R&D MANAGEMENT

What "V" of the big data influence SMEs' open innovation breadth and depth? An empirical analysis

Giulio Ferrigno*, **Saverio Barabuffi**, **Enrico Marcazzan** and **Andrea Piccaluga**

Institute of Management and L'EMbeDS Department, Scuola Superiore Sant'Anna, Piazza Martiri della Libertà, 33, Pisa, Italy. giulio.ferrigno@santannapisa.it, saverio.barabuffi@santannapisa.it, enrico.marcazzan@santannapisa.it, andrea.piccaluga@santannapisa.it

The open innovation (OI) paradigm has garnered relevant attention in recent years. Against this backdrop, this study explores the impact of a relatively recent phenomenon, such as Big Data, in terms of Volume, Velocity, and Variety, on small and medium enterprises' (SMEs') OI search. In fact, while issues related to Big Data have been often examined in the context of high-tech firms, the effects on SMEs' OI search strategies have not been extensively studied. This paper addresses this gap by developing a quantitative analysis on a sample of 123 Italian SMEs. The findings reveal that Big Data significantly influences SMEs' OI breadth, leading to increased external collaborations. In parallel, they do not affect SMEs' OI depth. Moreover, the impact varies among the different "3Vs" of Big Data, suggesting that some characteristics have a more pronounced effect on SMEs' OI strategies. Drawing on these insights, this study contributes to the understanding of the interplay between Big Data characteristics and SMEs' OI, offering hopefully valuable implications for both OI and Big Data literature and proposing avenues for further research and practice.

1. Introduction

Open Innovation (OI) has attracted the attention of both scholars and practitioners in the last two decades (Gassmann et al., 2010; Randhawa et al., 2016; Enkel et al., 2020; Marullo et al., 2022). In particular, since its introduction by Chesbrough (2003), extensive research has been conducted about OI breadth (i.e., the number of sources used to innovate) and OI depth (i.e., the extent to which the sources are involved in the innovation process) (Laursen and Salter, 2006; Messeni Petruzzelli et al., 2022; Nieto et al., 2023). Whereas OI offers both financial and non-financial benefits (Dahlander and Gann, 2010) regardless of firm's size (Spithoven et al., 2013), most studies have investigated OI in large companies, thereby producing an imbalance in our understanding of OI (Spithoven et al., 2013). As a matter of fact, despite some recent attention to small and medium enterprises (SMEs) (Kraus et al., 2020; Barrett et al., 2021; Carrasco-Carvajal et al., 2023), searching for new partners (i.e., OI breadth) and developing intense relationships with them (i.e., OI depth) can be more challenging for this type of companies, given their size, resource constraints, and managerial objectives (Spithoven et al., 2013; Dooley and O'Sullivan, 2018). According to the European

Commission, SMEs are companies that employ <250 people and have a turnover of <€50 million or a balance sheet total of <€43 million. They represent an important industrial component in many countries. For example, in Italy, SMEs account for over 90% of the workforce (ISTAT, 2022).

Nowadays, the world is overwhelmed by Big Data generated every second, with the growth rate increasing approximately 10 times every five years (Del Vecchio et al., 2018). Over 2025, data are expected to grow to more than 180 zettabytes (Statista, 2023). 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 analyzed (Kaisler et al., 2013; Ward and Barker, 2013). The ability to aggregate, elaborate, and analyze Big Data is becoming a key competitive advantage and resource for firms of different sizes, including SMEs (Del Vecchio et al., 2018). As such, scholars have explored the impact of Big Data characteristics mainly in terms of (1) *Big Data Volume* – that is, the amount of data (Wamba et al., 2015), (2) *Big Data Velocity* – that is, the speed of generating and analyzing data (Ghasemaghaei et al., 2017), and (3) *Big Data Variety* – that is, the diversity of data types, including both structured (e.g., numbers) and unstructured (e.g., pictures) data (Ghasemaghaei et al., 2017; Ghasemaghaei and Calic, 2020) on firms' performance (Ghasemaghaei, 2019; Cappa et al., 2021).

The impact of Big Data on SMEs' OI performance has been analyzed through qualitative studies. For instance, Urbinati et al. (2020) have investigated how digital technologies, including Big Data, can foster SMEs' OI firms' performance. However, to the best of our knowledge, research has neglected to develop quantitative studies to examine the relationship between Big Data characteristics and SMEs' OI search strategies, which is the object of this paper. The theoretical importance of this relationship is confirmed in the R&D Management work by Enkel et al. (2020, p. 165), who claim that a research area that needs further investigation is to understand "which influence has big or linked data on OI." This relationship may exhibit both positive and negative aspects. On one hand, there is a potential positive connection since Big Data can offer insights into effectively framing business challenges and identifying relevant external knowledge and resources for internal integration (Cepa, 2021). Conversely, there is also the possibility of a negative association, wherein greater values of Big Data characteristics of within firms could lead to a reduced inclination toward adopting OI search strategies because analyzing large amounts of data can be confusing and yield only a few useful insights for OI collaboration (Ghasemaghaei and Calic, 2019). This is why in this paper we focus on the impact of Big Data characteristics on SMEs' OI search strategies through the following research question: *what forms of Big Data are relevant for SMEs' open innovation search and breadth?*

To answer our research question, we focused our attention on developing some hypotheses about the effects of Big Data Volume, Big Data Velocity, and Big Data Variety (Johnson et al., 2017; Ghasemaghaei and Calic, 2019; Pedota, 2023) on SMEs' OI breadth and depth (Laursen and Salter, 2006).[1]

To verify our hypotheses, we conducted an econometric analysis (i.e., a negative binomial regression) on a sample of 123 Italian SMEs. We decided to focus on this specific country for two reasons. First, in the last decade, the Italian Government has launched many plans and ad-hoc interventions to stimulate firms' digitalization and the adoption of Industry 4.0 technologies, including Big Data (Messeni Petruzzelli et al., 2022). Second, considering the peculiar structure of the Italian economy, strongly based on SMEs and the recent national policies aiming at accelerating digitalization, Italy represents an interesting country to catch the complexity of SMEs behaviors toward the adoption of Big Data (Martinelli et al., 2021).

We found that each Big Data characteristic does impact SME's OI breadth. More specifically, Big Data Velocity and Variety both have a positive impact on SME's OI breadth, while Big Data Volume has a negative impact. However, none of them affects SMEs' OI depth. As a net result, our findings suggest that Big Data leads SMEs to activate external collaborations but not to intensify them. Drawing on these findings, our study aims to provide some theoretical contributions to OI, Big Data, and SMEs research, which are reported in the concluding section of the paper.

The remainder of the paper is organized as follows. In Section 2, we explore the links between Big Data characteristics and SMEs' OI breadth and depth and we develop our six research hypotheses. Then, in Section 3, we illustrate the methodology, the survey's data collected, and the empirical approach. Finally, we present the results (Section 4) and we discuss them, together with the main conclusions and implications, in Section 5.

2. Theoretical background and research hypotheses

2.1. Big Data and SMEs

The world is flooded with Big Data, and their growth rate is about 10 times every 5 years (Del Vecchio et al., 2018). Big Data refers to datasets so large and complex they create significant challenges for traditional data management and analysis tools in practical timeframes (Deloitte Consulting, 2012; Ward and Barker, 2013). Introduced by Laney (2001), Big Data Volume, Velocity, and Variety are the three main characteristics of Big Data, also known as the "3Vs" of Big Data (Tan and Zhan, 2017; Ghasemaghaei and Calic, 2020). First, *Big Data Volume* relates to large sizes of data that are collected and analyzed by companies (Ghasemaghaei and Calic, 2020). Second, *Big Data Velocity* refers to the speed of generating and analyzing data (Ghasemaghaei and Calic, 2020): the rate of data generation has been enhanced thanks to the increasing power and number of computers and digital devices, generating new possibilities for data analysis (Ghasemaghaei and Calic, 2020). Last, *Big Data Variety* indicates the diversity of data types, including both structured (e.g., numbers) and unstructured (e.g., pictures) data (Ghasemaghaei et al., 2017; Ghasemaghaei and Calic, 2020). Taken together, these three characteristics also represent guidelines that describe how an organization can develop new knowledge, improve decision-making processes, and better address consumers' needs.[2]

The ability to aggregate, process, and analyze big data is becoming a key competitive advantage and critical resource, especially for SMEs, and can be analyzed according to the Resource Based View (RBV) (Wernerfelt, 1984; Barney, 1991; Del Vecchio et al., 2018; Cappa et al., 2021). In fact, big data represent a valuable asset that can be difficult for competitors to replicate or substitute (Sena et al., 2019). As a result, SMEs across various industries are increasingly embracing digitalization and leveraging big data to enhance decision-making processes and innovate their products and services (McAfee et al., 2012; McKinsey Global Institute, 2016; Garg et al., 2017; Batistič and van der Laken, 2019). Some interesting examples of SMEs that are interested in developing Big Data technologies can be found in the Big Data Value Association (BDVA's) (2017) report and in the BDVA webpage (https://bdva.eu/members/).

2.2. OI search strategies in SMEs

Since the inaugural work of Chesbrough (2003), OI has gained increasing attention by both scholars and practitioners (Bogers et al., 2017; Dahlander et al., 2021), especially in the innovation/R&D management field (Ferrigno et al., 2023). OI "supports the firm's capabilities to insource external ideas and resources, to co-develop new products and processes with external partners, and to market internal ideas that fall outside the firm's current business model" (Enkel et al., 2009; Messeni Petruzzelli et al., 2022, p. 617) and impacts both firms' financial and non-financial performance (Carrasco-Carvajal et al., 2023; Martín-Peña et al., 2023).

More specifically, there is a strong interest in firms' search strategies (Greco et al., 2015; Sá et al., 2023) and a well-consolidated literature supports the idea that firms can adopt two different OI search strategies: (1) *OI breadth*, that is, the number of sources used to innovate; and (2) *OI depth*, that is, the extent to which the sources are involved in the innovation process (Laursen and Salter, 2006; Messeni Petruzzelli et al., 2022). Both OI search strategies influence firms' propension toward the acquisition of external knowledge (Laursen and Salter, 2006; Garriga et al., 2013) from several sources, including suppliers, customers, competitors, financial companies, consulting companies, other private companies, universities and research centers, and other public organizations (Messeni Petruzzelli et al., 2022).

Relying on external sources of knowledge is particularly important for SMEs (Messeni Petruzzelli et al., 2022; Del Sarto et al., 2023), which have several constraints in terms of resources, technological assets, skills, and expertise that can be devoted to the development of innovative products, services, or processes (Dahlander and Gann, 2010; Spithoven et al., 2013; Dahlander et al., 2021). Opening up to the external environment could help firms to solve or mitigate these issues (Carrasco-Carvajal et al., 2023).

Alongside theoretical reasonings, the implementation of OI search strategies in SMEs undoubtedly offers practical benefits. The breadth strategy suggests collaborating with universities and innovation hubs to address knowledge gaps and resource limitations (Kiel et al., 2017), fostering managerial readiness for new technologies (Zangiacomi et al., 2020). In some cases, horizontal partnerships with competitors can enhance operational and financial awareness (Shin et al., 2014). Conversely, the depth strategy emphasizes close ties with technology providers for knowledge assimilation (Messeni Petruzzelli et al., 2022), fostering strong relationships to facilitate learning processes (Terjesen and Patel, 2017) and leveraging customer and supplier relationships for product/service development (Nguyen et al., 2015). In sum, intensive cooperation with external partners enables SMEs to acquire the necessary knowledge over time, enhancing their capabilities.

However, this particular *genus* of firm (i.e., SME) may exhibit lower levels of external search breadth and depth (Laursen and Salter, 2006). Extant research has shown that SMEs do not have sufficient resources, skills, and competences to search for and profit from a wide range of external partners (Aslesen and Freel, 2012); it can be therefore difficult for them to expand their partners' network, especially in low-tech industries. Moreover, SMEs have less resources to cooperate with a large variety of partners and develop intense relationships with them (Messeni Petruzzelli et al., 2022).

Therefore, using the RBV, in this study, we posit that SMEs can leverage data (Akhtar et al., 2019) to address this deficiency and effectively implement OI search strategies and related practices. In the next section, we discuss in detail what has been written about the effects of the characteristics of Big Data on OI search in the context of SMEs.

