There is a lot of evidence and research that supports the benefits of a strategic alliance between two or more companies to develop new products and services (Harbison and Pekar, 1998; Uddin and Lecturer, 2011). The main advantage is a better response to the challenges of globalization, complexity and uncertain business environments (Išoraitė, 2009; Kale and Singh, 2009).

[1] https://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf

Summarized as “the use of purposive inflows and outflows of knowledge to accelerate internal innovation, and expand the markets for external use of innovation, respectively” (Chesbrough, 2003), the concept of Open Innovation emerges as one of the major challenges to sustaining innovation.

Big Data refers to any set of data that, with traditional systems, would require large capabilities in terms of space of storage and time to be analyzed (Kaisler et al., 2013; Ward and Baker, 2013).

In this context, Business Intelligence and Analytics (BI&A), and the related field of Big Data analytics, have increasing seen their importance in academic and business communities as a way to analyze Big Data for better insights and help on effective decision-making (Chen, 2012b). In this way, an “open” approach is necessary to manage the huge amount of data generated by different sources in real time which is proving to be one of the most important challenges of the last 15 years.

In the Open Innovation paradigm, companies can share external, as well as internal, ideas and knowledge deploying outside as well as in-house, to put on the market new products and services. In other words, the boundary between companies and their surrounding environment becomes more porous and more cooperative. Co-creation and inter-firm relationships become an essential part of the innovation process.

However, looking ex-post, over the last 15 years of the Big Data analysis evolution, we can identify a common development process traceable in the Open Innovation paradigm. In particular, it is remarkable to analyze the birth of the most important platform for managing and exploiting large (Big) data sets across computer clusters, called Hadoop. Let us look briefly at the Hadoop framework evolution.

From 2000, Google began developing many customer-oriented solutions to improve the core company activity: the internet search engine. Around 2004, Google understood that the disruptive effect of the huge amount of data generated would be an important topic for companies, public authorities, stakeholders and users. For this reason, in the same year, Google published a white paper about an in-house processing tool framework called MapReduce (Dean and Ghemawat, 2008). The following year, another IT company, Yahoo, released an open-source tool based on Parallel computing framework (Dobre, 2014). This new tool, implemented by Yahoo, was able to execute algorithm tasks together on a cluster of machines or supercomputers’ infrastructures to manage Big Data tasks. The tool was called Hadoop (White, 2012).

Since the Yahoo implementation other companies, such as Facebook, LinkedIn, eBay, Hortonworks and Cloudera, have contributed to the Hadoop project, making it, as we now know, the Apache Hadoop framework[2].

Hadoop ecosystem can supports the decision process of a board-level by gaining new information and knowledge, delivering new products and services, and developing powerful knowledge networks from a physical space to virtual space. In order to provide consistent information, the system uses a dynamic processes and tight collaborations among different stakeholders around the world (Hsiao-Kang and Chun, 2015)

Hadoop presents a different open-source projects and tools that provide various adaptable services, like on-demand workspace, interaction, information sharing or collective problem solving, turning them into its big

[2] List of institutions that using Apache Hadoop for educational or production: https://wiki.apache.org/hadoop/PoweredBy. List of companies that offer services or commercial support, and/or tools and utilities related to Hadoop: https://wiki.apache.org/hadoop/Distributions%20and%20Commercial%20Support

advantage. That means, for each project, tools or problems, it is possible to find a community of developers, experts and users to ask questions, fix bugs and implement new features.
Same practical examples of this kind of collaboration and co-creations processes are represented by Cloudera and Cask strategic[3] partnership. Both companies aim to create a better business value by turning data analytics directly into action using Hadoop to help customers and stakeholders to overcome the challenges in setup new applications, and to accelerate the value creation from operational analytics. Another interesting practical case of collaboration using Hadoop framework is in the manufacture industry sector (L. Hsiao-Kang, 2016).
The integration across a hyperconnected manufacturing collaboration system to reduce the complexity of extracting, processing and analyzing information, data and products, generate new opportunities and present new challenges for businesses across every industry sector (J.A. Harding, 2013).
The birth of the Hadoop framework shows how the open collaboration (co-creation) was a key factor in implementing and solving technological issues, but also highlights how to face common problems and find solutions in terms of business opportunities, by looking to the outside community.
The process described above, and the possibility to meet/bring other companies and stakeholders in a virtual space, represents for us the logical conceptual connection between the Open Innovation paradigm and Big Data analysis. This intimate connection with the enabling process of the Hadoop platform, itself born "open", represents an ideal set of strategies to leverage and enrich an Open Innovation and knowledge-sharing process. Most organizations' value and competitiveness depend on the development, use and distribution of knowledge-based competencies. Therefore, developing strong knowledge networks becomes a strategic way to access the right information and find the most valuable collaboration.
Indeed, if the birth of the open platform for Big Data analysis can very well sustain an Open Innovation strategy, can be useful to observe the main implications of this challenging ecosystem in terms of trends, opportunities, and threats for SMEs and large corporations.
Our study aims to provide an overview of the use of Big Data for Open Innovation strategy.
To answer these question, we review the main published academic works and certain relevant business applications. As well as contributing to updating the academic debate, this study examines the potential value that Big Data can offer to Open Innovation strategies for both large corporations and SMEs.
The rest of this paper is organized as follows: In the next section, we discuss the literature background structured into two main paragraphs: (i) a critical review of the literature related to Big Data, (ii) reviewing the literature on Open Innovation under the light of Big Data. Then we present an overview of the main trends, opportunities and challenges coming from the use of Big Data for Open Innovation strategies. Then some practical implications coming from the use of big data for open innovation strategies are discussed. In the last section, we discuss the main conclusions, limitations, and ideas for future research.

2. Literature review

[3]https://www.cloudera.com/more/news-and-events/press-releases/2015-02-11-cloudera-and-cask-announce-strategic-partnership.html

The literature background section has been structured into two main paragraphs focused on the topics of Big Data and Open Innovation. With the aim to provide a comprehensive reading of the fragmented debate on Big Data, as resulting from the merging of Business Management and Information System backgrounds, the first paragraph presents a critical review of the literature related to Big Data, to assure the larger comprehension of the phenomenon as well as its implications for organizations. Focusing on the challenge related to the creation of value from Big Data, the second paragraph has the objective of reviewing the literature on Open Innovation at the light of Big Data by exploring its meaning and implications for the open innovation strategies of both SMEs and large corporations.

2.1. Big Data at a glance

Big Data is a top trend in the debate of academics and practitioners (Gandomi and Haider, 2015). Overcoming the ubiquity of the word (De Mauro et al, 2016; Ward and Baker, 2013), Big Data represents a promising frontier in the future agenda of researchers and scholars in the fields of business management and information systems. In a spotlight article published in the Harvard Business Review, assumed as a manifesto of the Big Data movement, McAfee and Brynjolfsson (2012) claimed that Big Data are the cause of a new millennium industrial revolution, with a large set of unexplored implications and meaning. Looking at dynamics of industrial investments, Big Data confirmed in 2015 a trend of growth, and this demonstrates, according to Heudecker, research director at Gartner, that the topic of Big Data is becoming mainstream (Gartner, 2015).

Assumed as the most representative synthesis of the complexity characterizing the current socio-economic scenario and its configuration as a knowledge-based economy, Big Data refers to any set of data that, with traditional systems, would require large capabilities in terms of space of storage and time to be analyzed (Kaisler et al., 2013; Ward and Baker, 2013).

Big Data can be seen as the result of an evolutionary process in the field of IT technologies that started in the 1960s by moving from the phases of data processing (1960s) to information about applications (1970s-80s), from the emergence of decision support model (1990s) to data warehousing and mining (2000s) (Manning, 2013).

Two other components contributed to start of the Big Data era. The first was the increase of socialization and the advent of social-network platforms like Facebook, Twitter, and Pinterest. Socialization and data sharing also became easier thanks to the launch on the market of smart devices, such as smartphones, tablets, and wearables. All these smart technologies, followed by many open platforms, rapidly boosted the Big Data era and its potential applications (Lohr, 2012).

An example of the data proliferation magnitude is shown in table 1.

Source	Production
Apple	• Approximately 47,000 applications are downloaded per minute
Facebook	• Every minute, 34,722 Likes are registered • 100 terabytes (TB) of data are uploaded daily
Google	• The site gets over 2 million search queries per minute • Every day, 25 petabytes (PB) are processed

Instagram	• Users share 40 million photos per day
Twitter	• The site has over 645 million users • The site generates 175 million tweets per day
WordPress	• Bloggers publish nearly 350 new blogs per minute
YouTube	• Users upload 100 hours of new videos per minute • Each month, more than 1 billion unique users access YouTube

Table 1 - Growth rate of unstructured data
Source: N. Khan, et al. (2014). Big Data: Survey, Technologies, Opportunities, and Challenges.

The second component that contributed to the birth of the Big Data era was the possibility of accessing and storing data and programs over the Internet instead of on a personal computer's hard drive. Known as cloud computing (Marston et al., 2011), it is one of the best ways for both individuals and firms to store and access data and programs easily, cheaply, at any time and from any device.

As argued by Vance (2011), the decrease of costs in storage and the large diffusion of cloud solutions made available by well-known providers, such as Amazon, Google and Microsoft, have positively influenced the adoption of Big Data technologies and approaches. In this scenario, some open-source solutions such as Hadoop have started to be configured as standards for storing and processing large and differentiated datasets (Hashem et al., 2015). The availability of cloud computing solutions for Big Data management is an opportunity for all companies, especially SMEs, often limited by scarce financial and organizational resources (Vajjhala and Ramollari, 2016).

Due to the above perspectives, Big Data can give businesses, both SMEs and large corporations, the invaluable opportunity to perform novel, dynamic and scalable data analysis more quickly than ever before in human history.

