



**Sinergie SIMA**  
Management Conference



# **Management of sustainability and well-being for individuals and society**

*Conference Proceedings*

*Short Papers*

*Parma (Italy)*

**13-14 June 2024**

Sinergie-SIMA Management Conference Proceedings  
*Management of sustainability and well-being for individuals and society*  
13-14 June 2024  
University of Parma - Italy

ISBN 978-88-94-7136-5-7

The Conference Proceedings are published online on <https://www.sijmsima.it>

© 2024 FONDAZIONE CUEIM  
Via Interrato dell'Acqua Morta, 26  
37129 Verona - Italy



# **Management of sustainability and well-being for individuals and society**

**13-14 June 2024**

## *Conference Proceedings*

Short Papers

edited by

*Arabella Mocciaro Li Destri, Marta Ugolini,  
Angeloantonio Russo and Savino Santovito*

# The “5Vs” of Big Data and the Propensity to Generate Patents with Big Data Technologies in SMEs

ENRICO MARCAZZAN<sup>1</sup> SAVERIO BARABUFFI<sup>2</sup> GIULIO FERRIGNO<sup>3</sup> ANDREA PICCALUGA<sup>4</sup>

## Abstract

*In the contemporary digital landscape, the exponential growth of data, commonly referred to as Big Data, has become indispensable for businesses worldwide. Understanding the fundamental characteristics of Big Data, encapsulated by the “5Vs” (Volume, Variety, Velocity, Veracity, and Value), is crucial for deriving value and fostering innovation. This research explores the nexus between these “5Vs” and the propensity of Small and Medium-sized Enterprises (SMEs) to generate patents based on Big Data technologies. The study employs a knowledge-based view (KBV) framework to investigate this relationship, focusing on Italian SMEs. Through a questionnaire-based survey involving 135 SMEs, the research analyzes the impact of each of the “5Vs” on the generation of patents with Big Data technologies. The methodology utilizes multinomial probit analysis to assess the likelihood of developing different levels of patents in relation to the “5Vs” of Big Data. Findings suggest that while the “Variety” of data positively influences SMEs’ propensity to generate patents, other factors such as “Volume”, “Velocity”, “Veracity”, and “Value” show varying degrees of significance. Notably, the study reveals a counterintuitive relationship where a higher volume of data impedes SMEs’ patent generation activities. Similarly, factors like real-time data processing speed, data quality, and value extraction from Big Data show no significant impact on patent generation. This research underscores the importance of understanding the nuanced interplay between different dimensions of Big Data and innovation outcomes, particularly for SMEs. The findings suggest that SMEs should prioritize enhancing the diversity and sources of their data over sheer volume or speed of processing. Moreover, the study emphasizes the need for SMEs to consider the potential “dark side” effects of Big Data, advocating for a shift towards leveraging “Smart Data” rather than focusing solely on Big Data. In conclusion, this work contributes novel insights by highlighting the complex relationship between the “5Vs” of Big Data and SMEs’ patent generation activities. The findings offer valuable guidance for SMEs seeking to navigate the digital landscape effectively, innovate, and maintain competitiveness in the market.*

**Key words:** big data; patents; SMEs.

**Framing of the research.** *In the digital age in which we live, the amount of data generated every day is growing exponentially (Cappa et al., 2021). This immense amount of data, commonly known as Big Data, has become an invaluable resource for businesses and institutions around the world (Ferrigno et al., 2023). Specifically, by Big Data we mean “any set of data that, with traditional systems, would require large capabilities in terms of storage space and time to be analyzed” (Del Vecchio et al., 2018, p. 6). However, managing, analyzing and deriving value from this mass of information requires not only advanced technological tools, but also a thorough understanding of its fundamental characteristics, the so-called “5Vs”: Volume, Variety, Velocity, Veracity and Value (Acciarini et al., 2023). “Volume” refers to the amount of data (Ghasemaghahi and Calic, 2019). With the advent of the Internet of Things (IoT), social media, and other digital devices, the volume of data has exploded to unprecedented levels. This enormous flow of information offers endless opportunities for analysis and innovation. “Variety” refers to the different types of data collected and processed by an enterprise (Ghasemaghahi, 2021; Pedota, 2023). This diversity includes text, images, audio, video, and more. The ability to integrate and analyze a wide range of data is critical to fully understand the context and gain meaningful insights for firms’ R&D activities (Ferrigno et al., 2023). The “Velocity” of Big Data refers to “the speed and the frequency of processing and integrating data” (Ghasemaghahi and Calic, 2020, p. 150; Ghasemaghahi et al., 2017). With the world becoming increasingly connected and real-time, the ability to analyze data in real time becomes essential to seize immediate opportunities and make informed decisions. “Veracity”, on the other hand, is defined as*