2.3. *The "3Vs" of Big Data and SMEs' OI breadth and depth*

Despite the notable emphasis on Big Data and the promising contributions they can bring to SMEs' OI search strategies (Del Vecchio et al., 2018), extant studies have not investigated the impact of Big Data on SMEs' OI breadth and depth, which is particularly relevant for R&D management. For instance, the R&D management work by Enkel et al. (2020) suggests that a research area that needs further investigation is the one devoted to examining "which influence has big or linked data on OI" (Enkel et al., 2020, p. 165). Similarly, Urbinati et al. (2020) argue that we have only limited knowledge about the connection between Big Data and OI.

Drawing on these intuitions, in this article, we posit that the main characteristics of Big Data may influence SMEs' OI breadth and depth. The relationship between Big Data characteristics and OI search strategies could be both positive, since Big Data can help to provide information on how to appropriately frame business issues and opportunities and the related external knowledge and resources to internalize (Cepa, 2021), and negative, since the higher the values of the firms' Big Data characteristics, the lower the propensity toward the adoption of OI search strategies (Keupp and Gassmann, 2009). In this sub-section, we hypothesize that each "V" of Big Data impacts SMEs' OI breadth and depth. A snapshot of the research hypotheses is reported in Figure 1.

Existing studies suggest that, without a sufficient amount of data and an appropriate technology to analyze them, it is quite difficult for SMEs to create and/or implement innovative solutions (Sivarajah et al., 2017; Del Vecchio et al., 2018; Ghasemaghaei, 2019). As a matter of fact, Big Data can stimulate the search for best practices and more efficient solutions outside the firm, relying on the experience of external actors (Messeni Petruzzelli et al., 2022). However, the literature is not unanimous toward the benefits derived by the data Volume (Ghasemaghaei and Calic, 2020; Cappa et al., 2021; Ghasemaghaei, 2021), suggesting the existence of a phenomenon known as "InfoObesity" (Whitler, 2018; Cappa et al., 2021), which implies that very large amounts of data could have negative effects on SMEs' OI breadth. First, SMEs might lack both financial and human resources devoted to collect and store many data (Giotopoulos et al., 2017; Agostini and Filippini, 2019; Horváth and Szabó, 2019; Eller et al., 2020). In this regard, some scholars found that relevant costs should be sustained not only to collect and store data but also to appropriately analyze them (Cappa et al., 2021). Moreover, SMEs' employees could have important cognitive limitations to process large amounts of data (Prescott, 2016). Therefore, paradoxically, in a context characterized by lack of resources (i.e., SMEs), Big Data Volume might negatively impact the likelihood of adopting interorganizational relationships and, in turn, the search breadth of the OI strategy.

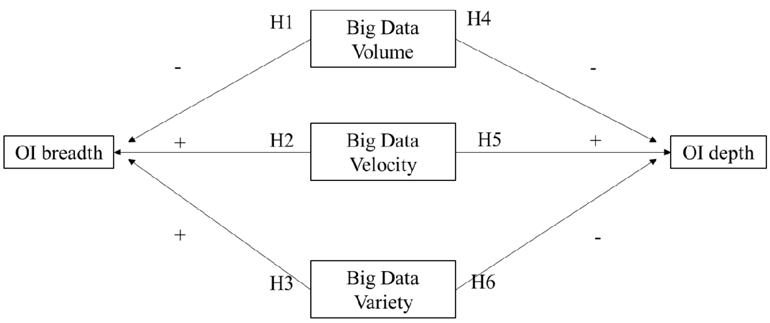

Figure 1. Theoretical model.

Second, analyzing large amounts of data can be confusing and yield only a few useful insights (Ghasemaghaei and Calic, 2019). This is especially true in the case of SMEs because they do not have reliable information on how and where to find the appropriate knowledge (Del Vecchio et al., 2018).

As such, Big Data Volume could represent an internal source of information that may obstacle fruitful forms of collaboration. In this sense, there could be enough information within the firm that can be used to develop internal innovative solutions without relying on external knowledge (Keupp and Gassmann, 2009). Therefore, in line with the above arguments, we hypothesize the following relationship:

H1 Big Data Volume exhibits a negative relationship with SMEs' OI breadth.

Several studies show that processing data in real time can allow firms to access external knowledge (Erevelles et al., 2016; Cepa, 2021). First, a fast process of data analysis can promptly reveal SMEs' needs and the collaboration opportunities that can be activated to satisfy them (Johnson et al., 2017). In this regard, some scholars suggest that Big Data Velocity is very helpful in continuously solving business issues (Ghasemaghaei and Calic, 2019). Moreover, Cepa (2021) states that architectural designs enabling a fast process of data analysis are crucial to collaboration dynamics. In this sense, the speed of data collected and analyzed could lead to an increase of OI breadth because fast process of data analysis can instantly unveil what an SME lacks and how to fill this gap (Johnson et al., 2017). Practically speaking, up-to-date information about the needs of the firm can affect the search for best practices and solutions according to technological, innovation-hubs, and research centers, which in turn improve the SME's attitude toward the adoption of new technologies (Zangiacomi et al., 2020).

Second, in a context characterized by a limitation of resources, like in SMEs, the related fast decision-making process could accelerate the search for adequate inter-organizational relationships to acquire the appropriate resources and knowledge (Keupp and Gassmann, 2009). This is particularly true in dynamic sectors with high competition, where quickly understanding consumer preferences might foster a competitive advantage (Dagnino et al., 2021). As previously mentioned, in order to access to a rich set of diverse and complementary resources, SMEs can collaborate with a vast range of external partners (thus implementing an OI breath strategy) (Gopalakrishnan and Damanpour, 1994; Messeni Petruzzelli et al., 2022). In this regard, Erevelles et al. (2016) find that having access to current and updated data can help SMEs to improve their entire decision-making process (Erevelles et al., 2016). This process might stimulate the search for the appropriate external partner. Therefore, we hypothesized that:

H2 Big Data Velocity exhibits a positive relationship with SMEs' OI breadth.

Existing studies also suggest that the search for appropriate interorganizational relationships and the resulting OI breadth strategy can emerge from the different data types a firm has (Keupp and Gassmann, 2009). Since "with more heterogeneous knowledge, managers have a larger number of hooks to recognize the value of further knowledge" (Pedota, 2023, p. 4), valuable knowledge can be found outside the boundaries of the company (Carrasco-Carvajal et al., 2023) and, in the case of SMEs, collaboration with other entities can help them overcome the lack of knowledge for innovative activities (Narula, 2004; Chesbrough, 2011; Dahlander et al., 2021; Messeni Petruzzelli et al., 2022). Furthermore, a varied amount of information means that SMEs have non-redundant information (Burt, 1992), which allows them to adequately frame their problems, stimulating the search for the right ideas and solutions (Urbinati et al., 2020). In fact, heterogeneous information allows SMEs to improve their focus of attention and, consequently, the perception of the external environment since they have a rich picture of the environment and, consequently, of the opportunities to exploit (Pedota, 2023). Practically speaking, this need can stimulate, for instance, the co-organization, with universities and research centers, of educational and training events on a wide variety of topics, like mathematics, engineering, programing, and data analysis and processing, which are useful in the context of big data and, broadly speaking, Industry 4.0 for exchanging viewpoints, ideas, and best practices among the employees and other participants with a diverse knowledge base.

Based on the arguments above, we hypothesize that SMEs relying on diverse data sources can better build and capitalize on a broad range of collaboration opportunities, thus implementing a breadth of OI strategies. These opportunities emerge from the degree of heterogeneity related to the knowledge that an SME possesses. A heterogeneous degree of knowledge leads to stimulating the search for the right ideas and creative solutions within the company, thus influencing collaboration opportunities (Pedota, 2023). Therefore, we hypothesize that:

H3 Big Data Variety exhibits a positive relationship with SMEs' OI breadth.

Moving to OI depth, extant studies point out that the "3Vs" of Big Data seem to have less powerful insights in the firms' exploitation orientation than the effects they may have on exploratory activities (Johnson et al., 2017). More specifically, prior works suggest that both (technology) exploration and exploration are related to OI practices (Carrasco-Carvajal et al., 2023). However, one crucial factor that seems to distinguish exploration and exploitation is related to the activities based on external network that affect the exploitation side (Carrasco-Carvajal et al., 2023). In this sense, exploitation orientation "prompts firms to focus on the refinement of existing routines" (Lisboa et al., 2011; Johnson et al., 2017, p. 645). As previously anticipated, the effect of Big Data "Volume" could have both positive and negative effects on OI search strategies. Considering OI depth, on one side, the characteristics of SMEs can lead to an increase in the propensity to activate external collaboration as the "Volume" of data increases. As the volume of managed data increases, SMEs may find it difficult to support all the activities required to analyze and derive value from the collected data on their own. In response, they may be inclined to seek collaborations with external partners who have additional resources, such as advanced analytical skills, scalable IT infrastructure, and data management capabilities (e.g., companies specializing in data analytics or expert consultants), as previously mentioned above. On the other side, with an increase in the "Volume" of data, the complexity of information can increase dramatically. This could make it more difficult for SMEs to manage and analyze data effectively. If data are not properly organized and structured, it could become complicated for SMEs to share and exchange information with external partners. Data complexity (Luqman et al., 2024) could then limit the depth of external collaborations, as it may require more time and resources to properly integrate and interpret the shared data. Moreover, as the volume of data increases, so does the risk of data security and privacy breaches (Iqbal et al., 2018). SMEs need to ensure that data shared with external partners are adequately protected to prevent unauthorized access or accidental disclosure of sensitive information. This security concern could lead SMEs to be more reluctant to share sensitive data with external partners, thus limiting the depth of collaborations. Finally, with an increase in the volume of big data, SMEs may face challenges related to data standardization and interoperability (Han and Trimi, 2022). If data are collected, stored, and managed in different formats or using incompatible systems and platforms, it could be difficult for SMEs to integrate and share data effectively with external partners. This lack of standardization and interoperability could limit the depth of external collaborations as it may be difficult for the parties involved to work together synergistically and consistently.

In sum, even if a bigger volume of data could enable firms to better perceive the external environment, at the same time, the marginal effect of an added piece of data decreases its value proportionally (Laney, 2001). This can lead to non-useful information for the firm, decreasing the return in exploitation-related activities. This effect, in turn, can diminish the propensity to intensify (i.e., going in-depth) the activities based on an external network (Carrasco-Carvajal et al., 2023), leading to a negative effect on OI depth. Consequently, we surmise that:

H4 Big Data Volume exhibits a negative relationship with SMEs' OI depth.

The speed of integrating and analyzing data might affect the routine-based and repetitive approach to organizational changes (Rust et al., 2002). In this sense, "analyzing data in real time helps firms to quickly generate insights about what is happening now, what is likely to happen in the future, and what actions they need to take to get the optimal results" (Ghasemaghaei and Calic, 2019, p. 72).

Faster data management can facilitate communication and information sharing with external partners (Yildiz et al., 2024). For example, by using tools and technologies that enable instant data sharing, SMEs can collaborate more effectively with their partners, exchanging information in real time and facilitating shared decision making. This improved communication and data sharing can increase the depth of external collaborations, enabling greater transparency and synergy among the parties involved (Messeni Petruzzelli et al., 2022). Moreover, high-speed data processing enables the SME to quickly adapt to changing market needs and new opportunities (Ferrigno et al., 2023), even for collaboration with external partners. By constantly monitoring data and market trends, the SME can quickly identify opportunities for collaboration and act promptly to take advantage of them. This agility in adapting can foster the development of deeper and longer-lasting partnerships with external partners, enabling the SME to remain competitive and innovative in its industry.

In sum, this speed might improve the perception of the external environment, affecting the orientation toward exploitation activities (like collaborations) (Johnson et al., 2017). Building on this set of arguments, we hypothesize that:

H5 Big Data Velocity exhibits a positive relationship with SMEs' OI depth.