A set of critical dimensions arose as suitable perspectives for the comprehension and management of the Big Data paradigm. In a first stage, those variables were identified by the three Vs of Volume, Velocity, and Variety (Laney, 2001; McAfee and Brynjolfsson, 2012). As result of a perspective more focused on the ICTs' dimension, Volume is related to the space of storage required by servers and databases, currently estimated in exabytes (10^{18}) although this is a trend in continuous growth (Anshari and Lim, 2016; Kaisler et al., 2013). Velocity, identified as the speediness of creation, sharing and storage (Kailser et al., 2013) is also representative of the obsolescence of the data available that in some industries are more meaningful of the volume (McAfee and Brynjolfsson, 2012), while Variety refers to the fragmentation of types of data available that can be text, images, videos, audio, etc (McAfee and Brynjolfsson, 2012; Kailser et al., 2013).

In addition to these, a second generation of Vs has been identified to offer the managerial challenges associated with Big Data. Aimed at comprehending how to transform data in valuable inputs to enhance companies' competitiveness, the Vs of Veracity, Variability and Value have been more recently assumed as new challengeable dimensions with larger managerial implications (Gandomi and Haider, 2015). Veracity in Big Data recalls the need of assuring reliable and confident interpretation of data; Variability is instead related to the management of changes occurring in the data, due to the continuous updating as well as to the perspectives of interpretation adopted (Fan and Bifet, 2013). Finally, Value is the most challengeable dimension of the Big

Data phenomenon, addressing the usefulness of data for decision making (Kaisler et al., 2013) and the improvment of the overall business performances (McAfee and Brynjolfsson, 2012).

Deepening the value dimension, critical issues regards the heterogeneity, accuracy and scalability of unstructured data and their integration process (Bifet et al.,2010; Intel, 2012).

Big Data analysis involves analytical approaches based on software solutions and sophisticated statistical algorithms such as machine learning, sentiment analysis, business analytics, clustering, social network analysis, (De Mauro, et al. 2016). Useful for quickly generating models to explain and predict meaningful trends by shaping the fast-moving universe of data, those approaches require time and a sufficiently large dataset for training in order to assure high responsiveness and punctual analytical performances (Gerry et al., 2014). Despite the intuition related to the benefits of Big Data management in contexts of both SMEs and large corporations, its practice continues to be populated by successful stories of latest ones (Marr, 2016), advantaged by more sophisticated human and organizational structures.

There are no industries that remain uninterested in Big Data, neither there are areas of business excluded from its potential application, and this is the reason behind its identification as a strategic factor in firms' competitiveness. The impact on the daily choices of individuals and organizations, both public and private, is disruptive, as radical as their effect on society (Jin et al., 2013).

The ability of firms to aggregate, elaborate and analyze the data is becoming a key competitive advantage and resource. However, few studies have evidenced the role and the outcomes that Big Data analysis could bring to firms in terms of innovation, efficiency, productivity, quality, and customer satisfaction (Groves, et al., 2016).

The review of the literature previously recalled highlights the fragmentation of contributions characterizing the debate on Big Data and the need of more investigations in the areas of convergence resulting from the merging of studies in management and information systems. In this regard, De Mauro et al. (2016) have tried to provide a first systematization of the discussion on Big Data by connecting the technological dimension with the managerial one and highlighting the value creation from the transformation of such large information asset as the main challenge for companies (De Mauro et al., 2016). However the complexity of the phenomenon as well as the need of deepening its implications for the value creation of companies call for a major comprehension.

With the aim to contribute at this goal, the following table presents a synthesis of the different perspectives of studies resulting from the critical review of the literature conducted, in terms of research focus, description and main references.

Focus	Description	Main References

Form and Nature of Data	- Structured, semi-structured and unstructured; - Text, video, audio, pictures, codes, satellite images etc.	Secundo, et al., 2017; Khan, et al., 2014; Hurwitz et al., 2013; Kailser et al., 2013; McAfee and Brynjolfsson, 2012; Ohlhorst, 2012; etc.
Sources	- From and about physical world, generated by sensors, scientific observation, also known as “machine generated” or “Internet of Things”; - Data from and about human society, generated by social networks, web, marketing, also known as “human generated” or “Internet of People”.	De Mauro, et al., 2016; Jin et al., 2015; Bi et al., 2014; Feki et al., 2013; Chen et al., 2012; etc.
Tools	- Proprietary (SAS, SPSS, etc) vs Open Sources (Hadoop, R, etc.) tools for data extraction, storage, processing, etc	De Mauro, et al., 2016; Hashem et al., 2015; Kailser et al., 2013; Shvachko et al., 2010; etc.
Approaches	- Machine learning, business analytics, semantic clustering, social network analysis, data mining, etc.	De Mauro et al., 2016: Gerry et al., 2014; Wu, et al., 2014; Chen et al., 2012; Sebastiani, 2002; etc.
Value creation	- effectiveness in decision-making; improvement of companies’ overall performance; renewal of companies’ intangible assets; better positioning in the industry; higher customer satisfaction; marketing analytics; innovation in products; innovation in business models, etc	Secundo et al., 2017; Abbassi et al., 2016; Gandomi and Haider, 2015; Jin et al., 2015; Bi et al., 2014; Del Vecchio et al., 2014; Mayer-Schönberger and Cukier, 2013; Kaisler et al., 2013; McAfee and Brynjolfsson, 2012; Brown et al., 2011; Laney, 2001; etc.

Table 2. An overview of the literature on Big Data

The critical review of the literature resumed in the previous table demonstrates the actuality of the topic in the scientific debate. The references collected under the issues of form and nature data, sources, tools and approaches offer a wider consolidated evidence for the challenges associated to the Vs of volume, variety, and velocity, as first three key dimensions of Big Data (Laney, 2001) as well as for the other two Vs of veracity and variability. As for value, Big Data can contribute at the value creation process through several roots identified into making more effective the decision-making (Kaisler et al., 2013), the improvement of companies’ overall performance (McAfee and Brynjolfsson, 2012), the renewal of companies’ intangible assets (Secundo et al., 2016; 2017), the better positioning in the industry (Kailser et al., 2013), higher customer satisfaction (Brown et al., 2011), marketing analytics (Xu et al., 2015), innovation in products (Mayer-Schönberger and Cukier, 2013) or business models (Brown et al., 2011). Despite the different perspectives recalled under the issue of value creation and the identification of potential contributions for the innovation process, how the huge amount of data available can be managed for executing an open innovation strategy and create sustainable innovation performance needs more work.

In order to assure a major comprehension of this implication, the next paragraph presents a critical reading of literature on open innovation through the lens of Big Data.

2.2 Open Innovation in the age of Big Data: SMEs vs Large Corporations

The value perspective of Big Data approach can result in a plurality of elements to assure the competitive positioning of companies. Focusing on the meanings previously recalled, the most challengeable dimension of value is identifiable as the contribution at the conception and execution of firms’ open innovation strategies.

Open Innovation is the result of an evolutionary path in theory and practice related to innovation management, and even though it cannot be considered new (the first conceptualization by Henry Chesbrough was in 2003) it continues to recall the interest for scholars and researchers in the field of innovation management (Spithoven et al., 2013; Gassmann et al., 2010).

Defined as “the use of purposive inflows and outflows of knowledge to accelerate internal innovation, and expand the markets for external use of innovation, respectively’’ (Chesbrough, 2006), Open Innovation is a new paradigm shift from previous innovation approaches, defined as closed innovation and characterized by traditional forms of research and development performed by companies (Gassman et al, 2010). The relevance of Big Data in the debate on Open Innovation results from the a set of elements that make actual the same paradigm of innovation through the usage of internal and external knowledge flows, identified by Dahlander and Gann (2010), into the social and economic changes occurring in consumption and production, the globalization as influencing the labor organization, the emergence of new and complex challenges into the issue of intellectual property, and the viral diffusion of new ICTs.

It is important to note that with open innovation the role of R&D departments continues to be relevant even if integrated with the huge amount of data available externally. This scenario confirms as the process of value creation cannot be afforded by companies singularly or autonomously (Prahalad and Venkatraman, 2008; Prahalad and Krishnan, 2008), but results are more effectively achieved by matching the organizational assets with external knowledge and skills in a network perspective (Gulat, et al., 2000; Iansiti and Levien, 2004). Assuming Big Data as the synthesis of the complex ocean of information and data created by and within companies’ networks, Open Innovation is the more suitable approach to assure the transformation of these data into inputs for the innovation process of companies. This can benefit of the new managerial and analytical approaches available to afford the challengeable dimensions of Big Data in terms of volume, variety and velocity as well as to allow their full exploitation into knowledge assets. The characteristics of Big Data match positively with the principles of Open Innovation defined by Chesbrough (2003). Specifically, it is possible to see in the light of Big Data that Chesbrough’s assumptions on good ideas are widely distributed: the absence of ideas’ monopoly, the lack of assurance of commercial advantage due to the timing of discovery, that the goodness of a business model can be preferred to the technological performances, and finally the perishables of intellectual property are all confirmed both in the practice as in the theory. As a multidimensional phenomenon, examinable along different perspectives (external R&D, mechanisms for intellectual propriety, etc.), open innovation can assume different forms (Dahlander and Gaan, 2010).

Identified as a conceptual paradigm useful for understanding and orchestrating the evolutionary path of industrial innovation (Chesbrough, 2006), Open Innovation presents characteristics associated with the companies’ industry and size. Starting from the pioneer experience of Procter & Gamble, the first to institutionalize the practice of opening the R&D department to external actors, the adoption of open innovation approaches by large companies is nowadays a consolidated pattern diffused in almost all the industries (Gassman et al., 2010). The literature on Open Innovation continues to be largely supported by empirical evidence related to large corporations. Best practices have been identified in the industry of manufacturing

(Laursen and Salter, 2006), healthcare (Hughes and Wareham, 2010), pharmaceutical (Bianchi et al., 2011), automotive (Ili et al., 2010), food (Sarkar and Costa, 2008; Bigliardi and Galati, 2013), etc.

The adoption of open innovation by SMEs is not only limited in practice due to resource constraints; it also has low consideration in theory (Spithoven et al., 2013). Positive examples of Open Innovation in SMEs can be identified in the cases of the so-called "born globals," small companies able to perform a rapid path of growth on global markets (Gasmann et al., 2010).