---

<sup>1</sup> Sant’Anna School of Advanced Studies, Pisa  
e-mail: enrico.marcazzan@santannapisa.it

<sup>2</sup> Sant’Anna School of Advanced Studies, Pisa  
e-mail: saverio.barabuffi@santannapisa.it

<sup>3</sup> Sant’Anna School of Advanced Studies, Pisa  
e-mail: giulio.ferrigno@santannapisa.it

<sup>4</sup> Sant’Anna School of Advanced Studies, Pisa  
e-mail: andrea.piccaluga@santannapisa.it

“reliability and credibility of data” (Ghasemaghahi and Calic, 2019; p. 73). Data quality is critical to ensure that analyses and decisions based on them are valid and reliable. Verification and management of data quality are therefore crucial in any analysis process. The fifth V, “Value”, refers to “firm’s ability to create value from big data” (Cappa et al., 2021, p. 53; Ghasemaghahi, 2021). In this sense, the value of Big Data is not intrinsic, but depends on the ability to interpret and use the data effectively to achieve desired goals.

The potential of Big Data has taken a central role in the development and growth of not only large organizations but also small and medium-sized enterprises (SMEs) (Vitale et al., 2010). SMEs are a key pillar of the global economy, contributing significantly to innovation and value creation (Müller et al., 2017). Understanding the relationship between the “5Vs” of Big Data and the propensity of SMEs to generate patents based on Big Data technologies is crucial to take full advantage of the opportunities offered by this digital revolution (Del Vecchio et al., 2018). Specifically, the importance of relying on patents stems from the fact that they can be valid measures of innovation and, when related to Big Data technologies, excellent measures to better understand the evolution of Big Data technologies (Saheb and Saheb, 2020). In addition, specifically for SMEs, patents can be considered as valid predictors of business performance (Müller et al., 2017).

**Purpose of the paper.** The purpose of the present “work-in-progress” is to understand which “Vs” have the greatest impact on the propensity to generate patents with Big Data technologies, considering the fact of the potential “dark side” effects of Big Data (Cappa et al., 2021). Through the lens of the knowledge-based view (KBV, Grant, 1996), the present research work aims to address the following research question:

*RQ: Which Vs of Big Data have an impact on the generation of patents based on Big Data technologies in SMEs?*

To address this research question, we give below some reasons for the relationship between “5Vs” and the propensity to develop patents on big data.

### **Volume**

The huge volume of available data offers multiple opportunities for innovation in various sectors (Cappa et al., 2021), including for SMEs (Del Vecchio et al., 2018). Companies can develop new technologies and solutions based on Big Data to improve their products, services, or operational processes, resulting in new ideas (Johnson et al., 2017) that can be protected through patents. In a competitive environment, access to big data can provide a significant strategic advantage (Ghasemaghahi and Calic, 2019; Ghasemaghahi, 2021). Firms that can effectively exploit the volume of available data can develop innovative and advanced solutions that set them apart from competitors, potentially leading to patents for their distinctive technologies. Nonetheless, managing large volumes of data can be expensive (Cappa et al., 2021; CoalaData, 2014; Trabucchi et al., 2017), especially for SMEs that are characterized by resource scarcity. As suggested by recent empirical work, the role of “Volume” on business outcomes may be characterized by non-significance (Ghasemaghahi, 2021; Ghasemaghahi and Calic, 2019) or even negative effects (Cappa et al., 2021). In line with the above, our first hypothesis is as follows:

*H1: Volume is negatively correlated with the propensity to develop patents with big data technologies.*

### **Velocity**

The speed at which data are generated and analyzed can affect a company’s ability to adapt quickly to changing market conditions (Ghasemaghahi and Calic, 2019). Technologies that enable real-time data processing and analysis can be invaluable in developing innovative solutions that improve operational efficiency or the ability to respond promptly to customer needs.