Finally, "exploitation demands efficiency and convergent thinking to harness current and familiar capabilities and continuously improve product offerings" (Wadhwa and Kotha, 2006; Johnson et al., 2017, p. 645). This leads a firm to avoid using Big Data Variety, since the related analytical issues do not meet the requirements of the decision makers in emphasizing both efficiency and depletion of development costs and time (Morgan and Berthon, 2008). Therefore, Big Data Variety can negatively affect OI depth.

If a SME has access to data from different sources, such as market data, financial data, and operational data, it can provide a completer and more in-depth overview of the challenges and opportunities for collaboration with external partners (Del Vecchio et al., 2018). As previously mentioned, through the analysis of data from a variety of sources, SMEs can identify emerging trends, hidden patterns, or unmet market needs, which can serve as insights for new collaborative initiatives and business strategies. By using detailed and contextualized information, the SME can tailor its approaches and offerings to best meet the needs and expectations of external partners (Wibowo et al., 2021). This ability to customize can help create deeper and longer-lasting relationships with partners as it demonstrates a commitment to mutual success and satisfaction. In line with the above arguments, we hypothesize the following relationship:

H6 Big Data Variety exhibits a negative relationship with SMEs' OI depth.

3. Data and methodology

To empirically investigate our hypotheses, we collected evidence on Big Data characteristics in Italian SMEs, defined according to the European Commission as companies that employ <250 persons and have a turnover lower than 50 million of euros or a balance sheet total lower than 43 million of euros. Italy is notable for its significant weight of SMEs, which account for over 90% of the national workforce (ISTAT, 2022). According to a study by its Ministry of the Economic Development (MISE, 2020), Italy exhibits a substantial percentage of firms transitioning toward Industry 4.0. In the last decade, the Italian Government has launched many plans and ad-hoc interventions to stimulate firms' digitalization and the adoption of Industry 4.0 technologies, including Big Data. Consequently, considering the peculiar structure of the Italian economy, mainly based on SMEs, and the recent national policies aiming at accelerating digitalization, Italy represents an intriguing backdrop for scrutinizing the multifaceted behaviors of SMEs pertaining to the adoption of Big Data (Martinelli et al., 2021).

After gaining an initial comprehension about the importance of Big Data in the Italian landscape, we developed and launched a survey drawing insights from extant literature (see Appendix A). Moreover, we conducted three preliminary tests aiming at enhancing the clarity of the questionnaire and fortifying the reliability of the data acquired (Crick et al., 2023). Specifically, in March 2023, we tested the survey involving both five academic experts in the fields of OI, Big Data, and SMEs and five decision makers and founders of SMEs operating in different sectors and obtained insights for refining the survey (Laursen and Salter, 2006). Additionally, in April 2023, an independent agency, not involved in the survey's administration, reviewed the questionnaire.

In May 2023, to build our sample, we accessed the Orbis database by Bureau van Dijk, which is known to be a reliable resource and has universal acceptability for investigating Big Data, OI, or SMEs (Costa et al., 2023; Rammer and Es-Sadki, 2023). Our search was tailored to identify Italian SMEs, defined as companies with <250 employees. This search yielded 19,722 SMEs. Subsequently, we extracted primary information about these SMEs, including company names, BvD identification number, telephone number, websites, and email addresses, and we organized them within an Excel spreadsheet.

In June 2023, we started the data collection. In particular, we targeted a randomly selected subset of Italian SME founders and decision makers across various industry sectors. The data collection process was carried out via the Computer-Assisted Telephone Interview (CATI) method, with the support of an external professional firm. Before launching the questionnaire, we embarked on an additional pre-test with 27 SMEs, whose feedback guided further refinements to the questionnaire. By August 2023, the companies contacted were 801, 591 of which did not engage in big data activities (discarded) and 210 engaged in big data activities. Of these companies engaged in big data activities, 55 did not complete the questionnaire, while 155 completed the questionnaire. Excluding the 27 responses gathered as pre-test, we successfully collected complete data from 128 firms. Following a meticulous review, we excluded five responses, as respondents indicated that their respective companies did not meet the SMES' classification criteria. Consequently, our final sample encompassed 123 SMEs.[3] Figure 2 illustrates the percentage frequency distribution of the 123 sampled firms across various regions in Italy.

Additionally, Figure 3 shows the percentage frequency distribution of respondents' age categories as reflected in the survey's data.

Finally, Figure 4 reports the percentage frequency distribution of the educational level of the respondents to the survey.

The variables implemented in our empirical analysis are sourced from the completed questionnaires.

3.1. Dependent variable

We followed previous works to measure OI breadth and OI depth (Laursen and Salter, 2006; Messeni Petruzzelli et al., 2022; Carrasco-Carvajal et al., 2023). First, we operationalize OI breadth as the number of sources (i.e., suppliers, customers, competitors, financial companies, consulting companies, other private companies, universities and research centers, other public organizations) that collaborate with the firm (in our case Italian SME) to stimulate innovative activities. Second, we operationalized OI depth as the number of sources with which the firm collaborates "very often" (Messeni Petruzzelli et al., 2022). According to Carrasco-Carvajal et al. (2023), these variables capture the role of external knowledge sources (Laursen and Salter, 2006; Ahn et al., 2017; Messeni Petruzzelli et al., 2022). The variables are computed summing, on one side, the number of sources and, on the other side, the number of sources with which a firm collaborates "very often." Therefore, their value lies between zero (absence of collaboration, min) and eight (max).

3.2. Independent variables

We followed previous studies on Big Data characteristics (Ghasemaghaei, 2019, 2021; Ghasemaghaei and Calic, 2020) to operationalize our explanatory variables, namely (1) *Big Data Volume*, (2) *Big Data Velocity*, and (3) *Big Data Variety*. In particular, in line with Ghasemaghaei and Calic (2019, 2020), who followed Johnson et al. (2017), we measured *Big Data Volume*, *Big Data Velocity*, and *Big Data Variety* according to a seven-point Likert scale (ranging from 1, "strongly disagree," to 7, "strongly agree"). The Likert scale enables us to indicate the amount of data collected (*Big Data Volume*; Wamba et al., 2015; Cappa et al., 2021), the speed and the frequency of processing and integrating data (*Big Data Velocity*; Ghasemaghaei and Calic, 2020), and the assortment of data per observation (*Big Data Variety*; Wamba et al., 2015; Cappa et al., 2021). By sourcing from the answers to the questionnaire, both *Big Data Volume* and *Big Data Velocity* consist of four items, whereas *Big Data Variety* consists of three items. Then, to operationalize them, we took the average score of the several responses (further details are reported in Appendix A).

3.3. Control variables

Consistent with prior research investigating the features of SMEs (Scuotto et al., 2017; Messeni Petruzzelli et al., 2022), our model incorporated several control variables sourced from the answers to the questionnaire. First, a crucial factor in explaining the adoption of Big Data within

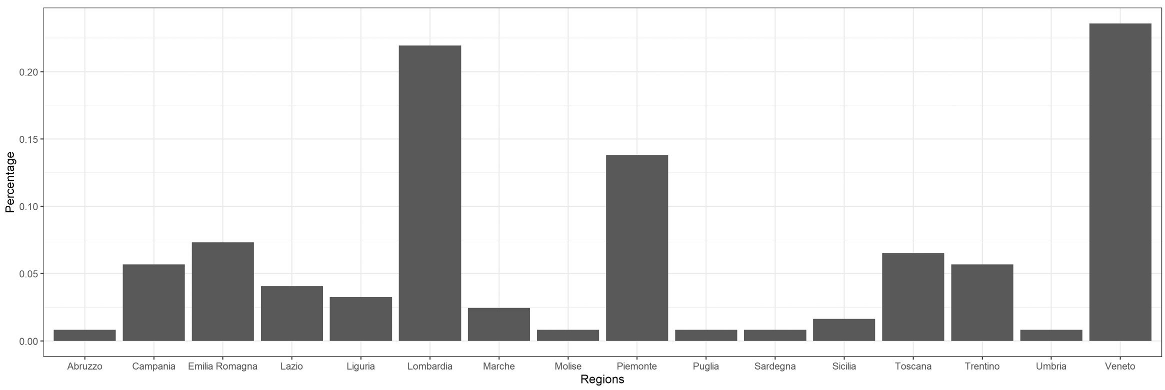

Figure 2. Percentage frequency of firms' regional distribution.

SMEs is their absorptive capacity, that is, the firm's ability to capture and effectively leverage external knowledge (Zahra and George, 2002). This absorptive capacity is quantified through the ratio of R&D expenditures to total revenues (Zahra and George, 2002; Cassetta et al., 2020). More concretely, we operationalized it as zero (*lower than 5%*), one (*between 5% and 10%*), two (*between 11% and 20%*), three (*between 21% and 30%*), or four (*higher than 30%*).

Second, we included SME's age at the time of the survey because of its influence on the adoption of Big Data (Kelly and Amburgey, 1991; Messeni Petruzzelli et al., 2022). On one hand, older SMEs may struggle with organizational inertia issues (Kelly and Amburgey, 1991). Conversely, younger SMEs can have a potential lack of human and financial resources (Messeni Petruzzelli et al., 2022).

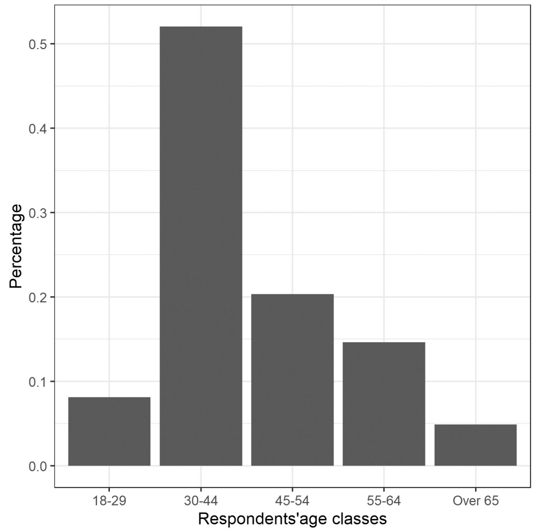

Figure 3. Percentage frequency distribution of respondents' age classes.

Third, we incorporated SMEs' size by introducing the number of employees referring to year 2023. In fact, previous studies found that the relative size of the company is very relevant among SMEs (Arbore and Ordanini, 2006; Horváth and Szabó, 2019).

Fourth, we considered the number of patents related to Big Data (Martinelli et al., 2021), a variable related to the data skills among employees (Damij et al., 2022).

Fifth, we assessed whether the SME operates in a high-tech industry or not, since recent studies have shown that the technological intensity of the business sector may exert influence on the likelihood of adopting Industry 4.0-related technologies (Messeni Petruzzelli et al., 2022). Lastly, by following Cappa et al. (2021), we controlled for the Industry R&D intensity by implementing the taxonomy according to firms' European Classification of Economic Activities (NACE) developed by Galindo-Rueda and Verger (2016) and by operationalizing it as a categorical variable: high, medium–high, medium, medium–low, and low R&D intensity.

Further details pertaining to the operationalization of these variables can be found in Appendix A. Table 1, instead, reports the abbreviations of all variables included in the empirical models and the descriptive statistics of discrete and continuous variables.

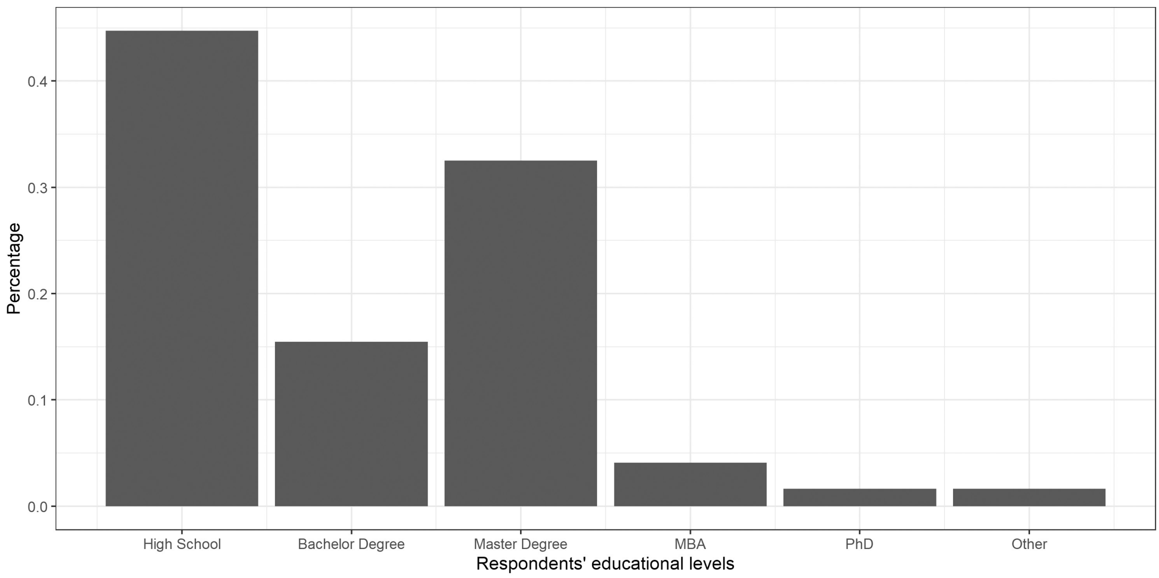

Figure 4. Percentage frequency distribution of respondents' educational level.

