ARTICLE

Big data for open innovation in SMEs and large corporations: Trends, opportunities, and challenges

Pasquale Del Vecchio[1] (iD) | **Alberto Di Minin[2]** | **Antonio Messeni Petruzzelli[3]** (iD) | **Umberto Panniello[3]** (iD) | **Salvatore Pirri[2]** (iD)

[1] Department of Engineering for Innovation, University of Salento, Lecce, Italy

[2] Department of Management, Scuola Superiore Sant'Anna, Pisa, Italy

[3] Politecnico di Bari, Bari, Italy

Correspondence
Umberto Panniello, viale Japigia 182, Politecnico di Bari, Bari, Italy 70125.
Email: umberto.panniello@poliba.it

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, no prior works have 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: small and medium-sized enterprises (SMEs) and big corporations, reviewing the major academic works published so far and analysing 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 & López, 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 ZB. To put it in perspective, 40 ZB is 40 trillion GB, 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, Chiang, Lindner, Storey, & Robinson, 2012).

In such a 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 & Pekar, 1998; Uddin & Lecturer, 2011). The main advantage is a better response to the challenges of globalization, complexity and uncertain business environments (Išoraitė, 2009; Kale & Singh, 2009).

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 storage space and time to be analysed (Kaisler, Armour, Espinosa, & Money, 2013; Ward & Barker, 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 analyse Big Data for better insights and help on effective decision making (Chen et al., 2012). 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 from 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 co-operative. 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 analyse 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 & Ghemawat, 2008). The following year, another IT company, Yahoo, released an open-source tool based on parallel computing (Dobre & Xhafa, 2014). This new tool was able to execute algorithm tasks together on a cluster of machines or supercomputer infrastructures to manage Big Data tasks. The tool was called Hadoop (White, 2012). Since the Yahoo implementation, other companies (e.g., Facebook, LinkedIn, eBay, Hortonworks and Cloudera) have contributed to the Hadoop project, making it, as we know it today, the Apache Hadoop framework.[2]

The Hadoop ecosystem can support 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 process and tight collaboration among different stakeholders around the globe (Lin & Chen, 2015). Hadoop presents a different open-source project and tools that provide various adaptable services, like on-demand workspace, interaction, information sharing or collective problem solving, turning them into its big advantage. That means, for each project, tools or problems, it is possible to find a community of developers, experts and users to ask questions of, fix bugs and implement new features.

Similar practical examples of this kind of collaboration and co-creations processes are represented by the Cloudera and Cask strategic[3] partnership. Both companies aim to create better business value by turning data analytics directly into action using Hadoop to help customers and stakeholders overcome the challenges in setting up new applications, and to accelerate value creation from operational analytics. Another interesting practical case of collaboration using the Hadoop framework is in the manufacturing sector (Lin, Harding, & Chen, 2016). Integration across a hyperconnected manufacturing collaboration system to reduce the complexity of extracting, processing and analysing information, data and products, generates new opportunities and present new challenges for businesses across every industry sector (Harding & Swarnkar, 2013).

The birth of the Hadoop framework shows how 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 into 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, the birth of the open platform for Big Data analysis can very well sustain an open innovation strategy, can be useful in observing 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 address this issue, 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 background literature structured into two main subsections: a critical review of the literature related to Big Data, and a review of the literature on open innovation in the light of Big Data. Then we present an overview of the main trends, opportunities and challenges stemming from the use of Big Data for open innovation strategies. Then some practical implications of 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

This section is structured into two main subsections focused on the topics of Big Data and open innovation. With the aim of providing a comprehensive reading of the fragmented debate on Big Data, resulting from the merging of Business Management and Information System backgrounds, the first subsection presents a critical review of the literature related to Big Data, to ensure the widest understanding 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 subsection reviews the literature on open innovation in 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 & Haider, 2015). Overcoming the ubiquity of the word (De

Mauro, Greco, & Grimaldi, 2016; Ward & Barker, 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*, and regarded as a manifesto of the Big Data movement, McAfee and Brynjolfsson (2012) claimed that Big Data are the instigators of a new millennium industrial revolution, with a large set of unexplored implications and meaning. Looking at the 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).

Regarded 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 storage space and time to be analysed (Kaisler et al., 2013; Ward & Barker, 2013).

Big Data can be seen as the result of an evolutionary process in the field of IT that started in the 1960s by moving from the phases of data processing (1960s) to information about applications (1970s–1980s), from the emergence of decision support models (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 in 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.

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, Li, Bandyopadhyay, Zhang, & Ghalsasi, 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 in storage costs 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 & 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.

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 & Brynjolfsson, 2012). As result of a perspective more focused on the dimensions of ICT, volume is related to the storage space required by servers and databases, currently estimated in exabytes (10^{18}), although this is a trend in continuous growth (Anshari & Lim, 2016; Kaisler et al., 2013). Velocity, identified as the speed of creation, sharing and storage (Kaisler et al., 2013) is also representative of the obsolescence of the data available that in some industries are more meaningful of the volume (McAfee & Brynjolfsson, 2012). Variety refers to the fragmentation of types of data available that can be text, images, videos, audio, etc. (Kaisler et al., 2013; McAfee & Brynjolfsson, 2012).