The ability to rapidly collect and analyze data can enable companies to identify emerging trends, market opportunities, or potential problems (Johnson et al., 2017). Technological solutions that facilitate this responsiveness can be patentable, thus protecting the firm’s investment in these advanced capabilities. Moreover, in a rapidly changing environment, the speed at which data become obsolete is a critical factor (D’Aveni et al., 2017; Johnson et al., 2017). Companies that develop technologies to manage and analyze large volumes of data in real time can reduce the risk of information obsolescence and gain a competitive advantage. These innovative solutions can be patented to protect their uniqueness and value. In line with the above, our second hypothesis is as follows:

*H2: Velocity is positively correlated with the propensity to develop patents with big data technologies.*

### **Variety**

The diversity of data types available requires innovative solutions to effectively integrate and analyze all these sources of information (Ghasemaghahi and Calic, 2020). Companies that develop technologies to manage the variety of Big Data extract value from this diversity can gain a competitive advantage. These solutions may be subject to patenting to protect their intellectual property.

Different types of data require specific analytical approaches to extract meaningful information. Companies that develop innovative algorithms or methodologies to analyze heterogeneous data can gain a competitive advantage (Ghasemaghahi, 2021). These solutions can be patented to protect their uniqueness and ensure a return on R&D investment. Moreover, the variety of Big Data provides opportunities to customize services and offerings to the specific

needs of customers (Johnson et al., 2017). Companies that develop technologies to collect, integrate, and analyze data from multiple sources can offer highly personalized services that increase customer satisfaction and loyalty. These innovative solutions may be subject to patenting to protect their intellectual property and maintain a competitive advantage in the market. In line with the above, our third hypothesis is as follows:

*H3: Variety is positively correlated with the propensity to develop patents with big data technologies.*

#### **Veracity**

Data quality and reliability are crucial to obtaining accurate and meaningful results from Big Data analysis (Cappa et al., 2021). Companies that develop technologies to assess and improve the veracity of data can gain a competitive advantage. These solutions may be subject to patenting to protect their intellectual property and ensure customer confidence in the results of data analysis. Data “Veracity” can be compromised by cybersecurity threats, such as unauthorized access, data breaches, or manipulation (Brook, 2019). Companies that develop technologies to protect the veracity and integrity of sensitive data can gain a competitive advantage. These innovative solutions may be subject to patenting to protect their intellectual property and ensure the security of customer information. Moreover, much of Big Data is unstructured data, such as text, images, or video. Managing and analyzing this information effectively requires innovative solutions (Ghasemaghahi and Calic, 2020). Companies that develop technologies to extract value from unstructured data can gain a competitive advantage. These solutions may be subject to patenting to protect their intellectual property and ensure a return on R&D investment.

*H4: Veracity is positively correlated with the propensity to develop patents with big data technologies.*

#### **Value**

Extracting value from Big Data through data analysis and interpretation can enable companies to identify new market opportunities, optimize business operations, and improve customer satisfaction (Cappa et al., 2021; Ghasemaghahi, 2021). Innovative technologies that harness the value of Big Data may be subject to patenting to protect their uniqueness and ensure a competitive advantage in the marketplace.

Using Big Data to optimize business processes, reduce costs, and improve operational efficiency can generate significant savings and competitive advantages (Brynjolfsson et al., 2011; Ghasemaghahi, 2019; Tambe, 2014). Companies that develop technologies to extract value from Big Data and improve their internal operations can gain a competitive advantage. These innovative solutions can be patented to protect their intellectual property and ensure a return on their R&D investment. Moreover, using Big Data to better understand customer needs and preferences can enable businesses to offer highly personalized services and improve customer satisfaction (Cappa et al., 2021; Johnson et al., 2017). Innovative technologies that allow Big Data to be analyzed and used to personalize services may be subject to patenting to protect their uniqueness and ensure a competitive advantage in the market.