Table 1. Abbreviations and descriptive statistics

Description	Abbreviations	Observations	Mean	SD	Min	Max
Open innovation breadth	OI breadth	123	3.11	1.87	1	8
Open innovation depth	OI depth	123	2.05	1.34	1	7
Average of Big Data Volume	Big Data Volume	123	4.96	1.46	1	7
Average of Big Data Velocity	Big Data Velocity	123	4.51	1.33	1	7
Average of Big Data Variety	Big Data Variety	123	4.7	1.46	1.33	7
R&D expenditures on revenues <5%	R&D exp <5%					
R&D expenditures on revenues between 5% and 10%	R&D exp 5%–10%					
R&D expenditures on revenues between 11% and 20%	R&D exp 11%–20%					
R&D expenditures on revenues between 21% and 30%	R&D exp 21%–30%					
R&D expenditures on revenues more than 30%	R&D exp >30%					
Firm's age	Age	123	40.93	21.91	3	120
Firm's size	Size	123	101.81	41.94	14	200
Dummy variable for patents developed by using Big Data	Patents Big Data					
Dummy variable for technological intensity industry	Tech int ind					
High R&D intensity industry	High R&D int ind					
Medium-high R&D intensity industry	Med-high R&D int ind					
Medium R&D intensity industry	Med R&D int ind					
Medium-low R&D intensity industry	Med-low R&D int ind					
Low R&D intensity industry	Low R&D int ind					

In Figures 5–8, we show the frequencies of categorical and dummy variables: R&D expenditures, Patents Big Data, Technology intensity industry, and Industry R&D intensity, respectively.

Table 2 illustrates the variance inflation factor (VIF) for each independent and control variable and the correlation matrix among the variables included in the empirical estimation. The highest correlation is between the three independent variables *Big Data Volume*, *Big Data Velocity*, and *Big Data Variety*. To address possible issues related to multicollinearity, we checked the value of VIF that resulted below a cut-off value of 10 for all the variables included in our regressions (Kutner et al., 2005).

3.4. Estimation method

In the empirical estimation, the dependent variables are discrete and non-negative count data. Accordingly, we propose a negative binomial regression approach by following previous literature analyzing the relationship between Industry 4.0 technologies and OI (Messeni Petruzzelli et al., 2022). The negative binomial model is a generalization of the Poisson model in which the Poisson parameter is represented as $\lambda(X) = Ve^{X'\beta}$, where V is a random variable with a Gamma distribution, and it allows to address overdispersion (Cameron and Trivedi, 2013; Hilbe, 2014). We computed our regression models by using bootstrapped cluster-robust standard errors at the regional level to correct for heteroskedasticity and to account within-group dependence in the presence of a small number of clusters (Cameron et al., 2008). The empirical models that we estimate, in general, are specified as follows:

$$Y = f(\mathit{Big\ Data\ Volume}, \mathit{Big\ Data\ Variety}, \mathit{Big\ Data\ Velocity}, \mathit{Control\ variables})$$

where Y represents *OI Breadth* and *OI Depth* in the empirical models we estimate.

4. Results

Table 3 reports the results of the negative binomial regression models on the variable *OI Breadth*. Column 1 shows the results of the baseline model

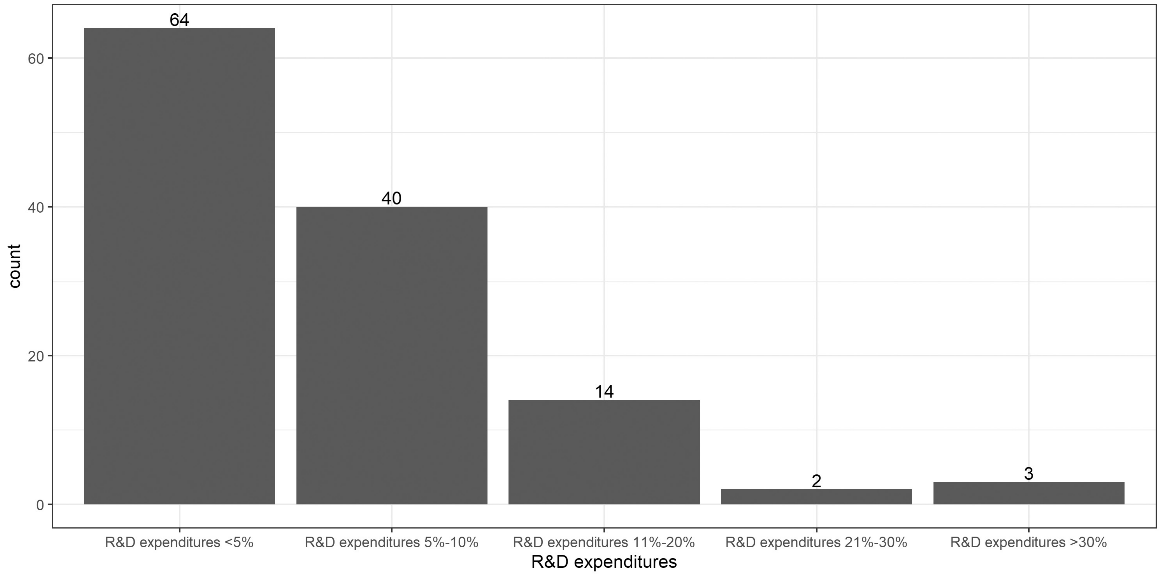

Figure 5. Frequency distribution of R&D expenditures' classes.

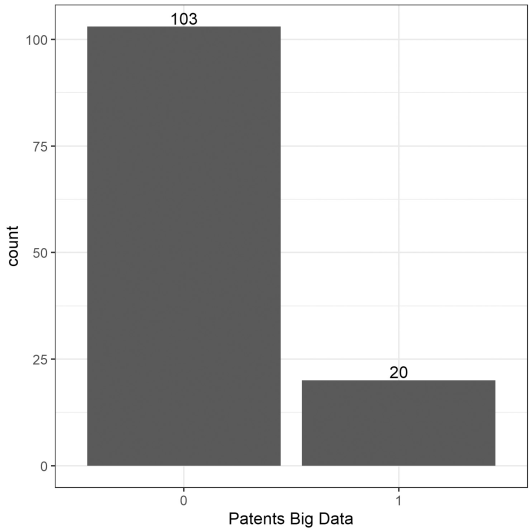

Figure 6. Frequency distribution of Patents Big Data.

in which we regress the 3 "Vs" of Big Data (i.e., *Big Data Volume*, *Big Data Variety*, and *Big Data Velocity*) on *OI Breadth*. Column 1 indicates that the estimated coefficient associated with *Big Data Volume* is negative but non-statistically significant ($\hat{\beta}=-0.0847$, p-value>0.1). *Big Data Velocity* is positive and non-statistically significant ($\hat{\beta}=0.0461$, p-value>0.1). Then, *Big Data Variety* is positive and statistically significant ($\hat{\beta}=0.0821$, p-value<0.1). However, the baseline model in Column 1 is affected by omitted variable bias, underestimating the coefficient of the independent variables. Hence, to solve this issue, we introduced a set of control variables. In Column 2, we added categorical variables to address the role of *R&D expenditures* on *OI Breadth* because the ratio of R&D expenditures to total revenues indicates firm's ability to capture and effectively leverage external knowledge (Zahra and George, 2002; Cassetta et al., 2020). In Column 3, we also controlled for *Age* and *Size* of the firms to account for the heterogeneity of Italian SMEs. In Column 4, we introduced the dummy *Patents Big Data* to control firms' application of patents developed using Big Data. In Column 5, we also assessed if the SMEs operate in a high-tech industry by introducing the dummy *Technology intensity industry*, while in Column 6, we controlled for the *Industry R&D intensity*. We did not find stable effects of the control variables. However, by controlling for omitted variables, from Column 2 to Column 6, we have been able to overcome the underestimation of the coefficients due to the presence of omitted variable bias in Column 1. Column 2 indicates that the estimated coefficient associated with *Big Data Volume* is negative and statistically significant ($\hat{\beta}=-0.0983$, p-value<0.1), while *Big Data Velocity* and *Big Data Variety* are positive and statistically significant ($\hat{\beta}=0.0533$, p-value <0.1; $\hat{\beta}=0.0934$, p-value<0.1, respectively). The magnitude and statistical significance of *Big Data Volume, Big Data Velocity, and Big Data Variety* then are confirmed from Column 3 to Column 6 and, as a result, H1, H2, and H3 are confirmed. More specifically, the impact of *Big Data Volume* on SMEs' *OI breadth* is negative and statistically significant, thereby confirming H1. In other terms, the availability of large size of data diminishes the development of inter-organizational relationships. Moreover, the introduction of omitted variables leads us to confirm H2 due to *Big Data Velocity* being positive and statistically significant. Hence, *Big Data Velocity* stimulates the search for the appropriate external partner in the Italian SMEs. Lastly, *Big Data Variety* is positive and statistically significant, confirming H3 and indicating that different types of data encourage the search for inter-organizational relationships.

Table 4 reports the results of the negative binomial regression models on the variable *OI Depth*. Contrarily to the findings pertaining to *OI Breadth*, even after accounting for potential omitted variables from Column 2 to Column 6, our analysis shows that *Big Data Volume*, *Big Data Velocity*, and *Big Data Variety* exhibit positive associations, yet these relationships are not statistically significant. Consequently, we do not find support for H4, H5, and H6. Results for hypotheses testing are reported in Table 5.

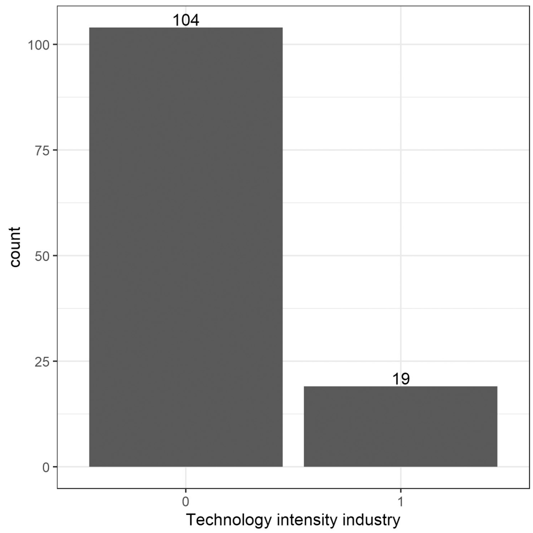

Figure 7. Frequency distribution of technology intensity industry.

5. Discussion

In this paper, we contribute to previous studies by considering the role of the "3Vs" of Big Data. Drawing on extant literature, we have examined the impact of Big Data Volume, Velocity, and Variety on SMEs' OI search strategies, being them OI breadth and depth. Both explanatory and dependent variables have been computed by following well-known measures in Big Data and OI literature (Ghasemaghaei, 2021; Messeni Petruzzelli et al., 2022). Albeit more objectives measures can be used (Acciarini et al., 2023), our paper might be the first attempt to develop a comprehensive recognition about the influence of Big Data characteristics on OI in the context of Italian SMEs.