In addition to these, a second generation of Vs has been identified that presents managerial challenges associated with Big Data. Aimed at comprehending how to transform data into valuable inputs to enhance companies' competitiveness, the Vs of veracity, variability and value have more recently been assumed as new challengeable dimensions with larger managerial implications (Gandomi & Haider, 2015). Veracity in Big Data recalls the need to assure reliable and confident interpretation of data. Variability is 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 & 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 improvement of overall business performance (McAfee & Brynjolfsson, 2012). Deepening the value dimension, critical issues concern the heterogeneity, accuracy and scalability of unstructured data and their integration process (Bifet, Holmes, Kirkby, & Pfahringer, 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 ensure high responsiveness and timely analytical performance (George, Haas, & Pentland, 2014). Despite the intuition related to the benefits of Big Data management in the context of both SMEs and large corporations, its practice continues to be populated by success stories of larger copanies (Marr, 2016), which have the advantage of more sophisticated human and organizational structures.

TABLE 1 Growth rate of unstructured data

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

Source: Khan et al. (2014).

There are no industries that remain uninterested in Big Data, nor are there 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, Wah, Cheng, & Wang, 2015).

The ability of firms to aggregate, elaborate and analyse the data is becoming a key competitive advantage and resource. However, few studies have provided evidence of the role and the outcomes that Big Data analysis could bring to firms in terms of innovation, efficiency, productivity, quality and customer satisfaction (Groves, Kayyali, Knott, & Kuiken, 2016).

The review of the literature highlights the fragmentation of contributions characterizing the debate on Big Data and the need for 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 assets as the main challenge for companies (De Mauro et al., 2016). However, the complexity of the phenomenon as well as the need to deepen its implications for the value creation of companies, call for much greater understanding.

With the aim to contribute to this goal, Table 2 presents a synthesis of the different perspectives of studies resulting from the critical review of the literature, in terms of research focus, description and main references. The critical review of the literature summarized in Table 2 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 of the challenges associated with the Vs of volume, variety and velocity, as the 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 to the value creation process in several ways: making decision making more effective (Kaisler et al., 2013), the improvement

of companies' overall performance (McAfee & Brynjolfsson, 2012), the renewal of companies' intangible assets (Secundo, Dumay, Schiuma, & Passiante, 2016; Secundo, Del Vecchio, Dumay, et al., 2017), the better positioning in the industry (Kaisler et al., 2013), higher customer satisfaction (Brown, Chui, & Manyika, 2011), marketing analytics (Xu, Frankwick, & Ramirez, 2015), innovation in products (Mayer-Schönberger & 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 ensure wider understanding of this implication, Section 2.2 presents a critical reading of the 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 the Big Data approach can result in a plurality of elements to assure the competitive positioning of companies. Focusing on the meanings previously mentioned, the most challenging 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 arouse the interest for scholars and researchers in the field of innovation management (Gassmann, Enkel, & Chesbrough, 2010; Spithoven, Vanhaverbeke, & Roijakkers, 2013).

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

TABLE 2 An overview of the literature on Big Data

Focus	Description	Main References
Form and nature of data	-Structured, semi-structured and unstructured. Text, video, audio, pictures, codes, satellite images, etc.	Secundo, Del Vecchio, Dumay, & Passiante, 2017; Khan et al., 2014; Hurwitz, Nugent, Halper, & Kaufman, 2013; Kaisler et al., 2013; McAfee & Brynjolfsson, 2012; Ohlhorst, 2012
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, Xu, & Wang, 2014; Feki, Kawsar, Boussard, & Trappeniers, 2013; Chen et al., 2012
Tools	-Proprietary (SAS, SPSS, etc.) vs Open source (Hadoop, R, etc.) tools for data extraction, storage, processing, etc.	De Mauro et al., 2016; Hashem et al., 2015; Kaisler et al., 2013; Shvachko, Kuang, Radia, & Chansler, 2010
Approaches	-Machine learning, business analytics, semantic clustering, social network analysis, data mining, etc.	De Mauro et al., 2016: George et al., 2014; Wu, Zhu, Wu, & Ding, 2014; Chen et al., 2012; Sebastiani, 2002
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, Del Vecchio, Dumay, et al., 2017; Gandomi & Haider, 2015; Jin et al., 2015; Bi et al., 2014; Del Vecchio, Passiante, Vitulano, & Zampetti, 2014; Mayer-Schönberger & Cukier, 2013; Kaisler et al., 2013; McAfee & Brynjolfsson, 2012; Brown et al., 2011; Abbasi et al., 2003; Laney, 2001

traditional forms of research and development performed by companies (Gassmann et al., 2010). The relevance of Big Data in the debate on open innovation results from a set of elements that make actual the same paradigm of innovation through the use of internal and external knowledge flows, identified by Dahlander and Gann (2010), into the social and economic changes occurring in consumption and production, globalization as influencing labour 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 that the process of value creation cannot be pursued by companies singularly or autonomously (Prahalad & Ramaswamy, 2004; Prahalad & Krishnan, 2008), but results are more effectively achieved by matching the organizational assets with external knowledge and skills in a network perspective (Gulati, Nohria, & Zaheerm, 2000; Iansiti & 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 from the new managerial and analytical approaches available to afford the challenging 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 about good ideas are widely distributed: the absence of a monopoly of ideas, 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 performance, and finally the perishables of intellectual property are all confirmed both in practice and in theory. As a multidimensional phenomenon, examinable along different perspectives (external R&D, mechanisms for intellectual propriety, etc.), open innovation can assume different forms (Dahlander & Gann, 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 industries (Gassmann et al., 2010). The literature on open innovation continues to be largely supported by empirical evidence related to large corporations. Best practice has been identified in the areas of manufacturing (Laursen & Salter, 2006), healthcare (Hughes & Wareham, 2010), pharmaceuticals (Bianchi, Cavaliere, Chiaroni, Frattini, & Chiesa, 2011), automotive (Ili, Albers, & Miller, 2010), and food (Bigliardi & Galati, 2013; Sarkar & Costa, 2008).