*H5: Value is positively correlated with the propensity to develop patents with big data technologies.*

**Methodology.** We collected the data of 135 Italian SMEs by submitting a questionnaire to gather information and data on both the Big Data characteristics, the patents with Big Data technologies done by the SME and other firm’s characteristics. More specifically, to analyze the relationship among the “5Vs” of Big Data and the propensity to generate patents with Big Data technologies in SMEs, we propose a multinomial probit (Keane, 1992; McCulloch and Rossi, 1994). since our dependent variables (“patents generated through Big Data Technologies”) is a categorical variable (in line with Messeni Petruzzelli et al., 2022). “Patents generated through Big Data Technologies” is subdivided into 4 classes: 1) “no patents”; 2) “1 patent”, 3) “between 2 and 4 patents”; 4) “more than 4 patents”. Then, we introduce as independent the 5 characteristics of Big Data: “Volume”, “Velocity”, “Variety”, “Veracity”, and “Value” which are operationalized according to a 7-point Likert scale. Then, we control for firm characteristics, such as “age”, “revenues”, “size”, and for regional and sectoral effects by introducing the dummy variables “region”, and “high technological sector”.

**Results.** By performing the econometric analysis, we estimate the likelihood of developing “1 patent”, “between 2 and 4 patents”, and “more than 4 patents” generated through Big Data Technologies with respect to none. Our preliminary analysis shows that “Volume” has a negative relationship for the category “1 patent” and “between 2 and 4 patents”, while it does not have any statistically significant effect on “more than 4 patents”. Then, we find that “Variety” has a positive and statistically significant effect on the category “between 2 and 4 patents”, and “more than 4 patents”. For the variables “Velocity”, “Veracity”, and “Value”, we do not find any statistical support for the three categories of patents, meaning that these Big Data characteristics do not have a statistically significant relationship with developing patents generated through Big Data technologies. We can conclude that we find support for our hypothesis H1 and H3, but not for H2, H4, and H5. The huge volume of data hinders the SMEs’ propensity of developing patents generated through Big Data Technologies, meaning that managing large volumes of data is expensive for SMEs (Cappa et al., 2021). However, managing the diversity of available data and sources improves the SMEs’ innovation performances and helps them maintain a competitive advantage in the market. Hence, SMEs can leverage the

heterogeneity of their data to improve their position in the innovative landscape and gain market power by protecting their intellectual property through patenting.

**Research limitations.** Our “work-in-progress” is characterized by some limitations. First, the research aims to investigate the relationship between the “5Vs” of Big Data and the propensity to generate patents with Big Data technologies individually, without investigating any internal relationships among the “5Vs”. Second, we focus only on patents with Big Data technologies as the dependent variable, without testing the impact of these patents on business performance. Third, our hypotheses are based on “correlation” between variables, without necessarily investigating “causal” relationships. These three main limitations open up insights for possible future research on the topic. In this sense, further research could find important interrelations between the “5Vs”, their effects on the business performance through patents with Big Data technologies, and “causal” relationships.

**Managerial implications.** Our work has several implications for SMEs. The “5Vs” of Big Data and their correlation with SMEs’ propensity to generate patents are key elements in ensuring success and competitiveness in today’s business landscape. Companies that understand and fully exploit the potential of Big Data can gain significant advantages and contribute significantly to innovation and economic growth. In this sense, instead of focusing on Big Data “Volume”, “Velocity”, “Veracity” and “Value”, SMEs should focus on improving the degree of “Variety” of Big Data. In this sense, SMEs should invest in the number of sources of big data.

**Originality of the paper.** The main originality of our “work-in-progress” relies on the fact that, as suggested by Plantec et al. (2023), we focus on data characteristics (i.e., “5Vs”) instead of technology (e.g., artificial intelligence) offering a complementary view of Plantec and colleagues’ work of the innovative activities of the firms. Moreover, the paper contributes to highlight the fact that Big Data are not good by themselves, but there are some “dark side” effects that must be taken into consideration adopting them. In this sense, our research suggests that it is better to rely on “Smart Data” rather than Big Data.