Our results show that each Big Data characteristic impacts SMEs' OI breadth, supporting Hypotheses 1, 2, and 3. More specifically, Big Data Velocity and Variety positively affect SMEs' OI breadth, while Big Data Volume has a negative impact. These outcomes are particularly interesting for three reasons. First, they confirm what has been found in the work by Ghasemaghaei and Calic (2020), that is: "Big Data is not always better data" (Ghasemaghaei and Calic, 2020, p. 158). Second, our findings enhance the generalizability of this argument in relationships, whereas the dependent variable is not firm's

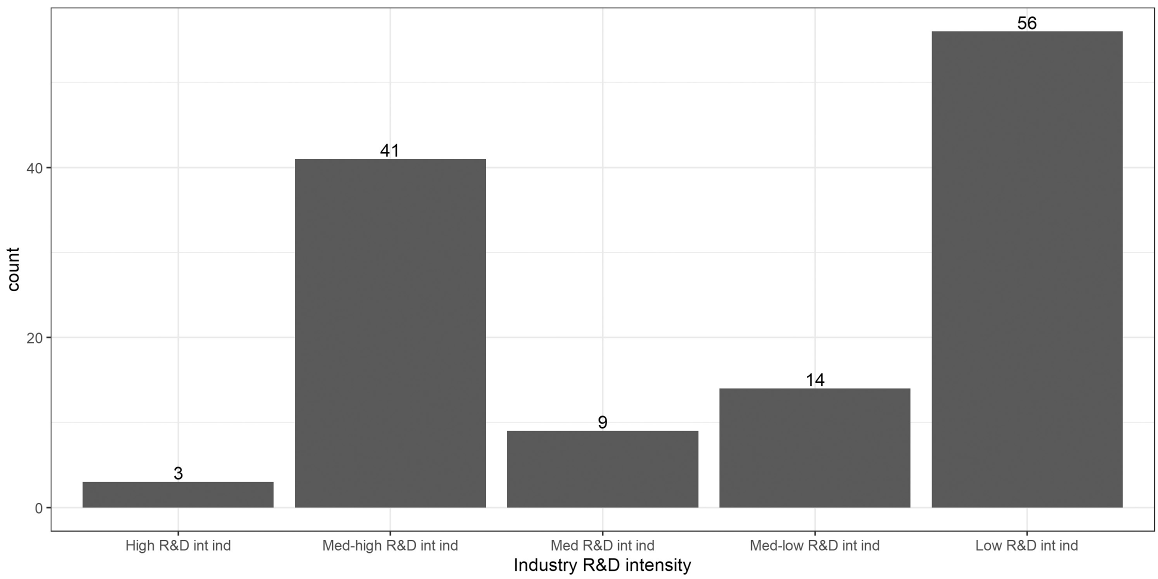

Figure 8. Frequency distribution of R&D intensity.

Table 2. VIF and correlation matrix

Variables	VIF	OI breadth	OI depth	Big Data Volume	Big Data Velocity	Big Data Variety	R&D exp <5%
OI breadth		1.0000					
OI depth		0.4573	1.0000				
Big Data Volume	2.28	−0.0217	0.1137	1.0000			
Big Data Velocity	1.53	0.0605	0.1037	0.5348	1.0000		
Big Data Variety	1.87	0.0995	0.1277	0.6512	0.3431	1.0000	
R&D exp <5%	1.21	−0.1315	−0.1470	−0.0573	−0.1247	0.0035	1.0000
R&D exp 5%–10%	1.26	0.0522	0.0669	0.0151	0.0323	0.0128	−0.6869
R&D exp 11%–20%	1.19	0.1507	0.1544	0.0356	−0.0228	−0.0481	−0.3757
R&D exp 21%–30%	1.09	−0.0380	−0.0454	0.0315	0.0980	−0.0156	−0.1594
R&D exp >30%	1.19	−0.0112	−0.0067	0.0286	0.2235	0.0588	−0.1848
Age	1.11	−0.1339	−0.1284	0.0149	0.0308	0.0418	0.1282
Size	1.13	0.0579	−0.1656	0.1991	0.1299	0.1897	−0.0313
Patents Big Data	1.13	0.1051	0.1078	0.0919	0.0895	0.1371	−0.1082
Tech int ind	1.08	−0.0171	−0.0331	0.0600	0.1114	0.0720	−0.0887
High R&D int ind	1.23	−0.0604	−0.0412	−0.0267	−0.0193	0.0693	−0.1848
Med-high R&D int ind	1.32	0.0296	−0.1166	0.0694	0.0539	−0.0317	0.0284
Med R&D int ind	1.15	−0.0498	0.0313	0.0526	−0.0249	0.0321	0.0829
Med-low R&D int ind	1.22	−0.0201	0.0291	−0.0432	−0.0139	0.0616	0.1617
Low R&D int ind	1.06	0.0339	0.0907	−0.0612	−0.0219	−0.0497	−0.1055

Variables	R&D exp 5%–10%	R&D exp 11%–20%	R&D exp 21%–30%	R&D exp >30%	Age	Size	Patents Big Data
R&D exp 5%–10%	1.0000						
R&D exp 11%–20%	−0.2539	1.0000					
R&D exp 21%–30%	−0.1077	−0.0589	1.0000				
R&D exp >30%	−0.1249	−0.0683	−0.0290	1.0000			
Age	0.0317	−0.1549	−0.0985	−0.0730	1.0000		
Size	−0.0049	0.0635	−0.0674	0.0425	−0.0339	1.0000	
Patents Big Data	0.0765	−0.0513	0.0593	0.1472	0.1705	0.0476	1.0000
Tech int ind	0.1379	−0.1091	0.0637	0.0340	0.0169	0.1027	−0.1150
High R&D int ind	0.2691	−0.0683	−0.0290	−0.0336	0.1287	−0.1122	0.0296
Med-high R&D int ind	−0.0117	−0.1112	−0.0027	0.1580	−0.0324	0.1411	−0.1255
Med R&D int ind	−0.0911	0.0573	−0.0496	−0.0575	0.0375	−0.0013	0.0689
Med-low R&D int ind	−0.1063	−0.1225	0.1256	−0.0603	0.0148	−0.1868	−0.0171
Low R&D int ind	0.0309	0.1710	−0.0337	−0.0698	−0.0454	0.0177	0.0799

Variables	Tech int ind	High R&D int ind	Med-high R&D int ind	Med R&D int ind	Med-low R&D int ind	Low R&D int ind
Tech int ind	1.0000					
High R&D int ind	0.2732	1.0000				
Med-high R&D int ind	−0.1124	−0.1320	1.0000			
Med R&D int ind	−0.1463	−0.0575	−0.2257	1.0000		
Med-low R&D int ind	−0.0105	−0.0603	−0.2368	−0.1030	1.0000	
Low R&D int ind	0.1001	−0.1622	−0.6370	−0.2772	−0.2909	1.0000

innovation performance but a firm's process-related variable, that is, OI breadth. Third, these findings reveal that both speed of data and a heterogeneous degree of knowledge lead a SME to stimulating the search for the right ideas and creative solutions enhancing the decision-making process within the

Table 3. OI breadth negative binomial results

	Dependent variables: OI breadth					
	(1)	(2)	(3)	(4)	(5)	(6)
Big Data Volume	−0.0847	−0.0983*	−0.1033*	−0.1039*	−0.1041*	−0.1106*
	(0.0547)	(0.0558)	(0.0573)	(0.0584)	(0.0610)	(0.0651)
Big Data Velocity	0.0461	0.0533*	0.0582**	0.0598**	0.0600**	0.0578*
	(0.0317)	(0.0273)	(0.0268)	(0.0285)	(0.0300)	(0.0312)
Big Data Variety	0.0821*	0.0934*	0.0941**	0.0877*	0.0879*	0.0974*
	(0.0489)	(0.0496)	(0.0474)	(0.0475)	(0.0485)	(0.0591)
R&D exp 5%–10%		0.1222	0.1185	0.1009	0.1021	0.1253
		(0.1282)	(0.1252)	(0.1229)	(0.1242)	(0.1464)
R&D exp 11%–20%		0.3352*	0.2925	0.2882	0.2875	0.3176
		(0.1740)	(0.1959)	(0.1928)	(0.1895)	(0.1968)
R&D exp 21%–30%		−0.0732	−0.1245	−0.1918	−0.1899	−0.1960
		(0.4144)	(0.4231)	(0.4948)	(0.4998)	(0.5191)
R&D exp >30%		−0.0661	−0.1182	−0.1965	−0.1960	−0.2485
		(0.5237)	(0.5085)	(0.5225)	(0.5222)	(0.4880)
Age			−0.0036	−0.0044	−0.0044*	−0.0041
			(0.0025)	(0.0027)	(0.0027)	(0.0025)
Size			0.0005	0.0004	0.0004	0.0001
			(0.0012)	(0.0012)	(0.0013)	(0.0010)
Patents Big Data				0.1901	0.1886	0.2148**
				(0.1191)	(0.1218)	(0.1051)
Tech int ind					−0.0096	
					(0.1173)	
Med-high R&D int ind						0.3700
						(0.2615)
Med R&D int ind						0.1977
						(0.3262)
Med-low R&D int ind						0.2837
						(0.3358)
Low R&D int ind						0.2673
						(0.3197)
Constant	0.9566***	0.8563***	0.9643***	0.9993***	0.9985***	0.7060**
	(0.2741)	(0.3011)	(0.3422)	(0.3453)	(0.3445)	−0.3417
Num. Obs.	123	123	123	123	123	123
AIC	488.33	491.35	493.05	493.07	495.06	499.2
BIC	502.39	516.66	523.99	526.81	531.62	544.2
Loglikelihood	−293.16	−236.67	−235.53	−234.53	−234.53	−233.6
Wald test	468.7***	497.1***	511.6***	523.3***	523.3	527.3***
Pseudo R^2	0.032	0.072	0.090	0.106	0.106	0.12

Bootsrapped cluster-robust standard errors with 1,000 replications.
*Statistical significance at 10% level.
**Statistical significance at 5% level.
***Statistical significance at 1% level.

company, thus influencing collaboration opportunities (Erevelles et al., 2016; Pedota, 2023).

However, contrary to the expectations, none of the "3Vs" affect SMEs' OI depth, rejecting Hypotheses 4, 5, and 6. We did not find any statistically significant effect. Similar results have been found in the work by Johnson et al. (2017), which evidences that a firm's exploitation orientation exerts no effect on the 3Vs of their big data usage. These results may be explicable through the recognition that possessing an adequate reservoir of data with requisite attributes, encompassing volume, real-time velocity, and diverse variety,

Table 4. OI depth negative binomial results

	Dependent variables: OI depth					
	(1)	(2)	(3)	(4)	(5)	(6)
Big Data Volume	0.0097	−0.0018	0.0033	0.0028	0.0028	0.0040
	(0.0829)	(0.0922)	(0.0860)	(0.0880)	(0.0921)	(0.0933)
Big Data Velocity	0.0303	0.0372	0.0477	0.0507	0.0506	0.0486
	(0.0769)	(0.0779)	(0.0745)	(0.0739)	(0.0771)	(0.0745)
Big Data Variety	0.0420	0.0518	0.0667	0.0599	0.0598	0.0620
	(0.0574)	(0.0605)	(0.0537)	(0.0568)	(0.0581)	(0.0645)
R&D exp 5%–10%		0.1533	0.1520	0.1301	0.1297	0.1693
		(0.1217)	(0.1235)	(0.1068)	(0.1041)	(0.1183)
R&D exp 11%–20%		0.3481	0.3423	0.3343	0.3345	0.3289
		(0.2407)	(0.2652)	(0.2635)	(0.2722)	(0.2803)
R&D exp 21%–30%		−0.1341	−0.2640	−0.3317	−0.3328	−0.3287
		(0.3926)	(0.2988)	(0.3612)	(0.3812)	(0.3603)
R&D exp >30%		−0.0094	−0.0464	−0.1186	−0.1193	−0.0666
		(0.4670)	(0.3716)	(0.4565)	(0.4635)	(0.4273)
Age			−0.0040	−0.0050	−0.0050	−0.0047
			(0.0030)	(0.0032)	(0.0032)	(0.0032)
Size			−0.0036**	−0.0037**	−0.0037**	−0.0038***
			(0.0015)	(0.0015)	(0.0015)	(0.0015)
Patents Big Data				0.2169	0.2174	0.1956
				(0.2578)	(0.2616)	(0.2509)
Tech int ind					0.0034	
					(0.1413)	
Med-high R&D int ind						0.2427
						(0.3218)
Med R&D int ind						0.3671
						(0.3314)
Med-low R&D int ind						0.3545
						(0.2733)
Low R&D int ind						0.3363
						(0.3174)
Constant	0.3310	0.2151	0.5997**	0.6382**	0.6385**	0.3125
	(0.2460)	(0.2455)	(0.2891)	(0.3164)	(0.3178)	(0.3635)
Num. Obs.	123	123	123	123	123	123
AIC	405.39	409.4	406.63	406.9	408.9	413.7
BIC	419.45	434.71	437.57	440.64	445.46	458.7
Loglikelihood	−197.69	−195.7	−192.32	−191.45	−191.45	−190.86
Wald test	133.4***	140.3***	151.8***	154.7***	154.7***	156.6***
Pseudo R^2	0.024	0.068	0.142	0.161	0.161	0.174

Bootsrapped cluster-robust standard errors with 1,000 replications.
*Statistical significance at 10% level.
**Statistical significance at 5% level.
***Statistical significance at 1% level.

within a SME, imparts a comprehensive understanding of the potential environmental threats and opportunities. This heightened awareness, in turn, may prompt SMEs to actively seek and acquire finely tailored external knowledge tailored to its specific context. Consequently, this obviates the necessity for intensifying interorganizational relationships in pursuit of domain-relevant knowledge, as the firm is already equipped with the capacity to address its distinctive informational requirements effectively.