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 (Gassmann 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 perform better at introducing 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 organizational and technological limitations to be overcome (Battistella & Nonino, 2012; Ndou, Del Vecchio, & Schina, 2011; van de Vrande, de Jong, Vanhaverbeke, & de Rochemont, 2009). However, in 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 a process of continuous renewal in the context of existing companies (Ebner, Leimeister, & Krcmar, 2009; Lazzarotti & Manzini, 2009; Secundo, Del Vecchio, Schiuma, & Passiante, 2017), as well as in the launch of new entrepreneurial ventures, such as industrial and academic spinoffs, start-ups, etc. (Chesbrough, 2006; Christensen, Olesen, & Kjær, 2005; Gilsing, Van Burg, & Romme, 2010). The first ones should benefit from 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 into the market by benefiting from greater evidence on competitors and customers (Secundo, Del Vecchio, Schiuma, et al., 2017).

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 centres (Zajac & Olsen, 1993; Powell, Koput, & Smith-Doerr, 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 (Chesbrough, 2006; Gatignon, Tushman, Smith, & Anderson, 2002; Hauser 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 & Elmquist, 2011) and opening new entrepreneurial opportunities (Eftekhari & Bogers, 2015). Moreover, it is interesting to understand how Big Data can be leveraged for developing inbound and outbound open innovation strategies (Dahlander & Gann, 2010); how they may impact on the companies' absorptive capacities (Cohen & Levinthal, 1990; Lane et al., 2006; Ooms, Bell, & Kok, 2015) as well as contribute to their competitive positioning (Chesbrough, 2011; Ollila & Elmquist, 2011).

Considering Big Data as an important approach to help companies to maximize their open innovation experience, it is important to be aware of 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 to human resources, privacy threats and intellectual property rights (IPR).

3 | BIG DATA FOR OPEN INNOVATION STRATEGIES

As discussed in the previous sections, recent research has called for investigation into 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 analyse the main industrial applications on this topic.

3.1 | Trends

As 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 analyse 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 analyse 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.

The number of 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, Chui, & Manyika, 2010) listed 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 for pioneers, such as Wikipedia or open source software developers, or the 70 percent of the executives surveyed by Bughin, Chui, and Miller (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 larger and more committed pool of users (George 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, Messeni Petruzzelli, Natalicchio, & Vanhaverbeke, 2013; Natalicchio, Messeni Petruzzelli, & Garavelli, 2014). However, as 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 requiring 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 flavour of the main trends in this topic. In particular, in 2015, Splunk split their Big Data open innovation competition into three categories: fraud/insider threats, social impact, and innovation.[5] The winner of the first category proposed an app that detects insiders and malicious behaviour 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, analyse 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 line in stores or pay services, a customized recommender system for suggesting interesting items to 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 would 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 occurring in the city itself, and an application measuring a 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 larger companies rather than SMEs, as 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 large amounts 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 is a significant trend for SMEs and large corporations as both of them need 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, Iera, & Morabito, 2010; Feki et al., 2013; Whitmore, Agarwal, & Da Xu, 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 & 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 & 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 additional evidence of this opportunity, Engie (formerly 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' project 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.

Drexler, Duh, Kornherr, and Korošak (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 company. 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 travelling, exploring foreign cultures, and visiting exotic countries. When the 'cup of information' model analysed the internal database of the insurance company, the software found that travelling and travel insurance 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 travellers: a package combining classical travel insurance with a new, high performance but low cost diagnostic device, which can considerably reduce the risk involved in 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 thus avoid illness) and, at the same time, gave customers a sense of security. A by-product 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 company had continuously to watch out for emerging technologies by screening patents. Similar to the case of the insurance company, the 'cup of information' model analysed 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 start-ups). 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 start-ups 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 & 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 & 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 to enter into, or improve their position within, a market, and large companies which are looking to be market leaders. Finally, SMEs also have 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, McGuire, Ariker, and Roggendorf (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; Enkel, Gassmann, & Chesbrough, 2009; Gassmann & Enkel, 2004) 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 (George, Osinga, Lavie, & Scott, 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 in order to effectively leverage 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, as these type of skills are difficult to find and, more important, expensive to acquire. In contrast, 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 Drexler 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, Drexler 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 results that are not useful 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 than 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-Schönberger & 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 to 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-Schönberger & 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, Gorgoglione, & Tuzhilin, 2016; Panniello, Hill, & Gorgoglione, 2016). 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-Schönberger & 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 Micheli & 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, Štiglic, & Korošak, 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 as both of them may suffer a reduction in 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 have their own challenges that become bigger and more complex when put together. This challenge is particularly relevant for SMEs as they have few resources for handling this aspect compared 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 of scattering resources. This challenge is particularly relevant for large companies because they may be more complex (in terms of processes and desired outputs) than 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 as both may access 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

The 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 discussions in previous sections.