In a comparative analysis of the open innovation patterns in SMEs and big corporations, Spithoven et al. (2013) have demonstrated how, in terms of effectiveness, SMEs are more performant in the introduction of new products while large corporations leverage on search strategies for their products' turnover.

Focusing on the different rates of adoption of open innovation in SMEs and large Corporations, the contributions of scholars and researchers have demonstrated how the diffusion of innovative models of interactions and technological platforms of collaboration can allow to overcome the organizational and technological limitations (Battistella and Nonino, 2012; Ndou et al., 2011; van de Vrande et al., 2009) However, at the light of Big Data, new solutions are required in order to make available the full exploitation of data available into the open innovation strategies of SMEs and large corporations.

The reading of open innovation through the lens of Big Data offers a further interesting area of speculation, represented by the opportunities emerging in the process of corporate entrepreneurship, as process of continuous renewal in the context of existing companies (Lazzarotti and Manzini, 2009; Ebner et al., 2009; Secundo et al., forthcoming), as well as in the launch of new entrepreneurial ventures, such as industrial and academic spinoffs, start-ups, etc. (Chesbrough, 2006; Christensen et al., 2005; Gilsing et al., 2010). The first ones should benefit of the large and distributed basis of data for a more effective implementation of their strategies of innovation and renewal in terms of products, processes, market and organizational structure. The second ones should reduce the margin of uncertainty related to the entry in the market by benefiting of larger evidences on competitors and customers (Secundo et al., forthcoming).

Open innovation in the age of Big Data highlights the centrality of the value networks in which firms operate as resulting from a complex reticulum of ties and relationships involving a plurality of stakeholders, mainly customers, competitors, public and private institutions, universities and research centers (Zajac and Olsen, 1993; Powell et al., 1996). Embedded into complex and distributed networks of actors, populated by knowledgeable stakeholders, firms need to adopt Big Data approaches and tools to acquire additional knowledge and skills for nurturing their innovation strategies (Gatignon et al., 2002; Hauser et al., 2006; Chesbrough et al., 2006).

However, while it is agreed that a huge amount of data exists, how it can opportunely be managed to sustain the innovation process of companies still receives little consideration. Focusing on this aspect, this paper aims to shed new light on the meaning of Big Data in an open innovation perspective, notably the way in which, in the experience of SMEs and big corporations, Big Data can support the conception and execution of an open innovation strategy for making companies more competitive (Chesbrough, 2011; Ollila and Elmquist, 2011) and opening new entrepreneurial opportunities (Eftekhari and Bogers, 2015). Moreover, it is interesting to understand how Big Data can be leveraged for developing inbound and outbound open innovation strategies (Dahlander and Gann, 2010); how they may impact on the companies' absorptive capacities (Cohen and Levinthal, 1990; Lane et al., 2006; Ooms et al., 2015) as well as contribute to their competitive positioning (Chesbrough, 2011; Ollila and Elmquist, 2011).

Considering Big Data as an important approach to help companies to maximize their open innovation experience, it is important to be aware about all the aspects related to the creation of value from Big Data presented above. In order to clarify these points, in the next section we discuss the main trends, opportunities and challenges faced by SMEs and large corporations dealing with Big Data for open innovation strategies. After that we recall these challenges highlighting some practical implications, such as those related the human resources, the privacy threats and the IPR.

3. Big Data for open innovation strategies

As largely discussed in previous sections, recent research calls for investigation as to how Big Data can be used for leveraging open innovation strategies. In particular, this analysis needs to distinguish between SMEs and large corporations due to their well-known structural differences. Therefore, in this section we provide a relevant list of the main trends, opportunities and challenges faced by SMEs and large corporations when dealing with Big Data for open innovation strategies. In particular, we review the main academic works published so far and analyze the main industrial applications on this topic.

3.1 Trends

As recently stated by Tom Rosamilia, senior vice president at IBM Systems and Technology Group[4], "Big Data accelerates the opportunity for new discovery while at the same time magnifying the challenge scientists face … the current approach to computing presumes a model of data repeatedly moving back and forth from storage to processor in order to analyze and access data insights, a process that is unsustainable with the onslaught of Big Data because of the amount of time and energy that massive and frequent data movement entails." He continues by saying that "the Big Data challenges can only be solved through open innovation… (since) the classic computing technologies will obviously continue to evolve but at a rate far short of the rate at which data is growing." In other words, firms (especially large corporations) are facing an increasing trend in terms of available data, although the existing technologies used by firms are not effective to analyze them. On the other hand, open innovation strategies are characterized by a large pool of resources (i.e. communities of users, or users coming from different sectors and with different skills or tools updated and upgraded, with continuous improvements coming from large communities) that can be the best (or the sole) way to manage Big Data.

Many initiatives characterized by the implementation of open innovation strategies through communities of customers who generate Big Data are rising and, as a result, McKinsey Quarterly (Bughin et al., 2010) listed

[4] http://www.dbta.com/Editorial/News-Flashes/US-Department-of-Energy-Taps-IBM-to-Develop-Supercomputers-to-Meet-Big-Data-Challenges-100569.aspx

the distributed co-creation of value as the first business trend to watch. Many large corporations are co-creating their value by leveraging communities of web participants to develop, market and support products and services, thus extending their reach and lowering the costs of serving customers. For example, this is the case of pioneers, such as Wikipedia or open source software developers, or the 70 percent of the executives surveyed by Bughin et al. (2009). Therefore, implementing open innovation strategies through communities of customers who generate Big Data can be a source of value, allowing firms to gather ideas and/or insights from a bigger and more committed pool of users (Gerry et al., 2014). As a result, the usage of open innovation strategies through communities of customers who generate big data has been translated into different initiatives, such as markets of ideas or crowdsourcing (Garavelli et al., 2013; Natalicchio et al., 2014). However, since the co-creation is a two-way process, these firms are called to provide feedback to stimulate continuing participation and commitment. Hence this implies several additional activities, such as coordinating the community of users, communicating with most of the users in the community or rewarding the deserving users. All these additional activities generate costs in terms of time and money, thus calling for a cost/benefit analysis before implementing such strategies.

This trending attention to the use of Big Data for open innovation strategies is also pushing firms, especially large corporations, to find new and efficient ways to pursue this goal. As a result, leading companies, such as the Spanish multinational bank BBVA and the American multinational software producer Splunk, offered prizes on this topic. In particular, these companies rewarded ideas in the open innovation Big Data field. Looking at the categories of these prizes and their winners gives us a good flavor of the main trends in this topic. In particular, Splunk, a famous American multinational, in 2015[5] split the Big Data open innovation competition into three categories: fraud/insider threats, social impact, and innovation.

The winner of the first category proposed an app that detects insiders and malicious behavior by assigning risk scores for suspicious events. Each risk event is related to a risk object, which could be a user or an employee abusing the system. If a risk object behaves badly over time, the risk score goes up such that it warrants further investigation. The innovation aims to make it easier to detect insiders and to avoid opening security incidents for every minor suspicious action, which is usually the cause of many false alerts. The winner of the second category proposed a solution to monitor, analyze and optimize Energy Load Profiles. Gathering these insights from the data could make it easier for utility companies to increase operational efficiency by optimizing their processes. Finally, the winner of the third category proposed a way to automatically change call routing based on business conditions and to roll back to normal call routing when appropriate.

Regarding the BBVA big data open innovation contest in 2014[6], there were three categories: one aiming to improve the everyday life of people through relevant information, one aiming to help companies and/or the public sector with decision-making, procedures, processes and/or products/services, and the last one aiming to convert data into understandable images in order to improve their usability. In the first category, the winners proposed a travel planner that helps users avoid waiting in lines in stores or pay services, a customized

[5] https://www.splunk.com/ https://www.splunk.com/ https://www.splunk.com/
[6] https://bbvaopen4u.com/en/actualidad/innova-challenge-big-data-example-open-innovation

recommender system for suggesting interesting items to the users visiting a new city, and a virtual assistant who could be asked any kind of question and would give real-time answers. In the second category, the winners proposed an application which predicts whether a business should be successful or needs improvements, an application assessing the social and economic impact on a city measured as a real-time function of the events operating in the city itself, and an application measuring the firm's direct marketing strategies and suggesting how these can be improved. Finally, in the third category, winners proposed a tool for visualizing where and how successful businesses are developed in specific areas, a tool for visualizing datasets from a number of perspectives and an interactive tool for visualizing several (business or social) performances in specific regions.

In conclusion, the main trend is in using open innovation strategies for handling and enhancing Big Data. In fact, the large pool of resources coming from open innovation strategies is the best solution for managing Big Data. This trend is particularly relevant for large companies than SMEs since the former are, in general, more affected by a proliferation of data and by a higher pool of committed users with respect to the latter. As a result, several large corporations are taking advantage of an "open" co-creation of value based on collaboration and communication with their customers who generate big amount of data that can be used as a source of value (such as markets of ideas, crowdsourcing, etc.). This approach is supported by providing customers with feedback to stimulate continuing participation and commitment. Finally, there also exists a significant trend in developing customer-based applications or decision support-based applications for companies. The former use Big Data coming from multiple "open" sources to help customers in their everyday life, the latter use Big Data to help firms in their decision-making processes and it is a significant trend for SMEs and large corporations since both of them need an automatic help for improving their decision-making processes when dealing with big data for open innovation strategies.

3.2. Opportunities

One of the main opportunities in Big Data for open innovation comes from the use of sensors or, more generally, all the devices characterizing the Internet of Things. The Internet of Things (IoT) is a new paradigm of information networks with the aim of expanding the potential of the conventional Web (Atzori et al., 2010; Feki et al., 2013; Whitmore et al., 2015). The rationale behind the IoT is in its denomination, as in "Internet" and "Things". The former reflects a network-oriented vision of communication, which entails the use of hardware, standards, and protocols characterizing the Web 2.0, whereas the latter tends to move the focus onto physical objects, rather than end users, as the "things" to be connected (Atzori et al., 2010). Put together, IoT semantically means a "world-wide network of interconnected objects uniquely addressable, based on standard communication protocols" (Bandyopadhyay and Sen, 2011:50). The IoT has thus been deemed the next logical evolution of the Web and a disruptive revolution in the ICT world (Feki et al., 2013), in that it provides us with instant and remote access to information about physical objects, thereby leading to innovative networked systems with higher efficiency and productivity (Bandyopadhyay and Sen, 2011; Bi et al., 2014). Therefore, multiple devices and/or sensors are used for real-time web-connected applications, thus also increasing the amount of data collected and the open innovation opportunities.