The above-discussed results underscore that Big Data can be a valuable resource that can bring

Table 5. Results for hypothesis testing

Hypotheses	Confirmed
3Vs of Big Data and SMEs' OI breadth	
H1: Big Data Volume is negatively correlated with SMEs' OI breadth	Yes
H2: Big Data Velocity is positively correlated with SMEs' OI breadth	Yes
H3: Big Data Variety is positively correlated with SMEs' OI breadth	Yes
3Vs of Big Data and SMEs' OI depth	
H4: Big Data Volume is negatively correlated with SMEs' OI depth	No
H5: Big Data Velocity is positively correlated with SMEs' OI depth	No
H6: Big Data Variety is negatively correlated with SMEs' OI depth	No

SMEs to activate external collaborations with different types of organizations, thereby overcoming the limitations that attain to their size, resource constraints, and managerial ambitions (Spithoven et al., 2013; Dooley and O'Sullivan, 2018): that is, when it is characterized by high velocity and variety. However, Big Data only contributes to the formation of new collaborations not to their strengthening. SMEs can eventually collaborate more often with their partners in the presence of a high value of veracity, further benefiting the "quality" of the collaboration. In the following subsections, we document papers' contributions and limitations and identify some avenues for further research.

5.1. Contributions

Based on the above results, this article provides three substantive theoretical contributions. First, our article augments the corpus of literature dedicated to OI by offering insightful revelations that facilitate a more nuanced comprehension of the nascent association between Big Data and OI, as delineated in extant studies (Enkel et al., 2020; Acciarini et al., 2023; Cappa et al., 2023). Differently from the findings advanced by Messeni Petruzzelli et al. (2022), who have posited that OI breadth engenders the propensity of SMEs to embrace Industry 4.0 technologies, encompassing Big Data analytics, our thesis posits, and substantiates, an opposite proposition. Specifically, in accordance with the empirical underpinnings of our research, it is evident that the salient attributes of Big Data – that is, Volume, Velocity, and Variety – exert a discernible influence upon the ambit of OI initiatives undertaken by SMEs. More importantly, the theoretical salience of our findings contributes to the R&D management field and more specifically to the work by Enkel et al. (2020). In fact, these authors underscore the need for deeper inquiry into "which influence has big or linked data on OI" (Enkel et al., 2020, p. 165). In our study, we tackle this issue by examining the impact of the 3Vs of Big Data on SMEs' OI search strategies. Moreover, our findings contribute to other studies close to the R&D management field that convey suggestions converging upon the notion that, within the milieu of SMEs, "Big Data can support the conception and execution of an OI strategy for making companies more competitive (Chesbrough, 2011; Ollila and Elmquist, 2011) and opening new entrepreneurial opportunities (Eftekhari and Bogers, 2015)" (Del Vecchio et al., 2018, p. 10). As such, our study also suggests that Big Data can represent a valuable internal resource for SMEs, thereby contributing to extant studies claiming the urgency to develop a RBV of Big Data (Akhtar et al., 2019).

Second, our article contributes to the Big Data literature by suggesting that the influence of Big Data on SMEs' OI is not always positive (Johnson et al., 2017; Ghasemaghaei and Calic, 2019). In other terms, taking into consideration that the Big Data characteristics can reveal the "dark side" of this technology thereby implies that their use has not always good outcomes (Ghasemaghaei and Calic, 2019). As such, our study suggests that while the velocity and variety of data, that SMEs process, bring them to form OI relationships with different partners, the volume of data processed implies a negative effect on OI breadth. As a net result, the influence of Big Data depends on the Big Data characteristic we consider. This finding seems to be in line with previous literature investigating the impact of Big Data characteristics on innovation performance (Ghasemaghaei and Calic, 2020). More specifically, our findings augment the generalizability of the thesis "Big Data is not always better data" (Ghasemaghaei and Calic, 2020, p. 158) in relationships, whereas the dependent variable is not firm's innovation performance but a firm's process-related variable, namely, OI breadth. Moreover, in line with previous studies, we found that both data velocity and data variety lead a SME to stimulate the search for the right ideas and creative solutions enhancing the decision-making process within the company, thereby influencing collaboration opportunities (Erevelles et al., 2016; Pedota, 2023).

Third, our article enriches the flourishing portion of literature investigating OI in the context of SMEs. The extant body of literature predominantly relies upon empirical substantiation gleaned from extensive investigations conducted within the milieu of large firms (Spithoven et al., 2013). Prevalent paradigms of best practices in the realm of OI have been extensively documented and applied in diverse sectors, including manufacturing (Laursen and Salter, 2006), healthcare (Hughes and Wareham, 2010), pharmaceuticals (Bianchi et al., 2011), automotive (Ili et al., 2010), and the food industry (Sarkar and Costa, 2008). It is noteworthy, however, that the diffusion and implementation of OI practices within the context of SMEs can be favored by the rapid adoption of Big Data. SMEs can reveal unexpected and surprising results since their limitations and flexibility might lead them to innovative solutions in this context. As such, the findings of our article seem to reveal some nuances that contribute to recent studies calling for more quantitative studies on the impact of Big Data on OI in the specific context of SMEs (Del Vecchio et al., 2018). Utilizing Big Data can serve as a valuable resource for SMEs, enabling them to initiate partnerships with a diverse range of organizations. This allows them to overcome the constraints associated with their size, resource limitations, and managerial aspirations, as emphasized by previous studies (Spithoven et al., 2013; Dooley and O'Sullivan, 2018). This is particularly true when Big Data is marked by rapid data generation and diverse sources. Nevertheless, it is essential to note that while Big Data facilitates the creation of new collaborations, it does not necessarily enhance existing ones.

This study also offers valuable practical insights for SMEs' owners-managers. In fact, our findings underscore how SMEs can benefit from leveraging Big Data to initiate and facilitate external collaborations, which can enhance their innovation efforts. However, SMEs' decision makers should be cautious about expecting Big Data to directly impact the depth of these collaborations. Moreover, they should carefully consider how to utilize Big Data in their innovation and collaboration strategies, taking into account that the impact may vary depending on the specific characteristics of the Big Data involved.

5.2. *Limitations and future research directions*

Although this paper may contribute to a better understanding of Big Data characteristics affecting SMEs' propension to adopt OI, a few limitations must be taken into account: first, the OI typology on which we based our study. We are aware that extant studies recognize other OI practices that we do not consider in our research. For instance, Gassmann et al. (2010) take a "process perspective" by discussing these practices in terms of inbound, outbound, and coupled OI processes. Future studies may replicate our approach by also considering alternative OI practices.

Second, the "Vs" of Big Data characteristics we consider. As noted by Ghasemaghaei (2021), both Big Data Veracity and Big Data Value are crucial characteristics that work as functions of Big Data Volume, Variety, and Velocity. It would be interesting to understand whether they mediate or moderate the relationships we found in our study or they are more related to OI depth.

Third, we have examined the effects of Big Data on SMEs' OI in a very specific context: Italian SMEs. We recognize that the generalizability of our results is limited to this type of organization operating in this specific country. Future studies could investigate what are the effects of Big Data characteristics on SMEs' OI in other countries. Likewise, this study focuses on SMEs. Future research may examine the impact of Big Data characteristics in other types of firms, such as older firms, large firms, or start-ups.

Fourth, these effects are analyzed separately instead of following an integrated logic. Albeit the VIF test has shown that there are no critical values of multicollinearity between Big Data Volume, Big Data Velocity, Big Data Variety, we are completely aware about the fact that there could be an important explanation effect by considering the interaction of the Vs (Cappa et al., 2021). Therefore, we suggest that future research may conduct a Qualitative Comparative Analysis to enrich our understanding of the impact of the 3Vs of Big Data on firm's innovation outcomes.

Fifth, other Industry 4.0 technologies may affect SMEs' OI. In this study, we have focused our attention on Big Data. It might be interesting to study whether the impact of other Industry 4.0 technologies on SMEs' OI differ by considering the technological features that epitomize a specific technology. For instance, future studies may explore whether Industry 4.0 base technologies and front-end technologies (Frank et al., 2019) impact differently on SMEs' OI.

Sixth, a limitation of the paper is its exclusive focus on SMEs' strategies for addressing Big Data, overlooking detailed insights into data collection and analysis methods. This raises important questions for future research: How do SMEs collect and analyze data? Which business functions predominantly engage with Big Data in SMEs? Exploring these areas could offer a more comprehensive understanding of Big Data's role in SMEs and its operational and strategic implications.

Finally, in this study, we have analyzed what Big Data characteristics affect SMEs' propensity to adopt OI. However, we have not investigated whether Big Data characteristics impact on OI can be reflected in SME's innovation performance. Future studies may conduct econometric analysis that can assess the mediation effects of Big Data characteristics on the relationship between the SMEs OI and their innovation performance.

Acknowledgments

We wish to thank participants to the 2023 SIMA Conference in Bari, the 2023 R&D Management Conference in Seville, and research seminars held at Scuola Superiore Sant'Anna for their constructive comments on earlier versions of this paper. We also acknowledge funding from the EU Horizon2020 research and innovation program under grant agreement no. 871042, SoBigData-PlusPlus. All flaws are ours. Open access publishing facilitated by Scuola Superiore Sant'Anna, as part of the Wiley - CRUI-CARE agreement.

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Acciarini, C., Cappa, F., Boccardelli, P., and Oriani, R. (2023) How can organizations leverage big data to innovate their business models? A systematic literature review. *Technovation*, **123**, 102713.

Agostini, L. and Filippini, R. (2019) Organizational and managerial challenges in the path toward industry 4.0. *European Journal of Innovation Management*, **22**, 3, 406–421.

Ahn, J.M., Minshall, T., and Mortara, L. (2017) Understanding the human side of openness: the ft between open innovation modes and CEO characteristics. *R&D Management*, **47**, 727–740.

Akhtar, P., Frynas, J.G., Mellahi, K., and Ullah, S. (2019) Big data-savvy teams' skills, big data-driven actions and business performance. *British Journal of Management*, **30**, 2, 252–271.

Arbore, A. and Ordanini, A. (2006) Broadband divide among SMEs: the role of size, location and outsourcing strategies. *International Small Business Journal*, **24**, 1, 83–99.

Aslesen, H.W. and Freel, M. (2012) Industrial knowledge bases as drivers of open innovation? *Industry and Innovation*, **19**, 7, 563–584.

Barney, J. (1991) Firm resources and sustained competitive advantage. *Journal of Management*, **17**, 1, 99–120.