4.1 | How Big Data can support open innovation strategy

As discussed above, 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 co-operative 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 that give rise to the most valuable opportunities.

Traditionally, the 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 & Haefliger, 2013).

The Open Business model can be focused and lead real-time Big Data processes to produce more and more 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 companies to gain a profitable customer needs predictive strategy. Data integration also enriches 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, Malone, D'Urso, Herman, & Woerner, 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 which is 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 (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 into a new approach and process. This development process is called Open Innovation 2.0 (OI2)[10] (Curley, 2016) (see Table 3). In OI2, innovation happens in ecosystems or networks and emphasizes the diversity of collaboration in an interdisciplinary approach, to improve the chances of creating 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. A few examples of how organizational changes and common innovation patterns can be applied include: (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).

Open Innovation 2.0 strategy in Big Data needs to become a discipline practised 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 (European Commission, 2016).

TABLE 3 How Open Innovation models have evolved: Main differences between Open Innovation and OI2

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

Source: Curley (2016).

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 (Noble, Durmusoglu, & Griffin, 2014).[11]

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 (Chandler, 2014; Loukides, 2010). 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 & Zhu, 2016), the field that comprises 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 and business strategy. Figure 1 illustrates data science and the connection of computer science, maths and statistics, and domain knowledge.

As Figure 1 shows, finding this complete set of skills in one person is extremely complicated. This is why Gil Press[12] defines a data scientist, data science professionals, as the new unicorn.[13] Considering the scarcity of these 'unicorns', it 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 company. 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.

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 outside 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, after the team building, concerns 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
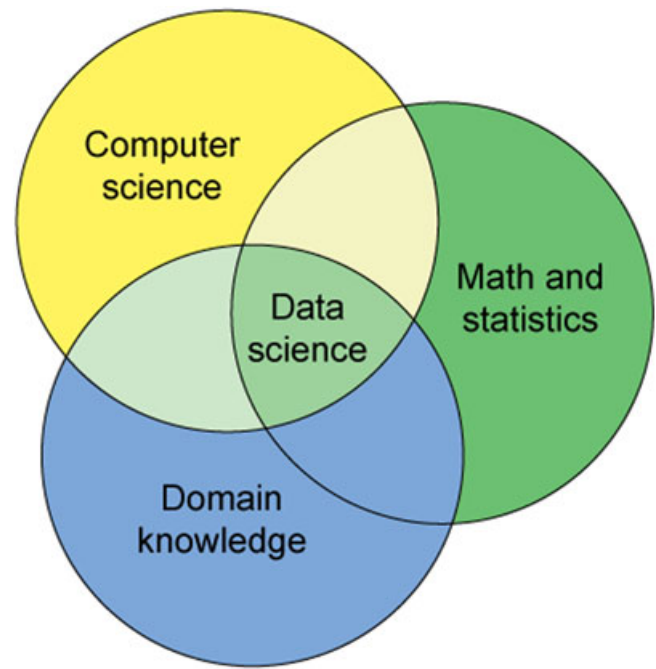
FIGURE 1 Key disciplines of the data scientist Source: Jones (2013) [Colour figure can be viewed at wileyonlinelibrary.com]

necessary. The third solution looks outside the company boundary. This last approach involves the company in an open innovation strategy and sets the priority on finding the right partners for collaboration, with the professional skills to fill in the competences missing from within the company's data scientist team (Patil, 2015).

4.3 | Privacy, threats or opportunities?

The emergence of IoT enables the collection of data quickly and direct from the source, as it can be gathered from devices like smartphones and tablets. As already mentioned, this ability is a great opportunity for 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 co-operative 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 & Gallagher, 2006). 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 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 (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 analyse 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 & Polonetsky, 2013b). For companies, the logical approach, from the legal threats to the sharing of data information 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 contests that allow customers and users to contribute to generate new ideas and solutions. Although this strategy can create a large source of new ideas, there is also the clear possibility for competitors to use them. In other words, developing an open innovation strategy means sharing technology and information. Therefore, this sharing process needs to be set up with a management defensive strategy regarding intellectual property or appropriability.

Let us look at some examples. Marcel Bogers describes the paradox caused by the natural tension between knowledge sharing and protection among companies (Bogers, 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, Eisenhardt, & Prescott, 2004). This counter-intuitive paradox between an 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 & Urban, 2012). Nevertheless, designing a good Big Data privacy strategy has become extremely important as an IPR defensive strategy has become 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 the use of Big Data for open innovation strategies. We focused our investigations on both SMEs and large corporations. The attention of scholars, policy makers and practitioners on 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 trending areas would be the most valuable according to the type of activities and the core business they intend to develop. These considerations should lead to the best solution in which the application of Big Data can drive a profitable open innovation strategy able to gain new business opportunities.

Through our analysis, we identified the main trends, opportunities and challenges faced by large companies and SMEs. We also described 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 creation of a new Open Business Model; (ii) the spin-out asset and the secondary markets; and (iii) through 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 point to fill the gap in the literature by addressing 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 defined. There is no field of application where Big Data cannot be applied, and the same is true for open innovation. This paper does not investigate potential new proactive collaboration or actual business processes within the same companies, deriving from the high fragmentation. Data business processes enable organizations to identify capabilities and roles required to ensure successful outcomes. This aspect needs further investigation. Recently approved regulations (e.g., European Parliament and the European Council, 2016) in terms of privacy might modify the strategy discussed, or create new approaches that we did not consider.