One example comes from the Thales Innovation Hub in Hong Kong[7]. Among other things, the hub is developing a big data platform using an adapted multi-sensor and data fusion strategy. The platform will be able to prototype smart transportation applications from diverse data sources, thus enabling current transportation challenges to be addressed, mainly around real-time crowd monitoring solutions and predictive maintenance. As an additional evidence of this opportunity, Engie (ex-GDF Suez, and a leading company in the energy market) recently proposed[8] through its open innovation lab[9] a context for projects using the Internet of Things and Big Data to support the development of the cities of tomorrow. In particular, this "cities of tomorrow" call for projects is focused on identifying innovative solutions in the field of the Internet of Things and Big Data to enable stakeholders, including citizens, businesses and local authorities, to better visualize, use and value the information they need. The project aims at finding innovative solutions for home comforts, smart cities, mobility, territories, energy, security, lighting, waste, water, and smart government.

Draxler et al. (2014) discussed how Big Data can be an opportunity for leveraging open innovation. In particular, the authors proposed the "cup of information" model, in which a huge amount of data coming from different sources (such as social networks, websites, blogs, etc.) can be used as the input of an open innovation process. This model can be used to develop new products, processes, and services, as well as to identify external partners and start new projects in an open innovation environment. The authors provided two empirical examples of the proposed model using two case studies: an insurance agency and a paper industry. In the first case, the model provided some insights for new product development, while in the second case it offered a quick and cheap approach for retrieving interesting patents for new product development.

In particular, regarding the first case, one major trend is the higher mobility of seniors, people who are retired and use their leisure time for traveling, exploring foreign cultures, and visiting exotic countries. When the "cup of information" model analyzed the internal database of the insurance company, the software found that traveling and travel insurances are relevant for that kind of company. Then it used "travel" as one of many keywords when doing its semantic search of the external data streams. The output was a tag cloud containing the connections of the trend "travel" with other trends. Surprisingly, the insurance company realized that there was a strong link to a technology called "paper microfluidics". A closer look revealed that this technology is a potential key enabling technology to develop mobile diagnostic devices (i.e., coupling this technology with a smartphone could offer completely new diagnostic methods for seniors even at remote destinations). Accordingly, the insurance company started a project aimed at offering a completely new product for senior travelers: a package combining classical travel insurance with a new, high performance but low cost diagnostic device, which can considerably reduce the risk of overseas travel for elderly people. Therefore, the company brought down costs (seniors who use that device will have better control of critical medical parameters and

[7]http://www.businesswire.com/news/home/20160328005783/en/Open-Innovation-Thales-Partners-Big-Data-Smarter
[8]http://openinnovation-engie.com/en/detail/opportunities/using-the-internet-of-things-and-big-data-to-support-the-development-of-the-city-of-tomorrow/1348
[9]http://openinnovation-engie.com/en/

thus avoid illness) and, at the same time, gave customers a sense of security. A byproduct of this approach is the identification of partners for the whole Open Innovation value chain (during the new product development). In the second case, the paper industry had continuously to watch out for emerging technologies by screening patents. Similar to the case of the insurance company, the “cup of information” model analyzed the internal input, which in this case was the company’s homepage, and found the initial trends relevant for the company. Then these terms were used for searching data from other sources, such as science databases, patent databases, the World Wide Web, blogs, and tweets. The output of this process was a measure of correlation between each internal and external input. The analysis revealed that the new term “LumeJet” showed good correlation with a number of internal terms like paper, print head, and digital printer. Digging deeper into the files, it became obvious that a company called LumeJet had developed a new inkless digital printer for ultra-high quality printed output. The company extended this insight into an in depth study of this technology in order to find out whether it posed opportunities or threats to the company. This case study demonstrated that this approach, based on open innovation and Big Data, is useful for exploring the activities of both large corporations and SMEs (or startups). In fact, on one hand, large corporations can be monitored quite easily via their websites, but it is difficult to find out what’s going on inside their R&D pipelines using traditional approaches. On the other hand, exploring the R&D activities of a large number of small but innovative companies requires significant effort, and potentially disruptive products of startups are also hard to detect early enough.

In many cases, the mix of big data and open innovation results in digital platforms where a myriad of other contributors are attracted and, when sufficiently rich, it can result in the formation of an ecosystem (Kenney and Zysman, 2016). For example, in the case of the app stores, complementary businesses are emerging. AppAnnie is a firm that ranks the revenue generated by apps; TubeMogul classifies YouTube “stars” and measures their reach, and several agencies are managing new YouTubers. This evidence highlights the opportunity for new business models working on Big Data for open innovation (or vice versa), thus making this mix a value proposition for their business (Kenney and Zysman, 2016). The aforementioned firms included services based on Big Data for open innovation as their “core” value proposition but, at the same time, it is possible to use Big Data for open innovation activities in order to open secondary markets in existing business models.

In conclusion, the main opportunity lies in the Internet of Things phenomenon. In fact, several devices are collecting open Big Data that can be used by SMEs for starting new businesses or by large companies for improving their existing businesses. Another relevant opportunity lies in the use of Big Data and open innovation strategies for developing new products which is a high value opportunity for both SMEs, which are looking for entering, or improving their position into, a market and large companies which are looking for leading the market. Finally, SMEs have also the huge opportunity to build new or innovative value propositions for their business models in the industrial ecosystem that is growing around the initiatives based on the use of Big Data for open innovation strategies.

3.3. Challenges

The first big challenge in the use of big data for open innovation strategies lies into the intersection between the challenges associated with the big data and the different types of open innovation strategies. In fact, McKinsey (2013) warned about the three key challenges of using Big Data: i) deciding which data to use (and where outside your organization to look), ii) handling analytics (and securing the right capabilities to do so), and iii) using the gained insights to transform the operations. In addition, open innovation strategies can be inbound, outbound and coupled (Chesbrough, 2003, 2006; Gassmann and Enkel, 2004; Enkel et al., 2009) with data coming from either inside or outside the organization (or even both). As a result, the first big challenge is about how to minimize the issues coming from the use of big data when dealing with different types of open innovation strategies. In addition, a relevant challenge is related to the skills and capabilities required for exploring the new research questions coming from the mix of Big Data and open innovation (i.e., open data warehouses and similar sources). In fact, data are now more easily available from multiple "open" sources, and they encourage researchers to access platforms and develop solutions for questions that have not received attention until now (Gerard et al., 2016). Therefore, firms need to develop skills related to data access and collection, data storage, data processing and, most of all, data analysis, reporting and visualization for effectively leveraging Big Data to enhance the benefits they may gain from open innovation strategies. As a result, the main challenge related to the use of big data (i.e., skills for handling them) is amplified when these data are applied for open innovation strategies due to the increasing complexity of the whole process. This challenge is particularly relevant for SMEs since these type of skills are difficult to find and, more important, expensive to acquire. On the contrary, large companies can benefit from acquiring these skills both in the form of outsourcing or acquisition inside the company itself.

Some important challenges in using Big Data for opening the innovation process are also provided by Draxler et al. (2014). In particular, the authors suggest defining the open innovation process in terms of where ideas are supposed to come from, how needs are being discovered, how prospective partners can be identified. They also suggest enabling the organization to access and handle big volumes of new data from multiple sources, especially from social media, and to select advanced analytic tools that help discover new ideas and predict outcomes of business decisions from these data. Finally, they suggest making sure that the outputs of the analytic models are translated into tangible actions such as improving the development of the next generation of products and creating innovative after-sales service offerings. In other words, Draxler et al. (2014) pointed out that one challenge of using big data for open innovation strategies lies in focusing the whole process on the innovation that the firms want to discover from the process. This initial step is mandatory for improving the efficacy of the big data analyses that otherwise may result in not useful results with respect to the open innovation goals pursued by the firm. This challenge is particularly relevant for large corporations because they may have a more complex structure and more articulated list of desired outputs with respect to SMEs thus making more difficult the alignment of big data analyses with the open innovation strategies.

Threats to privacy and security are often seen as the dark side of Big Data, but there is also a third danger: that of "falling victim to dictatorship of data, whereby we fetishize the information, the output of our analyses, and end up misusing it. Wielded unwisely, it can become an instrument of the powerful, who may turn it into a source of repression, either by simply frustrating customers and employees or, worse, by harming citizens" (Mayer Schoenberger and Cukier, 2013: 151). This aspect becomes a relevant challenge when dealing with big data for open innovation strategies since analyses are done on multiple and open sources that can be affected by several biases or by several inconsistencies thus leading into biased or false results. Therefore, when dealing with big data for open innovation strategies, another relevant challenge consists in balancing the trust in the results and the critique of them.

Related to this point, in a so-called Big Data world "knowing what, not why, is good enough" (Mayer-Schoenberger and Cukier, 2013), meaning that correlating massive, rapid, and versatile data streams yields fast and clear insights, and helps develop new products or services – such as in the case of Amazon's recommendation system which looks for associations between products. Recommender systems are automatic tools for profiling users and retrieving items that can be suggested to them for purchases. Among many algorithms, it was demonstrated that contextual recommender systems (i.e., systems that recommend specific items for users' specific contexts) or profit-based recommender systems (i.e., systems that recommend items that are both liked by the user and profitable for the firm) can improve business performance, such as sales (Panniello et al., 2016a; Panniello et al., 2016b). However, correlation does not necessarily imply causation, and therefore the usage of these systems calls for understanding what the results obtained from them mean (Mayer Schoenberger and Cukier, 2013) especially when analyses on these big data are done for open innovation strategies. Applying the system and using the result "as is" is not enough and can lead to bad innovation strategies. On the contrary, it is important to understand the rationale beyond the results in order to better use them. For example, recommending a specific book to a customer may lead to the purchase of the recommended book; but understanding why the customer likes the recommended book may lead to additional innovation opportunities (e.g., the user likes the book because he is a painter, and therefore it could be useful to innovate the design of the website home page or for innovating the bundle of products currently sold).