Barrett, G., Dooley, L., and Bogue, J. (2021) Open innovation within high-tech SMEs: a study of the entrepreneurial founder's influence on open innovation practices. *Technovation*, **103**, 102232.

Batistič, S. and van der Laken, P. (2019) History, evolution and future of big data and analytics: a bibliometric analysis of its relationship to performance in organizations. *British Journal of Management*, **30**, 2, 229–251.

Bianchi, M., Cavaliere, A., Chiaroni, D., Frattini, F., and Chiesa, V. (2011) Organisational modes for open innovation in the bio-pharmaceutical industry: an exploratory analysis. *Technovation*, **31**, 1, 22–33.

Big Data Value Association. (2017) *SMEs in the European Data Economy – A Representative Sample of the BDVA Community*. Available at: https://www.big-data-value.eu/wp-content/uploads/2018/01/SMEs-Brochure-2017.pdf [Accessed 23th March 2024].

Bogers, M., Zobel, A.-K., Afuah, A., Almirall, E., Brunswicker, S., Dahlander, L., Frederiksen, L., Gawer, A., Gruber, M., Haefliger, S., Hagedoorn, J., Hilgers, D., Laursen, K., Magnusson, M.G., Majchrzak, A., McCarthy, I.P., Moeslein, K.M., Nambisan, S., Piller, F.T., Radziwon, A., Rossi-Lamastra, C., Sims, J., and Ter Wal, A.L.J. (2017) The open innovation research landscape: established perspectives and emerging themes across different levels of analysis. *Industry and Innovation*, **24**, 1, 8–40.

Burt, R.S. (1992) *Structural Holes: The Social Structure of Competition*. Cambridge, MA: Harvard University Press.

Cameron, A.C., Gelbach, J.B., and Miller, D.L. (2008) Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics*, **90**, 3, 414–427.

Cameron, A.C. and Trivedi, P.K. (2013) *Regression Analysis of Count Data*, Vol. **53**. Cambridge, UK: Cambridge University Press.

Cappa, F., Boccardelli, P., Oriani, R., and Acciarini, C. (2023) How can organizations leverage value from big data? A systematic literature review. *Technovation*, **123**, 1–18.

Cappa, F., Oriani, R., Peruffo, E., and McCarthy, I. (2021) Big data for creating and capturing value in the digitalized environment: unpacking the effects of volume, variety, and veracity on firm performance. *Journal of Product Innovation Management*, **38**, 1, 49–67.

Carrasco-Carvajal, O., Castillo-Vergara, M., and García-Pérez-de-Lema, D. (2023) Measuring open innovation in SMEs: an overview of current research. *Review of Managerial Science*, **17**, 2, 397–442.

Cassetta, E., Monarca, U., Dileo, I., Di Berardino, C., and Pini, M. (2020) The relationship between digital technologies and internationalisation. Evidence from Italian SMEs. *Industry and Innovation*, **27**, 4, 311–339.

Cepa, K. (2021) Understanding interorganizational big data technologies: how technology adoption motivations and technology design shape collaborative dynamics. *Journal of Management Studies*, **58**, 7, 1761–1799.

Chen, C.P. and Zhang, C.Y. (2014) Data-intensive applications, challenges, techniques and technologies: a survey on big data. *Information Sciences*, **275**, 314–347.

Chesbrough, H. (2011) *Open Services Innovation: Rethinking Your Business to Grow and Compete in a New Era*. San Francisco, CA: Jossey-Bass.

Chesbrough, H.W. (2003) *Open Innovation: The New Imperative for Creating and Profiting from Technology*. Boston, MA: Harvard Business Press.

Costa, A., Crupi, A., De Marco, C.E., and Di Minin, A. (2023) SMEs and open innovation: challenges and costs of engagement. *Technological Forecasting and Social Change*, **194**, 122731.

Crick, J.M., Crick, D., and Ferrigno, G. (2023) Coopetition and the marketing/entrepreneurship interface in an international arena. *International Journal of Entrepreneurial Behavior & Research*, forthcoming.

Dagnino, G.B., Picone, P.M., and Ferrigno, G. (2021) Temporary competitive advantage: a state-of-the-art literature review and research directions. *International Journal of Management Reviews*, **23**, 1, 85–115.

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

Dahlander, L., Gann, D.M., and Wallin, M.W. (2021) How open is innovation? A retrospective and ideas forward. *Research Policy*, **50**, 4, 104218.

Damij, N., Hafner, A., and Modic, D. (2022) Activity-to-skills framework in the intellectual property big data era. *IEEE Transactions on Engineering Management*, **71**, 13251–13265.

Del Sarto, N., Ferrigno, G., Parida, V., and Di Minin, A. (2023) Do start-ups benefit from coworking spaces? An empirical analysis of accelerators' programs. *Review of Managerial Science*, **17**, 2471–2502.

Del Vecchio, P., Di Minin, A., Petruzzelli, A.M., Panniello, U., and Pirri, S. (2018) Big data for open innovation in SMEs and large corporations: trends, opportunities, and challenges. *Creativity and Innovation Management*, **27**, 1, 6–22.

Deloitte Consulting. (2012) *The Insight Economy: Big Data Matters – Except When It Doesn't*. Available at: https://www2.deloitte.com/content/dam/Deloitte/global/Documents/Deloitte-Analytics/dttl-analytics-us-ba-insight-economy-10012012.pdf [Accessed 28th April 2024].

Dooley, L. and O'Sullivan, D. (2018) Open innovation within the low-technology SME sector. In: Editors, Vanhaverbeke, W., Frattini, F., Roijakkers, N., & Usman, M., *Researching Open Innovation in SMEs*. Singapore: World Scientific Publishing, pp. 249–271.

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

Eller, R., Alford, P., Kallmünzer, A., and Peters, M. (2020) Antecedents, consequences, and challenges of small and medium-sized enterprise digitalization. *Journal of Business Research*, **112**, 119–127.

Enkel, E., Bogers, M., and Chesbrough, H. (2020) Exploring open innovation in the digital age: a maturity model and future research directions. *R&D Management*, **50**, 1, 161–168.

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

Erevelles, S., Fukawa, N., and Swayne, L. (2016) Big data consumer analytics and the transformation of marketing. *Journal of Business Research*, **69**, 2, 897–904.

Ferrigno, G., Crupi, A., Di Minin, A., and Ritala, P. (2023a) 50+ years of R&D management: a retrospective synthesis and new research trajectories. *R&D Management*, **53**, 5, 900–926.

Ferrigno, G., Del Sarto, N., Piccaluga, A., and Baroncelli, A. (2023b) Industry 4.0 base technologies and business models: a bibliometric analysis. *European Journal of Innovation Management*, **26**, 7, 502–526.

Frank, A.G., Dalenogare, L.S., and Ayala, N.F. (2019) Industry 4.0 technologies: implementation patterns in manufacturing companies. *International Journal of Production Economics*, **210**, 15–26.

Galindo-Rueda, F. and Verger, F. (2016) OECD taxonomy of economic activities based on R&D intensity. In: *OECD Science, Technology and Industry Working Papers*, 2016/04. Paris: OECD Publishing.

Garg, A., Grande, D., Miranda, G.M.L., Sporleder, C., and Windhagen, E. (2017) *Analytics in Banking: Time to Realize the Value*. McKynsey&Company. Availabe at: https://www.mckinsey.com/industries/financial-services/our-insights/analytics-in-banking-time-to-realize-the-value [Accessed 28th April 2024].

Garriga, H., Von Krogh, G., and Spaeth, S. (2013) How constraints and knowledge impact open innovation. *Strategic Management Journal*, **34**, 9, 1134–1144.

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

Ghasemaghaei, M. (2019) Are firms ready to use big data analytics to create value? The role of structural and psychological readiness. *Enterprise Information Systems*, **13**, 5, 650–674.

Ghasemaghaei, M. (2021) Understanding the impact of big data on firm performance: the necessity of conceptually differentiating among big data characteristics. *International Journal of Information Management*, **57**, 102055.

Ghasemaghaei, M. and Calic, G. (2019) Does big data enhance firm innovation competency? The mediating role of data-driven insights. *Journal of Business Research*, **104**, 69–84.

Ghasemaghaei, M. and Calic, G. (2020) Assessing the impact of big data on firm innovation performance: big data is not always better data. *Journal of Business Research*, **108**, 147–162.

Ghasemaghaei, M., Hassanein, K., and Turel, O. (2017) Increasing firm agility through the use of data analytics: the role of fit. *Decision Support Systems*, **101**, 95–105.

Giotopoulos, I., Kontolaimou, A., Korra, E., and Tsakanikas, A. (2017) What drives ICT adoption by SMEs? Evidence from a large-scale survey in Greece. *Journal of Business Research*, **81**, 60–69.

Gopalakrishnan, S. and Damanpour, F. (1994) Patterns of generation and adoption of innovation in organizations: contingency models of innovation attributes. *Journal of Engineering and Technology Management*, **11**, 2, 95–116.

Greco, M., Grimaldi, M., and Cricelli, L. (2015) Open innovation actions and innovation performance: a literature review of European empirical evidence. *European Journal of Innovation Management*, **18**, 2, 150–171.

Han, H. and Trimi, S. (2022) Towards a data science platform for improving SME collaboration through industry 4.0 technologies. *Technological Forecasting and Social Change*, **174**, 121242.

Hilbe, J.M. (2014) *Modeling Count Data*. Cambridge, UK: Cambridge University Press.

Horváth, D. and Szabó, R.Z. (2019) Driving forces and barriers of industry 4.0: do multinational and small and medium-sized companies have equal opportunities? *Technological Forecasting and Social Change*, **146**, 119–132.

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

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

Iqbal, M., Kazmi, S.H.A., Manzoor, A., Soomrani, A.R., Butt, S.H., and Shaikh, K.A. (2018) A study of big data for business growth in SMEs: opportunities & challenges. In: *2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*. Sukkur: IEEE, pp. 1–7.

ISTAT (2022) *Annuario Statistico Italiano 2022*. Roma: Istituto Nazionale Di Statistica.

Johnson, J.S., Friend, S.B., and Lee, H.S. (2017) Big data facilitation, utilization, and monetization: exploring the 3Vs in a new product development process. *Journal of Product Innovation Management*, **34**, 5, 640–658.

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

Kelly, D. and Amburgey, T.L. (1991) Organizational inertia and momentum: a dynamic model of strategic change. *Academy of Management Journal*, **34**, 3, 591–612.

Keupp, M.M. and Gassmann, O. (2009) Determinants and archetype users of open innovation. *R&D Management*, **39**, 4, 331–341.

Kiel, D., Müller, J.M., Arnold, C., and Voigt, K.I. (2017) Sustainable industrial value creation: benefits and challenges of industry 4.0. *International Journal of Innovation Management*, **21**, 8, 1740015.

Kraus, S., Kailer, N., Dorfer, J., and Jones, P. (2020) Open innovation in (young) SMEs. *The International Journal of Entrepreneurship and Innovation*, **21**, 1, 47–59.

Kutner, M.H., Nachtsheim, C.J., Neter, J., and Li, W. (2005) *Applied Linear Statistical Models*. New York: McGraw-Hill.

Laney, D. (2001) 3D data management: controlling data volume, velocity and variety. *META Group Research Note*, **6**, 70, 1.

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

Lisboa, A., Skarmeas, D., and Lages, C. (2011) Entrepreneurial orientation, exploitative and explorative capabilities, and performance outcomes in export markets: a resource-based approach. *Industrial Marketing Management*, **40**, 8, 1274–1284.

Luqman, A., Wang, L., Katiyar, G., Agarwal, R., and Mohapatra, A.K. (2024) Unpacking associations between positive-negative valence and ambidexterity of big data. Implications for firm performance. *Technological Forecasting and Social Change*, **200**, 123054.

Martinelli, A., Mina, A., and Moggi, M. (2021) The enabling technologies of industry 4.0: examining the seeds of the fourth industrial revolution. *Industrial and Corporate Change*, **30**, 1, 161–188.

Martín-Peña, M.L., Sánchez-López, J.M., Kamp, B., and Giménez-Fernández, E.M. (2023) The innovation antecedents behind the servitization–performance relationship. *R&D Management*, **53**, 3, 459–480.