Despite the limitations, this paper enriches the debate on the trends, opportunities and challenges of the Big Data framework in the 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 the 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 problem of possible appropriability under the Big Data domain needs to be further investigated by scholars, researchers and businesses.

Another important issue that requires further investigation is the effect of Big Data analysis and open innovation on economies of scale. In general, large companies are more able to obtain economies of scale from their size and increase efficiency better than SMEs. The expansion of economies of scale can positively impact at any stage of productivity and reduce the average cost of 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 do not have internal resources to draw on when the economy takes a turn for the 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.

ENDNOTES

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

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

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

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

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

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/

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 to co-create and drive structural changes far beyond the scope of what one organization could do alone (Curley & Salmelin, 2013). This new model based on principles of integrated collaboration and co-created shared value also encompasses user-oriented innovation models able to take full advantage of the exploits of disruptive technologies of the digital era – such as cloud computing, Big Data, IoT –leading to experimentation and prototyping in a real-world setting (Curley, 2016). Open Innovation 2.0, unlike Open Innovation, is not just something 'restricted' to the innovation funnel, like the 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.

11 http://www.ibm.com/analytics/us/en/technology/cloud-data-services/data-scientist/

12 Writer about technology, entrepreneurs and innovation. Manager Partner at gPress.

13 http://www.forbes.com/sites/gilpress/2015/10/03/these-are-the-skills-you-need-to-eventually-become-a-240000-unicorn-data-scientist/#7082b7385701

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, 71–81.

Anshari, M., Alas, Y., & Guan, L. (2016). Developing online learning resources: Big data, social networks, and cloud computing to support pervasive knowledge. *Education and Information Technologies*, 21, 1663–1677.

Asplund, C. (2012). *Beyond 'Triple Helix' – Towards 'Quad Helix'.* London: Bearing Consulting.

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

Baden-Fuller, C., & 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., & Nonino, F. (2012). Open innovation web-based platforms: The impact of different forms of motivation on collaboration. *Innovations*, 14, 557–575.

Bi, Z., Xu, L. D., & 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, 22–33.

Bifet, A., Holmes, G., Kirkby, R., & Pfahringer, B. (2010). MOA: Massive online analysis. *Journal of Machine Learning Research*, 11, 1601–1604.

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

Bogers, M. (2011). The open innovation paradox: Knowledge sharing and protection in R&D collaborations. *European Journal of Innovation Management*, 14, 93–117.

Brown, B., Chui, M., & Manyika, J. (2011). Are you ready for the era of 'Big Data'? *McKinsey Quarterly*, 4(1), 24–35.

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

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

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., & Robinson, J. M. (2012). Business intelligence and analytics: From Big Data to Big Impact. *MIS Quarterly*, 36, 1165–1188.

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

Chesbrough, H. (2006). Open innovation: A new paradigm for understanding industrial innovation. In H. Chesbrough, W. Vanhaverbeke, & J. West (Eds.), *Open innovation: Researching a new paradigm* (pp. 1–12). Oxford, UK: Oxford University Press.

Chesbrough, H. W. (2011). Bringing open innovation to services. *MIT Sloan Management Review*, 52(2), 85.

Christensen, J. F., Olesen, M. H., & Kjær, J. S. (2005). The industrial dynamics of Open Innovation—Evidence from the transformation of consumer electronics. *Research Policy*, 34, 1533–1549.

Cohen, W., & Levinthal, D. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly*, 35, 128–152.

Conner, M. (2012). Data on Big Data. Retrieved from http://marciaconner.com/blog/data-on-big-data/

Curley, M. (2016). Twelve principles for open innovation 2.0. *Nature*, 533.

Curley, M., & Salmelin, B. (2013). Open Innovation 2.0 – A new paradigm. EU Open Innovation and Strategy Group.

Dahlander, L., & Gann, D. M. (2010). How open is innovation? *Research Policy*, 39, 699–709.

De Mauro, A., Greco, M., & Grimaldi, M. (2016). A formal definition of Big Data based on its essential features. *Library Review*, 65, 122–135.

De Micheli, C., & Stroppa, A. (2013). Twitter and the underground market. 11th Nexa Lunch Seminar, Turin, 22 May. Retrieved from http://nexa.polito.it/nexacenterfiles/lunch-11-de_micheli-stroppa.pdf

Dean, J., & Ghemawat, S. (2008). MapReduce: Simplified data processing on large clusters. *Communications of the ACM*, 51, 107–113.

Del Vecchio, P., Passiante, G., Vitulano, F., & 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, 12–30.

Dobre, C., & Xhafa, F. (2014). Parallel programming paradigms and frameworks in Big Data era. *International Journal of Parallel Programming*, 42, 710–738.

Drexler, G., Duh, A., Kornherr, A., & Korošak, D. (2014). Boosting open innovation by leveraging big data. In C. H. Noble, S. S. Durmusoglu, & A. Griffin (Eds.), *Open Innovation: New product development essentials from the PDMA* (pp. 299–318). Chichester, UK: John Wiley & Sons.