Another relevant challenge, strictly correlated with the previous ones, is the call for distinguishing between reliable and false data. Recent investigations about brand recognition show that many prominent brands are using fake followers or fake opinions for their promotion, or to discredit their opponents (De Michelli and Stroppa, 2013). The consequences of unrecognizable fake input can lead to misleading conclusions being drawn from the fake data. There is already some research on how to enhance the detection of fake opinion profiles, based on content originated by such profiles using additional features from quantitative psycholinguistic text analytics tools (Duh et al., 2013). In any case, it is of paramount importance to detect such fake activities to ensure that the web remains a trusted source of valuable information. All the aforementioned challenges associated with the risk of false data are particularly relevant since the border between private and open elements (such as data) could become difficult to recognize when dealing with big data and open innovation strategies at the same time. In fact when dealing with open elements (such as open sources of data) the deception issue (i.e., ad-hoc modifications of data coming from involved subjects) can be more relevant with respect to the case when no open elements are used at all. This challenge is relevant for

both SMEs and large corporations since both of them may have a reduction of their reputation or both of them may miss their goals due to a wrong strategy based on false results or due to a wrong interpretation of them.
Therefore the main challenge lies in the mix of issues coming from the mix of big data and open innovation strategies. In fact, both of them have their how challenges that become bigger and more complex when put together. This challenge is particularly relevant for SMEs since they have few resources for handling this aspect with respect to large companies. In addition, another relevant challenge is to focus the whole process on the desired innovation output. In fact, all the analyses done on the data need to be driven by the open innovation strategy in order to reduce the risk to scatter resources. This challenge is particularly relevant for large companies because they may be more complex (in terms of processes and desired outputs) with respect to the SMEs. Another relevant challenge when dealing with big data and open innovation consists in the risk of biased or false analyses due to the mix between private and open elements (such as data) thus resulting in misleading or wrong results. This challenge is relevant for both SMEs and large corporations since both of them may incur in false or ad-hoc modified data thus resulting in misleading open innovation strategies and, in turn, damaging their business (in terms of missed goals or reduction of their reputation).

4. Open Innovation implications for Big Data applications

Big Data framework can change the way companies organize their external collaborations, catch new information from customers and their environment, and choose strategic industrial processes for products and services. In previous sections, we identified a relevant list of the main trends, opportunities and challenges faced by large corporations and SMEs when dealing with Big Data analyses for open innovation strategies.
This section aims at discussing the main practical implications coming from the discussion made in previous sections.

4.1. How Big Data can support Open Innovation strategy.

As discussed earlier in this paper, Open Innovation is one of the principal ways to reach faster and better creative solutions and facilitate problem-solving through the open flow of ideas, development of a knowledge-sharing cooperative process inside and outside the organization’s environment. To unlock and sustain most of the benefits of the Open Innovation paradigm, Big Data analysis represents an incredible technological opportunity to grasp valuable information that has been hidden until now.
Business opportunities created by Big Data can help firms to share knowledge and redefine relationships between companies. It is the “intimate nature” of Big Data and the information gleaned from them which give rise to the most valuable opportunities.
Traditionally, Open Innovation literature has stressed the importance of changing the “classic” business model to a more profitable, knowledge-sharing Open Business Model to maximize the potential complementary relationships between companies.
The interplay creates an important role in business model choice in searching to establish a link between technology innovation and competitive advantage. At the same time, choosing and setting up the right Open

Business Model can determine the nature of complementarity between business opportunities and technology that is able to gain profitable value and monetization (Baden-Fuller and Haefliger, 2013).
The Open Business model can be focalized and lead real-time Big Data process to produce more and better valuable insights for market gain, for customer intelligence and for effective marketing campaigns. Web Data-scraping, for instance, can drive information to understand what happens regarding the quality and quantity of products and services delivered. Intelligent marketing feedback enables companies to answer variations in customers' needs more quickly.
Applying a balanced combination of all these Open Business Model applications through Big Data analysis can be profitable for the companies to gain a step forward customer approach. Data-integration also enriches the Customer Relationship Management (CRM) and can be useful for creating or identifying new systems of suppliers, distributors, services, and commerce providers. This system is called a business-web or "b-web" (Weill, 2005).
Another way in which Big Data and Open Innovation can lead innovation is the outbound strategy process. The outward transfer of technology in open exploitation processes can create spin-out able to catch new opportunities, for instance from a secondary market. In particular, considering the technological aspect, the possibility of creating spin-out represents a valuable asset to explore new possibilities not previously investigated carefully, or understood well, by the company (Lichtenthaler, 2009).
In light of recent developments in the field of outbound Open Innovation strategy, companies have started seeing this as a prerequisite to effective extraction of value information from Big Data analysis for external knowledge exploitation and finding a strategic fit alliance with other companies. To be effective an outbound strategy and spin-out needs to bring together various assets and resources to explore commercialization opportunities for products and/or services in secondary markets. The combination of huge amounts of data and the potential for employees to develop new solutions can be helpful to establish technology standards outside the companies. The likelihood of having new products and attracting improvements from outside make the technology more appealing, thanks to the specialization process in the company's assets.
A rising number of technology-based firms are taking advantage of the opportunities to utilize internal technological assets by licensing out their underutilized technologies or sustaining spin-off companies. Big Data can improve all initiatives that require a company's extensive coordination in order to be effective in time, supporting activities and preventing the company from hollowing out potential profitable advantages for its own business units (Kutvonen, 2011).
One more strategic objective achievable thanks to the outbound process is enabling a company's technological specialization assets to induce strategic alliances. A lot of research has been devoted to strategic alliances and innovation partnerships, such as the motives for, and the impacts of, collaboration. According to Solesvik and Westhead (Solesvik and Westhead, 2010), the selection of the right partner is probably the most crucial aspect of Open Innovation success (Segers, 2015).
Many scholars, researchers and most of the scientific literature about Open Innovation strategy are focused on understanding how organizational changes within a company would benefit from the application of an Open

Innovation paradigm; for instance, re-organizing the R&D department; IPR management; spin-out opportunities, etc.

Big Data analysis combined with Open Innovation strategies can sustain the overall business value and positively impact efficiency and intelligence operations. All these potential benefits require companies to adopt specific organizational changes to profit from a Big Data strategy. To solve organizational challenges in the digital era, where ideas, communication and information run fast, the concept of Open innovation also needs to evolve in a new approach and process. This development process is called Open Innovation 2.0[10] (Curley, 2016).

In the OI2 innovation happens in ecosystems or networks and emphasizes the diversity of collaboration in interdisciplinary approach, to improve chances to create breakthrough innovations.

Creating awareness of the concept of Open Innovation 2.0 means it is necessary to implement organizational changes and develop some collaborative common patterns to move from concept to action. (i) The importance of online engagement platforms; (ii) More partnerships between industry, public government, academia and citizens, (iii) New product service systems able to shift from just delivering the product to related services (Curley, 2016). These are just a few examples of how organizational changes and common innovation patterns can be applied.

Open innovation	Open innovation 2.0
Independency	Interdependency
Cross-licensing	Cross-fertilization
Bilateral	Ecosystem
Linear, leaking	Nonlinear mash-up
Bilateral	Triple or quadruple helix
Validation, pilots	Experimentation
Management	Orchestration
Win–win game	Win more–win more
Out of the box	No boxes!
Single discipline	Interdisciplinary
Value network	Value constellation

Table 3: How Open innovation modes have evolved. Main difference between Open Innovation and OI2.
Source: Curley, M. (2016). Twelve principles for open innovation 2.0.

Open Innovation 2.0 strategy in Big Data needs to become a discipline practiced by many companies and organizations rather than a few companies who have mastered it. For this reason the EU research commissioner, Carlos Moedas, proposed fostering investment and efforts to establish throughout the European Union a policy to implement an Open Innovation 2.0 strategy roadmap which would develop private-public partnerships and sustain innovation (EC - Open Innovation 2.0).

[10] Open Innovation 2.0 (OI2) is a new paradigm based on a Quadruple Helix Model (Asplund, 2012) where civil society joins with government, business and academia work together to co-create and drive structural changes far beyond the scope of what one organization could do alone (Curley and Salmelin, 2013). This new model based on principles of integrated collaboration and co-created shared value can encompasses also user-oriented innovation models able to take full advantages from the exploits of disruptive technologies of digital era – such as cloud computing, Big Data, IoT – and leading to experimentation and prototyping in a real world setting (Curley, 2016).
Open Innovation 2.0, differently form Open Innovation, is not just something “restricted” to the innovation funnel, like linear model that organizations use in their innovation processes. But rather something that widens the scope and adds an essential component, the ecosystem innovation-networks, to the traditional approaches and accelerates collective learning.

4.2. Data scientist is the new unicorn

It has become extremely important for managers to understand exactly what kind of competence they are looking for to fill the data scientist position in the firm. With the rise of Big Data, we have seen the advent of a new professional profile: The data scientist. Many organizations already have fixed vacancies for data scientists, like the chief data officer (CDO). As mentioned before, one definition of data scientist comes from IBM: *"A data scientist represents an evolution from the business or data analyst role. What sets the data scientist apart is strong business acumen, coupled with the ability to communicate findings to both business and IT leaders in a way that can influence how an organization approaches a business challenge. Good data scientists will not just address business problems, they will pick the right problems that have the most value to the organization*[11]"(A.Griffin et al., 2014).