Marullo, C., Ahn, J.M., Martelli, I., and Di Minin, A. (2022) Open for innovation: an improved measurement approach using item response theory. *Technovation*, **109**, 102338.

McAfee, A., Brynjolfsson, E., Davenport, T.H., Patil, D.J., and Barton, D. (2012) Big data: the management revolution. *Harvard Business Review*, **90**, 10, 60–68.

McKinsey Global Institute (2016) *The Age of Analytics: Competing in a Data-Driven World*. McKinsey&Company. Available at: https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-age-of-analytics-competing-in-a-data-driven-world [Accessed 16th April 2024].

Messeni Petruzzelli, A.M., Murgia, G., and Parmentola, A. (2022) How can open innovation support SMEs in the adoption of I4.0 technologies? An empirical analysis. *R&D Management*, **52**, 4, 615–632.

MISE (2020) *Relazione Sugli Investimenti Di Sostegno Alle Attività Economiche E Produttive 2020*. Roma: Ministero Dello Sviluppo Economico.

Morgan, R.E. and Berthon, P. (2008) Market orientation, generative learning, innovation strategy and business performance inter-relationships in bioscience firms. *Journal of Management Studies*, **45**, 8, 1329–1353.

Narula, R. (2004) R&D collaboration by SMEs: new opportunities and limitations in the face of globalisation. *Technovation*, **25**, 153–161.

Nguyen, T.H., Newby, M., and Macaulay, M.J. (2015) Information technology adoption in small business: confirmation of a proposed framework. *Journal of Small Business Management*, **53**, 1, 207–227.

Nieto, M.J., Santamaria, L., and Bammens, Y. (2023) Digitalization as a facilitator of open innovation: are family firms different? *Technovation*, **128**, 102854.

Olabode, O.E., Boso, N., Hultman, M., and Leonidou, C.N. (2022) Big data analytics capability and market performance: the roles of disruptive business models and competitive intensity. *Journal of Business Research*, **139**, 1218–1230.

Ollila, S. and Elmquist, M. (2011) Managing open innovation: exploring challenges at the interfaces of an open innovation arena. *Creativity and Innovation Management*, **20**, 4, 273–283.

Pedota, M. (2023) Big data and dynamic capabilities in the digital revolution: the hidden role of source variety. *Research Policy*, **52**, 7, 104812.
Prescott, M.E. (2016) Big data: innovation and competitive advantage in an information media analytics company. *Journal of Innovation Management*, **4**, 1, 92–113.
Rammer, C. and Es-Sadki, N. (2023) Using big data for generating firm-level innovation indicators-a literature review. *Technological Forecasting and Social Change*, **197**, 122874.
Randhawa, K., Wilden, R., and Hohberger, J. (2016) A bibliometric review of open innovation: setting a research agenda. *Journal of Product Innovation Management*, **33**, 6, 750–772.
Rust, R.T., Moorman, C., and Dickson, P.R. (2002) Getting return on quality: revenue expansion, cost reduction, or both? *Journal of Marketing*, **66**, 4, 7–24.
Sá, T., Ferreira, J.J., and Jayantilal, S. (2023) Open innovation strategy: a systematic literature review. *European Journal of Innovation Management*.
Sarkar, S. and Costa, A.I. (2008) Dynamics of open innovation in the food industry. *Trends in Food Science & Technology*, **19**, 11, 574–580.
Scuotto, V., Santoro, G., Bresciani, S., and Del Giudice, M. (2017) Shifting intra-and inter-organizational innovation processes towards digital business: an empirical analysis of SMEs. *Creativity and Innovation Management*, **26**, 3, 247–255.
Sena, V., Bhaumik, S., Sengupta, A., and Demirbag, M. (2019) Big data and performance: what can management research tell us? *British Journal of Management*, **30**, 2, 219–228.
Shin, S.J., Woo, J., and Rachuri, S. (2014) Predictive analytics model for power consumption in manufacturing. *Procedia CIRP*, **15**, 153–158.
Sivarajah, U., Kamal, M.M., Irani, Z., and Weerakkody, V. (2017) Critical analysis of big data challenges and analytical methods. *Journal of Business Research*, **70**, 263–286.
Spithoven, A., Vanhaverbeke, W., and Roijakkers, N. (2013) Open innovation practices in SMEs and large enterprises. *Small Business Economics*, **41**, 537–562.
Statista. (2023) *Amount of Data Created, Consumed, and Stored 2010–2020, with Forecasts to 2025*. Petroc Taylor. Available at: https://www.statista.com/statistics/871513/worldwide-data-created/#:~:text=The%20total%20amount%20of%20data,replicated%20reached%20a%20new%20high [Accessed 28th April 2024].
Tan, K.H. and Zhan, Y. (2017) Improving new product development using big data: a case study of an electronics company. *R&D Management*, **47**, 4, 570–582.
Terjesen, S. and Patel, P.C. (2017) In search of process innovations: the role of search depth, search breadth, and the industry environment. *Journal of Management*, **43**, 5, 1421–1446.
Urbinati, A., Chiaroni, D., Chiesa, V., and Frattini, F. (2020) The role of digital technologies in open innovation processes: an exploratory multiple case study analysis. *R&D Management*, **50**, 1, 136–160.
Vitari, C. and Raguseo, E. (2020) Big data analytics business value and firm performance: linking with environmental context. *International Journal of Production Research*, **58**, 18, 5456–5476.
Wadhwa, A. and Kotha, S. (2006) Knowledge creation through external venturing: evidence from the telecommunications equipment manufacturing industry. *Academy of Management Journal*, **49**, 4, 819–835.
Wamba, S.F., Akter, S., Edwards, A., Chopin, G., and Gnanzou, D. (2015) How 'big data' can make big impact: findings from a systematic review and a longitudinal case study. *International Journal of Production Economics*, **165**, 234–246.
Ward, J.S. and Barker, A. (2013) Undefined by data: a survey of big data definitions. *Arxiv Preprint*, Arxiv:1309.5821.
Wernerfelt, B. (1984) A resource-based view of the firm. *Strategic Management Journal*, **5**, 2, 171–180.
Whitler, K.A. (2018) *Why Too Much Data Is a Problem and How to Prevent It*. Forbes. Available at: https://www.forbes.com/sites/kimberlywhitler/2018/03/17/why-too-much-data-is-a-problem-and-how-to-prevent-it/#348d5d77755f [Accessed 28th April 2024].
Wibowo, S., Suryana, Y., and Kaltum, U. (2021) Value creation with big data in marketing: an empirical evidence on SMEs. *Asian Journal of Business and Accounting*, **14**, 2, 173–196.
Yildiz, B., Çiğdem, Ş., Meidutė-Kavaliauskienė, I., and Činčikaitė, R. (2024) The nexus of big data analytics, knowledge sharing, and product innovation in manufacturing. *Journal of Business Economics and Management*, **25**, 1, 66–84.
Zahra, S.A. and George, G. (2002) Absorptive capacity: a review, reconceptualization, and extension. *Academy of Management Review*, **27**, 2, 185–203.
Zangiacomi, A., Pessot, E., Fornasiero, R., Bertetti, M., and Sacco, M. (2020) Moving towards digitalization: a multiple case study in manufacturing. *Production Planning & Control*, **31**, 2–3, 143–157.

Notes

[1] We decided to analyze the individual attributes (3Vs) of Big Data separately instead of following an integrated logic according to the empirical work by Ghasemaghaei and Calic (2020), which suggests this kind of approach in investigating the effects of the 3Vs of Big Data on innovation outcomes. This may be related to the fact that, for instance, relying on a huge amount of data does not necessarily mean relying on up-to-date data and/or data based on multiple sources. Moreover, up-to-date data do not necessarily mean multiple sources of data. In this sense, the 3Vs of Big Data represent the main basic characteristics upon which other Vs can be built, like, for instance, "veracity" and "value" (Cappa et al., 2021; Ghasemaghaei, 2021). The VIF test that we will provide later in this work supports this discourse. However, we are completely aware about the fact that there could be an important explanation effect

while considering the interaction of the Vs (Cappa et al., 2021). Therefore, the individual analysis can be considered both a limitation of our study and a fruitful avenue for further research. We are grateful to Reviewer 3 for suggesting us to clarify this aspect in our paper.

[2] Alongside these three dimensions, the research has recently suggested the introduction of other two Vs, namely Data Veracity and Data Value (Ghasemaghaei and Calic, 2020; Ghasemaghaei, 2021). *Data veracity* arises from high quality of data (Ghasemaghaei and Calic, 2020). *Data value*, instead, emerges from high data in terms of volume, velocity, and variety. Despite their importance, we do not include these Big Data characteristics in our study because “despite the differences in conceptualization of the BDAC construct, most scholars seem to settle on a three-dimensional approach: volume, variety, and velocity (McAfee et al., 2012; Chen and Zhang, 2014; Johnson et al., 2017; Vitari and Raguseo, 2020). In following these studies, therefore, we conceptualize BDAC as a three-dimensional construct, with the component elements volume, variety, and velocity of big data” (Olabode et al., 2022, p. 1219).

[3] To test whether the final sample size is reliable, we calculated the minimum number of necessary samples for a finite population. The equation for calculating the sample size is: Minimum sample size $= \frac{N}{1+\frac{z^2 \times \hat{p}(1-\hat{p})}{\varepsilon^2 N}}$, where N represents the population size, z is the z score for the confidence interval, ε is the margin of error, and $\hat{p}$ the population proportion which represents the percentage of the value associated to the survey. In our case, N is equal to 19,722, the confidence interval is set at 95%, while the margin of error ε is set at 8%. The population proportion of firms implementing big data is unknown, but we infer it by calculating the ratio of sampled firms' implementing big data over the total number of firms contacted. Then, we set the population proportion at 26.22% (210/801), which provide 116 firms as a minimum sample size for our survey.

Giulio Ferrigno is a Senior Assistant Professor at Sant’Anna School of Advanced Studies of Pisa. He has held visiting positions at the University of Cambridge, Tilburg University, and the University of Umea. His main research themes include strategic alliances, big data, and Industry 4.0. His works have been published in *Small Business Economics, Technological Forecasting and Social Change, International Journal of Management Reviews, R&D Management, Technology Analysis & Strategic Management, Review of Managerial Science, European Journal of Innovation Management, International Journal of Entrepreneurial Behavior & Research, Journal of Small Business, and Enterprise development*. He is an Associate Editor of *Technology Analysis & Strategic Management*.

Saverio Barabuffi is a post-doc research fellow at Sant’Anna School of Advanced Studies of Pisa. He obtained a PhD in Economics at Università degli Studi di Roma 3. He has held visiting positions at University of Sussex. Before starting the PhD program, he studied economics at the University of Pisa (master degree). His research fields include econometrics, big data, and artificial intelligence, with a special focus on China.

Enrico Marcazzan is a post-doc research fellow at Sant’Anna School of Advanced Studies of Pisa. He gained a PhD in Management at the University of Padova, Italy. Before starting the PhD program, he studied in the field of economics at the University of Verona (bachelor degree) and in management for social economy at the University of Bologna (master’s degree), Italy. His research fields include resilience and crisis management, with a special focus on the pre-crisis stage and the small and medium-sized enterprises, and topics at the intersection between Civil Economy and Management. His works have been published in *Journal of Small Business and Enterprise development*.

Andrea Piccaluga is Full Professor of Innovation management at the Institute of Management and delegate for Technology Transfer at Scuola Superiore Sant’Anna, Pisa. He is a Vice President of Netval (www.netval.it), the Italian network of University Technology Transfer Offices and Associate editor of the *R&D Management*. He has co-edited various special issues, including “R&D Management at a time of crisis: what are we learning from the response to the COVID-19 pandemic?”, published in *R&D Management Journal* and published several articles in leading management journals, including *Technovation*, *R&D Management Journal*, *Small Business Economics*, *California Management Review*, *Journal of Technology Transfer*, *Journal of Knowledge Management*, *Technology Analysis & Strategic Management*, *European Management Journal*, and *Creativity and Innovation Management*. His research and teaching mainly deal with R&D and innovation management, open innovation, and the collaboration between firms and research institutions in the innovation process.

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher’s web site.