Duh, A., Štiglic, G., & Korošak, D. (2013). Enhancing identification of opinion spammer groups. In *Proceedings of International Conference on Making Sense of Converging Media*, (pp. 326). New York, New York: ACM.

Ebner, W., Leimeister, J. M., & Krcmar, H. (2009). Community engineering for innovations: The ideas competition as a method to nurture a virtual community for innovations. *R&D Management*, 39, 342–356.

Eftekhari, N., & Bogers, M. (2015). Open for entrepreneurship: How open innovation can foster new venture creation. *Creativity and Innovation Management*, 24(4), 574–584.

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

European Commission (2016). Open Innovation 2.0 Yearbook – edition 2016. Retrieved from https://ec.europa.eu/digital-single-market/en/news/open-innovation-20-yearbook-edition-2016/

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

Fan, W., & Bifet, A. (2013). Mining Big Data: Current status, and forecast to the future. *SIGKDD Explorations*, 14(2), 1–5.

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

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

Garavelli, A. C., Messeni Petruzzelli, A., Natalicchio, A., & Vanhaverbeke, W. (2013). Benefiting from markets for ideas—an investigation across different typologies. *International Journal of Innovation Management*, 17. 1340017

Gartner. (2015). "Hype Cycle: Big Data is out, Machine Learning is in" by Geethika Bhavya Peddibhotla. Retrieved from: http://www.kdnuggets.com/2015/08/gartner-2015-hype-cycle-big-data-is-out-machine-learning-is-in.html

Gassmann, O., & Enkel, E. (2004). Towards a theory of Open Innovation: Three core process archetypes. *Proceedings of the R&D Management Conference (RADMA)*, Lisbon, Portugal.

Gassmann, O., Enkel, E., & Chesbrough, H. (2010). The future of open innovation. *R&D Management*, 40, 213–221.

Gatignon, H., Tushman, M. L., Smith, W., & Anderson, P. (2002). A structural approach to assessing innovation: Construct development of innovation locus, type, and characteristics. *Management Science*, 48, 1103–1122.

George, G., Haas, M. R., & Pentland, A. (2014). From the Editors: Big Data and management. *Academy of Management Journal*, 57, 321–326.

George, G., Osinga, E. C., Lavie, D., & Scott, B. A. (2016). Big Data and data science methods for management research. *Academy of Management Journal*, 59, 1493–1507.

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

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

Gulati, R., Nohria, N. & Zaheer, A. (2000). Strategic networks. *Strategic Management Journal*, 21, 203–215.

Harbison, J. R., & Pekar, P. (1998). *Smart alliances: A practical guide to repeatable success*. San Francisco, CA: Jossey-Bass.

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.

Hashem, I. A., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., Khan, S. U. (2015). The rise of "big data" on cloud computing: Review and open research issues. *Information Systems*, 47, 98–115.

Hendrickson, S. (2010). Getting Started with Hadoop with Amazon's Elastic MapReduce. EMR-Hadoop Meetup, Boulder/Denver, CO.

Hilbert, M., & López, P. (2011). The world's technological capacity to store, communicate, and compute information. *Science*, 332, 60–65.

Hughes, B., & Wareham, J. (2010). Knowledge arbitrage in global pharma: A synthetic view of absorptive capacity and open innovation. *R&D Management*, 40, 324–343.

Hurwitz, J., Nugent, A., Halper, F., & Kaufman, M. (2013). Big data for dummies. NY, USA: John Wiley & Sons.

Iansiti, M., & Levien, R. (2004). The keystone advantage: what the new dynamics of business ecosystems mean for strategy, innovation, and sustainability. Boston, MA: Harvard Business Press.

IDC. (2014). The digital universe of opportunities: Rich data and the increasing value of the Internet of Things. Retrieved from https://www.emc.com/leadership/digital-universe/2014iview/index.htm

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

Intel (2012). Big Data Analytics. Retrieved from 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*, 1(5), 39–46.

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

Jones, M. (2013). Data science and open source. Retrieved from https://www.ibm.com/developerworks/library/os-datascience/index.html

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

Kale, P., & Singh, H. (2009). Managing strategic alliances: What do we know now, and where do we go from here? *Academy of Management Perspectives*, 23, 45–62.

Kenney, M., & Zysman, J. (2016). The rise of the platform economy. *Issues in Science and Technology*, 32(3).

Khan, N., Yaqoob, I., Hashem, I. A. T., Inayat, Z., Ali, W. K. M., Alam, M., … Gani, A. (2014). Big Data: Survey, technologies, opportunities, and challenges. *The Scientific World Journal*. 2014, Article ID 712826

Kutvonen, A. (2011). Strategic application of outbound open innovation. *European Journal of Innovation Management*, 14, 460–474.

Laney, D. (2001). 3-D data management: Controlling data volume, velocity and variety. *Application Delivery Strategies*, File 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., & Salter, A. (2006). Open for innovation: The role of openness in explaining innovation performance among UK manufacturing firms. *Strategic Management Journal*, 27, 131–150.

Laursen, K., & Salter, A. (2014). The paradox of openness: Appropriability, external search and collaboration. *Research Policy*, 43, 867–878.