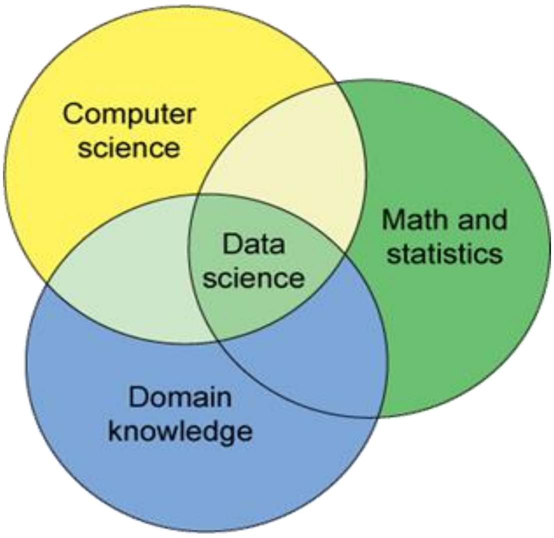

Figure 1 - Key disciplines of the data scientist.
Source: M. Jones (2013) Data science and open source. https://www.ibm.com/developerworks/library/os-datascience/

Generally speaking, the data scientist goes beyond the classic use of software tools and the management of strategic data and business analysis which we are used to in the world of innovation (Loukides, 2010; Chandler, 2014). The data scientist's responsibility begins with the design of a prototype with the technologies best suited to the problem under investigation, and then involves setting up an implementation strategy able to break down into new products or services the results obtained from value/key information previously hidden.

Data Science (Song, I. and Zhu, 2016), the field that comprise the combination of techniques used to extract data insight and information from a massive volume of data, requires a mixture of broad and multidisciplinary competence ranging from an intersection of mathematics, statistics, computer science, business strategy. Figure 1 illustrates data science and the connection of computer science, math and statistics, and domain knowledge.

As figure 1[12] shows, finding all these set of skills in one person is extremely complicated. This is why Gil Press[13] defines a data scientist, Data Science professionals, as the new unicorn[14]. Considering the scarcity of these "unicorns" is more effective and easier to create a group work environment by building a data science team with those competences (a mix of professionals) to bring more value into the companies. This context is similar to the environment found in the innovation process, which is usually a result of a work team characterized by multiple skills, rather than the result of isolated work from one brilliant employee.

[11] http://www.ibm.com/analytics/us/en/technology/cloud-data-services/data-scientist/
[12] M. Jones (2013) Data science and open source. https://www.ibm.com/developerworks/library/os-datascience/
[13] Writer about technology, entrepreneurs and innovation. Manager Partner at gPress.
[14]http://www.forbes.com/sites/gilpress/2015/10/03/these-are-the-skills-you-need-to-eventually-become-a-240000-unicorn-data-scientist/#7082b7385701

Three main solutions can be applied to deal with the scarcity of unicorns: The first two solutions involve more the inclusiveness inside the company; the third concerns the openness and collaboration out-side the company. The first solution looks within the company and aims to search out those with the skills and competence to create a date scientist team to face this organizational development by identifying key roles and responsibilities, including research leaders, data analysts, and project managers. The second step concerns, after the team building, the necessity to find out how to define areas of responsibility, foster effective and good communication in the team, and build compelling reports. Avoiding pitfalls or losing focus on objectives is necessary. The third solution looks outside the company boundary. This last approach gains the company an open innovation strategy and sets the priority on finding the right partners for collaboration, with the professional skills to fill in the competence missing from within the company's data scientist team (Patil, 2015).

4.3. Privacy, threats or opportunities?

The emerge of IoT enable the collection of data quickly and direct from the source, since it can be gathered from devices as smartphone and tablet. As already said, this ability is a great opportunity to SMEs and large corporations to extract information that will support the innovation process. However, it also poses major questions regarding privacy and security. In this paper, it is not our purpose to look into the implications connected to the privacy issues in Big Data analysis, but just to emphasize the importance of this delicate argument and highlight some possible cooperative and open solutions currently under discussion by legal authorities, policy makers, researchers, and businesses.

Privacy management and brand reputation has in the digital era been a powerful way of increasing or losing several business opportunities (West and Gallagher, 2012). For instance, as Eric von Hippel describes in his book *Democratizing Innovation* (von Hippel, 2005), one of the most powerful key components of innovation development with users is the potential for customers and users to share their experience on a web page, platform or social network. These channels foster the creation of Big Data, generating many unstructured data that are captured by organizations and companies to obtain insights about business opportunities. Getting a bad feedback or providing information from organizations that use data other than for the purposes declared can induce a negative perception of the company itself and alert other users about this abuse. This kind of side effect generated by users can also alert competent authorities to investigate the apparent misconduct in more depth: a sort of real-time crowd monitoring of lawful use of data.

Many authors and scholars have proposed various solutions and applications; nevertheless for our purposes in this paper, where the focus is on business strategy to leverage Open Innovation through Big Data analysis, an interesting approach comes from Tene and Polonetsky (Tene and Polonetsky, 2013a). Their principal solution does not take into consideration traditional legal answers, such as included consent, data minimization, and access. They propose a sort of 'sharing the wealth' strategy on data collection, providing individuals control with access to their data in an easy and usable format. This allows the advantage of applications to analyze users' own data and propose practical use of it. Tene and Polonetsky argue that this application of data will unleash innovation and create new business opportunities (Tene and Polonetsky, 2013b). For companies the logical approach, from the legal threats to the data information-sharing with users, can be more efficient in discovering new business opportunities and avoiding the same kind of privacy threats.

4.4. Intellectual Property an open world

One of the trends stressed above was the possibility to use Big Data analysis to foster Open Innovation strategy by creating communities and contest that allow customers and users to contribute to generate new ideas and solutions. Although this strategy can create a big source of new ideas, is this also an open possibility to competitors to use them. In other words, developing an Open Innovation strategy means sharing technology and information. Therefore, this sharing process needs to set up a management defensive strategy regarding intellectual property or appropriability.

Let us look same examples. Marcel Borges describes the paradox caused by the natural tension between knowledge sharing and protection among companies (Borges, 2011). He has observed how firms can protect their technological competencies and, at the same time, collaborate with other organizations to create and capture value in the era of Open Innovation (McEvily et al.; 2004).

This counterintuitive paradox between Open Innovation agreement and protection has seen an increasing importance for patents when launching technology on the market. The outcome of collaborative agreements depends also on the companies' ability to set up an efficient management defense strategy. Intellectual property rights, for example, seek to obtain value from the R&D investment and effort.

Using creative and efficient methods to exploit firms' IPR defensive strategy is one of the most challenging processes in seeking to enrich a profitable Open Innovation strategy (Schultz and Urban, 2012). Nevertheless, design a good Big Data privacy strategy has become extremely important as IPR defensive strategy has been recognized as an Open Innovation profitable strategy.

In conclusion, if on one hand issues regarding privacy threats in Big Data applications cannot be ignored, on the other hand managers and scholars need to consider how companies can develop an effective defense strategy of sharing information with users to leverage new competitive advantages and business opportunities.

5. Conclusion

In this paper, we described the emerging trends, opportunities, challenges and some practical implications of them, coming from the use of Big Data for Open Innovation strategies. We focalized our investigations on both SMEs and large corporations.

The attention of scholars, policy makers and practitioners for this topic has recently increased but, to the best of our knowledge, our study is the first broad overview on the topic of Big Data analysis for Open Innovation strategies. We reviewed the main academic works published so far and the main industrial applications on this topic.

Our findings provide some managerial implications. Companies should carefully take into consideration what kind of trends areas would be the most valuable according to the type of activities and the core business they

intend to develop. These attentive considerations would lead to the best solution in which Big Data application can drive a profitable Open Innovation strategy able to gain new business opportunities.
Through our analysis, we identified main trends, opportunities and challenges faced by large companies and SMEs. We also describe the situations in which Big Data analysis finds significant and profitable applications to boost the innovation process. In addition, from our study we identify three ways in which Big Data and Open Innovation can create new opportunities to sustain an Open Innovation strategy: i) the creations of new Open Business Model; ii) the Spin-out asset and the secondary markets; and iii) the organizational change. Following the Open Innovation paradigm, these principles must be strengthened and applied more effectively, creatively, and more innovatively.
This paper has provided an entry points to fill the gap in literature of the lack of a comprehensive overview of the use of Big Data for open innovation strategies. Nonetheless, wider synoptic overviews and in-depth empirical studies are required to examine all potential outcomes of the application of Big Data to foster a valuable innovation process through the Open Innovation paradigm. The fast growth rate of technological techniques and the “young” field of Big Data represent a limitation per se. Data sources, tools and approach in practical implications need more systematic research to be well define. There is no field of application where Big Data cannot be applied, and the same is true for the Open Innovation. This paper does not investigate potential new proactive collaboration or actual business process within the same companies, deriving from the high fragmentation. Data business process enables organizations to identify capabilities and roles required to ensure successful outcomes. This aspect needs future investigation. Recently approved regulations[15] in terms of privacy might modify the strategy discussed, or create new approaches that we did not consider.
Despite the limitation, this paper enriches the debate on the trends, opportunities and challenges on Big Data framework in Open Innovation paradigm faced by SMEs and large companies; furthermore, it sets up a framework for future research on the topic, particularly on some specific aspects of the technological and organizational dynamics and relationships.
In the contemporary fast-changing world, collaboration with other companies, businesses, governments, researchers and participants in co-creation projects is an extremely powerful way to obtain useful insights and information, to capture hidden value opportunities and create new business models, products, and services for the benefit of both customers and stakeholders.
It is important to notice that firms also need to be focused and receive a profitable return on their investment and efforts in an innovation process strategy. This essential commercial factor highlights an enigmatic phenomenon called paradox of openness (Laursen & Salter, 2014). Essentially, the return of efforts spread over the innovation process needs to be closed in the commercialization and, hopefully, the monetization phases. This end return of efforts needs to be protected by the companies involved.
This appropriability implication under the Big Data domain needs to be further investigated by scholars, researchers, and businesses.

[15] For instance, the EU Directive N° 697/2016 on Data Protection and Privacy.

Another important issue that requires future investigation is the effect of Big Data analysis and Open Innovation in the economies of scale. In general, large companies are more able to obtain economies of scale from their size and increase efficiency better than the SMEs. The expansions of the economies of scale can positively impact at any stage of productivity and reduce the average cost production.