Lazzarotti, V., & Manzini, R. (2009). Different modes of open innovation: A theoretical framework and an empirical study. *International Journal of Innovation Management*, 13, 615–636.

Lichtenthaler, U. (2009). Outbound open innovation and its effect on firm performance: Examining environmental influences. *R&D Management*, 39, 317–330.

Lin, H.-K., & Chen, C.-I. (2015). Big Data hadoop for cloud-based enterprise collaboration systems. *International Journal of Economics & Management Sciences*, 4, 300.

Lin, H.-K., Harding, J. A., & Chen, C.-I. (2016). A hyperconnected manufacturing collaboration system using the semantic web and Hadoop Ecosystem System. *Procedia CIRP*, 52, 18–23.

Lohr, S. (2012). The age of Big data. *The New York Times, Sunday Review*, February 11.

Loukides, M. (2010). What is data science? The future belongs to the companies and people that turn data into products. Retrieved from https://www.oreilly.com/ideas/what-is-data-science

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

Marr, B. (2016). *Big Data in practice: How 45 successful companies used Big Data Analytics to deliver extraordinary results*. Chichester, UK: Wiley.

Marston, S., Li, Z., Bandyopadhyay, S., Zhang, J., & Ghalsasi, A. (2011). Cloud computing – The business perspective. *Decision Support Systems*, 51, 176–189.

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

McAfee, A., & Brynjolfsson, E. (2012). Big Data: The management revolution. *Harvard Business Review*, 90, 61–68.

McCraw, T. (2007). *Prophet of innovation: Joseph Schumpeter and creative destruction*. Cambridge, MA: Harvard University Press.

McEvily, S. K., Eisenhardt, K. M., & Prescott, J. E. (2004). The global acquisition, leverage, and protection of technological competencies. *Strategic Management Journal*, 25, 713–722.

McGuire, T., Ariker, M., & Roggendorf, M. (2013), Making data analytics work: Three key challenges. McKinsey & Company. Retrieved from http://www.mckinsey.com/business-functions/business-technology/our-insights/making-data-analytics-work

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

Ndou, V., Del Vecchio, P., & 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, 437–453.

Noble, C. H., Durmusoglu, S. S., & Griffin, A. (2014). *Open Innovation: New product development essentials from the PDMA*. Chichester, UK: John Wiley & Sons.

Ohlhorst, F. J. (2012). *Big Data analytics: Turning Big Data into big money*. Hoboken, NJ: John Wiley & Sons.

Ollila, S., & Elmquist, M. (2011). Managing open innovation: Exploring challenges at the interfaces of an open innovation arena. *Creativity and Innovation Management*, 20, 273–283.

Ooms, W., Bell, J., & Kok, R. A. W. (2015). Use of social media in inbound Open Innovation: Building capabilities for absorptive capacity. *Creativity and Innovation Management*, 24, 136–150.

Panniello, U., Gorgoglione, M., & Tuzhilin, A. (2016). 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., Hill, S., & Gorgoglione, M. (2016). The impact of profit incentives on the relevance of online recommendations. *Electronic Commerce Research and Applications*, 20, 87–104.

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

Powell, W. W., Koput, K. W., & Smith-Doerr, L. (1996). Interorganizational collaboration and the locus of innovation: Networks of learning in biotechnology. *Administrative Science Quarterly*, 41(1), 116–145.

Prahalad, C. K., & Ramaswamy, V. (2004). Co-creation experiences: The next practice in value creation. *Journal of Interactive Marketing*, 18(3), 5–14.

Prahalad, C. K., & Krishnan, M. S. (2008). *The new age of innovation: Driving cocreated value through global networks*. New York, NY: McGraw-Hill.

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

Schultz, J., & 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*, 26, 1–69.

Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM Computing Surveys (CSUR)*, 34, 1–47.

Secundo, G., Del Vecchio, P., Dumay, J., & Passiante, G. (2017). Intellectual capital in the age of Big Data: Establishing a research agenda. *Journal of Intellectual Capital*, 18, 242–261.

Secundo, G., Del Vecchio, P., Schiuma, G., & Passiante, G. (2017). Activating entrepreneurial learning processes for transforming university students' idea into entrepreneurial practices. *International Journal of Entrepreneurial Behavior & Research*, 23, 465–485.

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

Segers, J.-P. (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*, 5, 1–17.

Shvachko, K., Kuang, H., Radia, S., & Chansler, R. (2010). The Hadoop distributed file system. MSST'10: *Proceedings of the 2010 IEEE 26th Symposium on Mass Storage Systems and Technologies (MSST)*. Piscataway, NJ: IEEE.

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

Song, I., & Zhu, Y. (2016). Big data and data science: What should we teach? *Expert Systems*, 33, 364–373.

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

Tene, O., & Polonetsky, J. (2013a). Judged by the Tin Man: Individual rights in the age of Big Data. *Journal of Telecommunications and High Technology Law*, 11, 351–368.

Tene, O., & Polonetsky, J. (2013b). Big Data for all: Privacy and user control in the age of analytics. *Northwestern Journal of Technology and Intellectual Property*, 11, 239–273.

Uddin, M., & Lecturer, B. (2011). Strategic alliance and competitiveness: Theoretical framework. *Journal of Arts Science & Commerce*, II, 43–54.