SMEs more often encounter difficulties in implementing economies of scale. Developing the size of the company involves investment and higher cost. Obtaining financial and economic resources is more difficult for SMEs than large corporations. Vulnerability and the need to remain competitive when trading conditions became more challenging is also more complicated for SMEs, because they don't have internal resources to draw on when the economy takes a turn for worse. These are just some examples of how economies of scale are generally more complicated to implement for SMEs (Vossen, 1998).

Nevertheless, in the digital era when the agility of companies, fewer bureaucratic procedures and attention to the customer's needs are of prime concern, this "agility" can become one of the most powerful tools for SMEs to remain in the market and create revenue and profits. This market agility strategy can very well "compensate" for the loss of ability to create economies of scale like the large companies.

References

Abbasi, F. Simunek, J. Van Genuchten, M. T. Feyen, J. Adamsen, F. J. Hunsaker, D. J. Shouse, P. (2003). Overland water flow and solute transport: Model development and field-data analysis. Journal of Irrigation and Drainage Engineering, 129(2), 71-81.

Asplund, C. (2012). Beyond “Triple Helix” – towards “Quad Helix.” The Bearing Wave, Bearing Consulting Ltd.

Atzori, L. Iera, A. and Morabito, G. (2010). The Internet of Things: A survey. Computer Networks 54, 2787-2805.

Baden-Fuller, C. and Haefliger, S. (2013). Business Models and Technological Innovation”. Long Range Planning 46; 419–426.

Bandyopadhyay, D. Sen, J. (2011). Internet of Things: Applications and Challenges in Technology and Standardization. Wireless Personal Communications 58, 49-69.

Battistella, C. and Nonino, F. (2012). Open innovation web-based platforms: The impact of different forms of motivation on collaboration. Innovation, 14(4), 557-575.

Bi, Z. Xu, L.D. and Wang, C. (2014). Internet of Things for Enterprise Systems of Modern Manufacturing. IEEE Transactions on Industrial Informatics 10, 1537-1546.

Bianchi, M. Cavaliere, A. Chiaroni, D. Frattini, F. Chiesa, V. (2011). Organisational modes for Open Innovation in the bio-pharmaceutical industry: An exploratory analysis. Technovation, 31(1), 22-33.

Bifet, A. Holmes, G. Kirkby, R. and Pfahringer, B. (2010). Moa: Massive online analysis. The Journal of Machine Learning Research, vol. 11, pp. 1601–1604.

Bigliardi, B. and Galati, F. (2013). Models of adoption of open innovation within the food industry. Trends in Food Science & Technology, 30(1), 16-26.

Bogers, M. (2011). "The open innovation paradox: knowledge sharing and protection in R&D collaborations", European Journal of Innovation Management, Vol. 14 Iss 1 pp. 93 – 117;(2011).

Brown, M. S. (2014). Big Data, Mining, and Analytics. Components of Strategic Decision Making. In S. Kudyba (Ed.), Big Data, Mining, and Analytics. Components of Strategic Decision Making (pp. 211-230). Boca Raton: CRC Press Taylor & Francis Group.

Bughin, J. Chui, M. and Manyika, J. (2010) Clouds, big data, and smart assets: Ten tech-enabled business trends to watch, McKinsey Quarterly, August 2010.

Bughin, J. Chui, M. and Miller, A. (2009) How companies are benefiting from Web 2.0: McKinsey Global Survey results, McKinsey Quarterly , September 2009.

Chandler, N. (2014). Business Intelligence and Performance Management Key Initiative Overview. Retrieved from: https://www.gartner.com/doc/2715117/business-intelligence-performance-management-key.

Chen, H. Chiang, R. H. L. Lindner, C. H. Storey, V. C. and Robinson, J. M. (2012);a. Business intelligence and analytics: From Big Data to Big Impact. MIS Quarterly Vol. 36 No. 4, pp. 1165-1188.

Chen, H. Chiang, R. H. L. Storey, V. C. (2012);b. Business intelligence and analytics: from big data to big impact. MIS Quarterly, Volume 36 Issue 4. Pg 1165-1188.

Chesbrough, H. (2003). Open innovation; The new imperative for creating and profiting from technology. Harvard Business School Press, Harvard: Boston.

Chesbrough, H. (2006). Open innovation: a new paradigm for understanding industrial innovation. Open innovation: Researching a new paradigm, 1-12.

Christensen, J. F. Olesen, M. H. and Kjær, J. S. (2005). The industrial dynamics of Open Innovation—Evidence from the transformation of consumer electronics. Research policy, 34(10), 1533-1549.

Christensen, K. S. (2004). A classification of the corporate entrepreneurship umbrella: labels and perspectives. International Journal of Management and Enterprise Development, 1(4), 301-315.

Cohen, W. and Levinthal, D. (1990). Absorptive Capacity: A new perspective on learning and innovation. Administrative Science Quarterly, 35: pp. 128-152.

Curley, M. (2016). Twelve principles for open innovation 2.0. Nature, Vol 533; Macmillan Publishers.

Curley, M. and Salmelin, B. (2013). Open Innovation 2.0 – A New Paradigm. (EU Open Innovation and Strategy Group).

Dahlander, L. and Gann, D. M. (2010). How open is innovation?. Research policy, 39(6), 699-709.

De Mauro, A. Greco, M.. and Grimaldi, M. (2016). A Formal Definition of Big Data Based on its Essential Features, Library Review, 65 (3), 122-135.

Dean, J. and Ghemawat, S. (2008). "MapReduce: simplified data processing on large clusters". Commun ACM, vol. 51, no. 1, pp. 107-113.

Del Vecchio, P. Passiante, G. Vitulano, F. and Zampetti, L. (2014). Big Data and Knowledge-intensive entrepreneurship: trends and opportunities in the tourism sector. Electronic Journal of Applied Statistical Analysis: Decision Support Systems and Services Evaluation, 5(1), 12-30.

Dobre, C. and Xhafa, F. (2014). "Parallel Programming Paradigms and Frameworks in Big Data Era," Int. J. Parallel Program., vol. 42, no. 5, pp. 710–738.

Draxler, H. N. Durmusoglu, S. S. and Griffin, A. (2014). Boosting Open Innovation by Leveraging Big Data, in "Open Innovation: New Product Development Essentials from the PDMA", Wiley Online Library, 299-318.

Dyché, J. and Levy, E. (2011).Customer Data Integration: Reaching a Single Version of the Truth. John Wiley & Sons.

Ebner, W. Leimeister, J. M. and Krcmar, H. (2009). Community engineering for innovations: the ideas competition as a method to nurture a virtual community for innovations. R&d Management, 39(4), 342-356.

Enkel, E. Gassmann, O. and H Chesbrough. (2009). Open R&D and open innovation:Exploring the phenomenon.R&D Management Journal, 39(4), 311–316.

Essentials from the PDMA. Wiley. Published by John Wiley & Sons, Inc.

European Parliament and the European Council: REGULATION (EU) 2016/679; 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

Feki, M.A. Kawsar, F. Boussard, M. and Trappeniers, L. (2013). The internet of things: the next technological revolution. Computer 46, 24-25.

Gandomi, A. and Haider, M., (2015), "Beyond the hype: Big data concepts, methods, and analytics", International Journal of Information Management, 35, pp. 137-144.

Garavelli, A.C. Messeni Petruzzelli, A. Natalicchio, A. and Vanhaverbeke, W. (2013), Benefiting from markets for ideas — an investigation across different typologies, International Journal of Innovation Management, 17 (6).

Gassmann, O. and Enkel, E. (2004), "Towards a Theory of Open Innovation: Three Core Process Archetypes." In: Proceedings of the R&D Management Conference (RADMA).

Gassmann, O. Enkel, E. and Chesbrough, H. (2010). The future of open innovation. R&d Management, 40(3), 213-221.

Gatignon, H. et al (2002). A Structural Approach to Assessing Innovation: Construct development of Innovation Locus, Type, and Characteristics. Management Science, 48: 1103-1122.

Gerard, G. Ernst, C. O., Dovev, L. and Brent A. S. (2016) Big Data And Data Science Methods For Management Research, Academy of Management Journal, 29 (5), 1493-1507.

Gerry G. Haas, M.R. and Pentland, A. (2014) Big Data and Management – Editorial. Academy of Management Journal, 57 (2), 321-326.

Gilsing, V. A. Van Burg, E. and Romme, A. G. L. (2010). Policy principles for the creation and success of corporate and academic spin-offs. Technovation, 30(1), 12-23.

Griffin, A. Noble, C. Durmusoglu, S. (2014). Open Innovation: New Product Development.

Groves, P., Kayyali, B., Knott, D., & Kuiken, S. V. (2016). The'big data'revolution in healthcare: Accelerating value and innovation.

Harbison, J.R. and Pekar, P. (1998). Smart Alliances. A Practical Guide to Repeatable Success. Jossey-Bass. Ricerca di Booz. Allen & Hamilton.

Harding, J.A. Swarnkar, R. (2013). Implementing collaboration moderator service to support various phases of virtual organisations. International Journal of Production Research 51: 7372-7387.

Hendrickson, S. (2010). Getting Started with Hadoop with Amazon's Elastic MapReduce. EMR.

Hilbert, M. and L´opez, P. (2011).The world's technological capacity to store, communicate, and compute information. Science, vol. 332, no. 6025, pp. 60–65.

Hsiao-Kang, Lin. and Chun-I Chen. (2015). Big Data Hadoop for Cloud-Based Enterprise Collaboration Systems. International Journal of Economics & Management Sciences. Volume 4;Issue 10.

Hsiao-Kang, Lin. Hardingb, J. A. Chun-I Chen. (2016). A Hyperconnected Manufacturing Collaboration System Using the Semantic Web and Hadoop Ecosystem System. Published by Elsevier in Changeable, Agile, Reconfigurable & Virtual Production Conference.

Hughes, B. and Wareham, J. (2010). Knowledge arbitrage in global pharma: a synthetic view of absorptive capacity and open innovation. R&d Management, 40(3), 324-343.

IDC.(2014).Analyze the future.http://www.idc.com/

Ili, S. Albers, A. and Miller, S. (2010). Open innovation in the automotive industry. R&D Management, 40(3), 246-255.