Vajjhala, N. R., & Ramollari, E. (2016). Big Data using cloud computing – Opportunities for small and medium-sized enterprises. *European Journal of Economics and Business Studies*, 4, 129–137.

van de Vrande, V., de Jong, J. P. J., Vanhaverbeke, W., & de Rochemont, M. (2009). Open Innovation in SMEs: Trends, motives and management challenges. *Technovation*, 29, 423–437.

Vance, A. (2011). The data knows. *Bloomberg Business week*, pp. 70–74.

von Hippel, E. (2005). *Democratizing innovation*. Cambridge, MA: MIT Press.

Vossen, R. W. (1998). Relative strengths and weaknesses of small firms in innovation. *International Small Business Journal*, 16, 88–94.

Ward, J. S., & 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., & 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., & Gallagher, S. (2006). Challenges of open innovation: The paradox of firm investment in open source software. *R&D Management*, 36, 319–331.

White, T. (2012). *Hadoop: The definitive guide*. Sebastopol, CA: O'Reilly Media.

Whitmore, A., Agarwal, A., & 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., & Ding, W. (2014). Data mining with Big Data. *IEEE Transactions on Knowledge and Data Engineering*, 26, 97–107.

Xu, Z., Frankwick, G. L., & Ramirez, E. (2015). Effect of big data analytics and traditional marketing analytics on new product success: A knowledge fusion perspective. *Journal of Business Research*, 69, 1562–1566.

Zajac, E. J., & Olsen, C. P. (1993). From transaction cost to transactional value analysis: Implications for the study of interorganizational strategies. *Journal of Management Studies*, 30, 131–145.

Pasquale Del Vecchio is Researcher and Lecturer in the Department of Engineering for Innovation at the University of Salento, Italy. He gained his PhD in eBusiness Management at the University of Salento in 2008. In 2007, he spent a period as visiting PhD student in the Center for Collective Intelligence at MIT's Sloan School of Management (Boston, MA). His research field concerns the issues of innovation and technology-driven entrepreneurship, with a specific focus on the phenomena of virtual communities of users and digital ecosystems. Currently, he is involved in a project related to the development of a model for the smart specialization of a regional destination as well as in educational initiatives in the field of entrepreneurial development. His research activities have been documented in several publications in international journals, conference proceedings, and book chapters.

Alberto Di Minin is Associate Professor of Management at the Scuola Superiore Sant'Anna in Pisa. He is also Research Fellow with the Berkeley Roundtable on the International Economy (BRIE), University of California – Berkeley and Social Innovation Fellow with the Meridian International Center in Washington, DC. He is currently the Italian representative on the SMEs & Access to Finance Programme Committee, for Horizon 2020, with the European Commission. He is the Co-Director of the Executive Doctorate in Business Administration Program at the Sant'Anna, the Director of the Confucius Institute of Pisa, and the Director of the Galilei Institute in Chongqing University. His research deals with Open Innovation, appropriation of innovation and science and technology policy. He also works on technology transfer, intellectual property, and R&D management.

Antonio Messeni Petruzzelli is Senior Lecturer in Innovation Management and co-founder of the Innovation-Management Group at the Politecnico di Bari. Dr. Messeni Petruzzelli holds a PhD in Management (2008) and a Laurea in Management Engineering (2003) from Politecnico di Bari. During the 2007–2008 academic year, he was a visiting researcher at the IESE Business School (Department of Strategic Management) of the University of Navarra (Barcelona, Spain). Dr. Messeni Petruzzelli is the author of 38 international publications and three international books on the topic of knowledge search and recombination. Specifically, his research lies at the intersection between innovation and knowledge management. His studies on this issue have been published in leading journals such as *Academy of Management Perspectives*, *Journal of Management*, *International Journal of Management Reviews*, *Long Range Planning*, *Technological Forecasting and Social Change*, *Journal of Organizational Behavior*, *Industry and Innovation*, *Technology Analysis & Strategic Management*, *Technovation*, and *European Management Review*. His studies have recently been awarded the Nokia Siemens Network Award in Technology Management for Innovation into the Future.

Umberto Panniello received his PhD in Business Engineering from the Politecnico di Bari, Italy. He has been a Visiting Scholar at the Wharton Business School of the University of Pennsylvania, Philadelphia, USA. He is an Assistant Professor in Management at the Politecnico di Bari, where he teaches e-business models, business intelligence, and marketing in the School of Engineering. His current main research interests are customer modeling, consumer behavior, customer re-identification, context-aware and profit-based recommender systems, and Social TV. He has also carried out research on knowledge discovery in databases to support the decision-making process.

Salvatore Pirri is a PhD student in Management of Innovation at Sant'Anna School of Advanced Studies in Pisa, Italy. He graduated

cum laude (2013) in Economics and Sciences of Public Government at the University of Calabria, and obtained a Postgraduate Masters in Management of Innovation at Sant'Anna School of Advanced Studies in 2015. Salvatore also gained working experience as a business analyst and junior project manager in Italy and elsewhere in Europe: at the Italian Embassy in Budapest (2012); at ENEA in Brussels (2014); at SELEX-LEONARDO in Genoa (2015). His current research interests concern an innovative approach to Big Data analysis applied in the domain of healthcare, to fostering new business models able to connect IT instruments for customers and firms. His research is mainly focused on managerial and public policy aspects of technological applications and innovation processes for pharmacovigilance and predictive fraudulent behavior.