Intel,BigDataAnalytics,(2012).http://www.intel.com/content/dam/www/public/us/en/documents/reports/data-insights-peer-research-report.pdf.

Išoraitė, M.(2009). Importance of strategic alliances in company's activity. INTELLECTUAL ECONOMICS, No. 1(5), p. 39–46.

Jean-Pierre Segers. (2015). The interplay between new technology based firms, strategic alliances and open innovation, within a regional systems of innovation context. The case of the biotechnology cluster in Belgium. Journal of Global Entrepreneurship.

Jin, X. Wah, B. W. Cheng, X. and Wang, Y. (2015). Significance and challenges of big data research. Big Data Research, 2: 59-64.

Jones, M. (2013). Data science and open source. Accessed, 15 October 2016; SN: https://www.ibm.com/developerworks/library/os-datascience/

Kaisler, S. Armour, F. Espinosa, J. A. and Money, W. (2013). Big data: issues and challenges moving forward. In System Sciences (HICSS), 2013 46th Hawaii International Conference on (pp. 995-1004). IEEE.

Kale, P. and Singh, H. (2009). Managing Strategic Alliances: What Do We Know Now, and Where Do We Go From Here?. Academy of Management Perspectives, (2009).

Kenney, M. and Zysman, J. (2016). The Rise of the Platform Economy. Issues in Science and Technology 32, no. 3.

Khan, N. and Yaqoob, I. et all. (2014).Big Data: Survey, Technologies, Opportunities, and Challenges. Hindawi Publishing Corporation, The Scientific World Journal. Article ID 712826.

Kutvonen, A. (2011). Strategic application of outbound open innovation. European Journal of Innovation Management.

Laney, D. (2001). 3-D data management: Controlling data volume, velocity and variety, Application Delivery Strategies by META Group Inc. (2001, February 6), p. 949 Retrieved from http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf.

Laursen, K. and Salter, A.J. (2014). The paradox of openness: Appropriability, external search and collaboration. Elsevier - Research Policy 43;(2014).

Laursen, K., and Salter, A. (2006). Open for innovation: the role of openness in explaining innovation performance among UK manufacturing firms. Strategic management journal, 27(2), 131-150.

Lazzarotti, V. and Manzini, R. (2009). Different modes of open innovation: a theoretical framework and an empirical study. International journal of innovation management, 13(04), 615-636.

Lichtenthaler, U. (2009). Outbound open innovation and its effect on firm performance: examining environmental influences. R&D Management 39, 4.

Lohr, S. (2012). The Age of Big data,. The New York Times, SundayReview. Accessed, 3 October 2016; SN: http://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html

Loukides, M. (2010). What is Data Science? The future belongs to the companies and people that turn data into products. O'Reilly Radar Rep.

Manning, P. (2013). Big data in history. Palgrave Macmillan, UK.

Marcia. (2012). Data on Big Data. http://marciaconner.com/blog/data-on-big-data/.

Marr. B. (2016). Big Data in Practice: How 45 successful companies used Big Data Analytics to deliver extraordinary results.

Marston, S. Zhi Li. Bandyopadhyay, S. Zhang, J. and Ghalsasi, A. (2001). Cloud computing - The business perspective. Decision Support Systems, (51).

Mayer-Schönberger, V. and Cukier, K. (2013). Big data: A revolution that will transform how we live, work, and think. Houghton Mifflin Harcourt.

McAfee, A. and Brynjolfsson, E. (2012). "Big Data: The Management Revolution", Harvard Business Review, October 2012, pp.61-68.

McCraw, T. (2007). Prophet of Innovation: Joseph Schumpeter And Creative Destruction, Harvard University Press.McEvily, S.K. Eisenhardt, K.M. and Prescott, J.E. (2004). The global acquisition, leverage, and protection of technological competencies. Strategic Management Journal, Vol. 25 Nos 8/9, pp. 713-22.

McKinsey (2013), "http://www.mckinsey.com/business-functions/business-technology/our-insights/making-data-analytics-work".

Natalicchio, A. Messeni Petruzzelli, A. and Garavelli, A. (2014), A literature review on markets for ideas: Emerging characteristics and unanswered questions, Technovation, 34, 65-76.

Ndou, V. Vecchio, P. D. and Schina, L. (2011). Open Innovation Networks: The Role Of Innovative Marketplaces For Small And Medium Enterprises' Value Creation. International Journal of Innovation and Technology Management, 8(03), 437-453.

Ohlhorst, F. J. (2012). Wiley and SAS Business Series: Big Data Analytics: Turning Big Data into Big Money. Somerset, NJ, USA: John Wiley & Sons.

Ooms, W. Bell, J. and Kok, R.A.W. (2015). Use of Social Media in Inbound Open Innovation: Building Capabilities for Absorptive Capacity. Creativity and Innovation Management, 24(1), 136-150.

Open Innovation 2.0 Yearbook - edition 2016. EUROPEAN COMMISSION. Source: https://ec.europa.eu/digital-single-market/en/news/open-innovation-20-yearbook-edition-2016/

Panniello, U. Gorgoglione, M. and Tuzhilin, A. (2016a). In CARSs we trust: how context-aware recommendations affect customers' trust and other business performance measures of recommender systems. Information Systems Research, 27, 182–196.

Panniello, U. Shawndra, H. and Gorgoglione, M. (2016b). The impact of profit incentives on the relevance of online recommendations, Electronic Commerce Research and Applications, Forthcoming.

Patil, D.J (2015)."Building data science teams": Data science teams need people with the skills and curiosity to ask the big questions. https://www.oreilly.com/ideas/building-data-science-teams

Sarkar, S. and Costa, A. I. (2008). Dynamics of open innovation in the food industry. Trends in Food Science & Technology, 19(11), 574-580.

Schultz, J. and Urban, J. M.(2012). Protecting Open Innovation: The defensive patent, license as a new approach to patent threats, transaction cost and tactical disarmament. Harvard Journal of Law & Technology, Volume 26, Number 1.

Sebastiani, F. (2002). Machine learning in automated text categorization. ACM computing surveys (CSUR), 34(1), 1-47.

Secundo , G. Dumay, J. Schiuma, G. and Passiante, G. (2016) "Managing intellectual capital through a collective intelligence approach: An integrated framework for universities", Journal of Intellectual Capital, Vol. 17 Iss: 2, pp.298 – 319.

Secundo, G. Del Vecchio, P. Dumay, J. and Passiante, G. (2017), "Intellectual capital in the age of Big Data: Establishing a research agenda", Journal of Intellectual Capital, 18, (2), (forthcoming – April 2017).

Secundo, G. Del Vecchio, P. Schiuma, G. and Passiante G. (forthcoming) "Activating entrepreneurial learning processes for transforming university students' idea into entrepreneurial practices", paper accepted for publication on the International Journal of Entrepreneurial Behaviour & Research, Special Issue "Entrepreneurial learning Dynamics in Knowledge Intensive Entreprises".

Shvachko, K. Kuang, H. Radia, S. and Chansler, R. (2010). Mass Storage Syst. Technol. MSST2010 in 2010 IEEE 26th Symp.

Solesvik, M. and Westhead, P. (2010). Partner selection for strategic alliances: case study insights from the maritime industry. Industrial Management & Data Systems, 110(6), 841-860.

Song, I. and Zhu, Y. (2016). Big data and data science: what should we teach?. Wiley Publishing Ltd. Expert Systems, Vol. 33, No. 4.

Spithoven, A. Vanhaverbeke, W. and Roijakkers, N. (2013). Open innovation practices in SMEs and large enterprises. Small Business Economics, 41(3), 537-562.

Tene, O. and Polonetsky, J. (2013)a."Judged by the Tin Man: Empowering Individuals in an Age of Big Data". J.TELECOM&HIGH TECH.L.351.

Tene, O. and Polonetsky, J. (2013)b. Big Data for All: Privacy and User Control in the Age of Analytics. Northwestern Journal of Technology and Intellectual Property J. Tech. & Intell. Prop 2013 vol: 11 (239).

Uddin, M. and Lecturer, B. (2011). Strategic Alliance and competitiveness: Theoretical framework”. Journal of Arts Science & Commerce. International Refereed Research Journal. Vol.– II, Issue 1,(2011).

Vajjhala, N.R. and Ramollari, E. (2016). Big Data using Cloud Computing – Opportunities for Small and Medium-sized Enterprises. European Journal of Economics and Business Studies, 4(1), 129-137.

Van de Vrande, V. de Jong, J.P.J. Vanhaverbeke, W. and de Rochemont, M. (2009). Open Innovation in SMEs: trends, motives and management challenges. Technovation, 29, 423-437.

Von Hippel, E. (2005). “Democratizing Innovation”, MIT Press.

Vossen, R.W. (1998). Relative strengths and weaknesses of small firms in innovation. International Small Business Journal, vol. 16, no. 3.

Ward, J. S. and Barker, A. (2013). Undefined by data: a survey of big data definitions. arXiv preprint arXiv:1309.5821.

Weill, P. Malone, T. W. D’Urso, V. T. Herman, G. and Woerner, S. (2005).Do Some Business Models Perform Better than Others? A Study of the 1000 Largest US Firms. MIT Center for Coordination Science Working Paper No. 226.

West, J. and Gallagher, S. (2006). Challenges of open innovation: the paradox of firm investment in Open source software. R&D Management 36, 3.

White, T. (2012). Hadoop: The Definitive Guide. Sebastopol, CA: O’Reilly Media, Inc.

Whitmore, A. Agarwal, A. and Da Xu, L. (2015). The Internet of Things - A survey of topics and trends. Information Systems Frontiers 17, 261-274.

Wu, X. Zhu, X. Wu, G. Q. and Ding, W. (2014). Data mining with Big Data. IEEE transactions on knowledge and data engineering, Vol. 26 No. 1,pp. 97-107.

Xu, Z. Frankwick, G.L. and Ramirez, E. (2015). Effect of big data analytics and traditional marketing analytics on new product success: a knowledge fusion perspective. Journal of Business Research.