In conclusion, our paper presents counterintuitive findings. First, generally speaking, it is common to think that SMEs can effectively manage and analyze this huge volume of data in order to have a significant competitive advantage. However, our results suggest that higher “Volume” of data can have negative effects on the propensity to generate patents with Big Data technologies. Therefore, SMEs should focus on other “Vs” instead of the first (i.e., “Volume”). Second, in general terms, SMEs that can react quickly to market changes and emerging trends, supported by real-time analysis of Big Data (i.e., “Velocity”), can develop innovative solutions that merit patent protection. However, our research shows that it does not have a significant effect on the propensity to generate patents with Big Data technologies. Third, SMEs that invest in accurate data collection and management (i.e., “Veracity”) can develop solutions based on reliable information, increasing the credibility of their inventions and the likelihood of obtaining patents. However, our results show that it has no significance. Fourth, the concept of “Value” emphasizes the importance of turning data into meaningful information and creating tangible value for the business and customers. However, based on the results of our study, big data patents are not influenced by SMEs’ identification and exploitation of the inherent value of Big Data. Finally, only SMEs that synergistically integrate and analyze these different types of data (i.e., “Variety”) can develop innovative, multidimensional solutions that are patentable. The “Variety” dimension represents the only positive and significant factor for our results.

## References

- ACCIARINI C., CAPPÀ F., BOCCARDELLI P., ORIANI R. (2023), “How can organizations leverage big data to innovate their business models? A systematic literature review”, *Technovation*, vol. 123, 102713.
- BROOK, C. (2019). What is data breach or cyber security insurance? Digital Guardian. Available at: <https://digitalguardian.com/blog/what-data-breach-or-cyber-security-insurance>.
- BRYNJOLFSSON E., HITT L.M., KIM H.H. (2011), “Strength in numbers: How does data-driven decision making affect firm performance?”, Available at SSRN 1819486, 33.
- CAPPÀ F., ORIANI R., PERUFFO E., 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*, vol. 38 n. 1, pp. 49-67.
- COOLADATA (2014). The true cost of building a big data solution. New York, USA: Cooladata, Available at: <https://www.cooladata.com/blog/true-cost>.
- D’AVENI R.A., DAGNINO G.B., SMITH K.G. (2010). “The age of temporary advantage”, *Strategic Management Journal*, vol. 31 n. 13, pp. 1371-1385.
- DEL VECCHIO P., DI MININ A., PETRUZZELLI A.M., PANNIELLO U., PIRRI, S. (2018), “Big data for open innovation in SMEs and large corporations: Trends, opportunities, and challenges”, *Creativity and Innovation Management*, vol. 27 n. 1, pp. 6-22.
- FERRIGNO G., CRUPI A., DI MININ A., RITALA P. (2023), “50+ years of R&D Management: a retrospective synthesis and new research trajectories”, *R&D Management*, vol. 53 n. 5, pp. 900-926.
- FERRIGNO G., DEL SARTO N., PICCALUGA A., BARONCELLI A. (2023), “Industry 4.0 base technologies and business models: a bibliometric analysis”, *European Journal of Innovation Management*, vol. 26 n. 7, pp. 502-526.
- GHASEMAGHAEI M. (2019), “Are firms ready to use big data analytics to create value? The role of structural and psychological readiness”, *Enterprise Information Systems*, vol. 13 n. 5, pp. 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*, vol. 57, 102055.

- GHAEMAGHAEI M., CALIC G. (2019), “Does big data enhance firm innovation competency? The mediating role of data-driven insights”, *Journal of Business Research*, vol. 104, pp. 69-84.
- GHAEMAGHAEI M., CALIC G. (2020), “Assessing the impact of big data on firm innovation performance: Big data is not always better data”, *Journal of Business Research*, vol. 108, pp. 147-162.
- GHAEMAGHAEI M., HASSANEIN K., TUREL, O. (2017), “Increasing firm agility through the use of data analytics: The role of fit”, *Decision Support Systems*, vol. 101, pp. 95-105.
- GRANT R.M. (1996), “Toward a knowledge-based theory of the firm”, *Strategic Management Journal*, vol. 17 n. S2, pp. 109-122.
- JOHNSON J.S., FRIEND S.B., LEE H.S. (2017), “Big data facilitation, utilization, and monetization: Exploring the 3Vs in a new product development process”, *Journal of Product Innovation Management*, vol. 34 n. 5, pp. 640-658.
- KEANE M.P. (1992), “A note on identification in the multinomial probit model”, *Journal of Business & Economic Statistics*, vol. 10 n. 2, pp. 193-200.
- MCCULLOCH R., ROSSI P.E. (1994), “An exact likelihood analysis of the multinomial probit model”, *Journal of Econometrics*, vol. 64 n. 1-2, pp. 207-240.
- MÜLLER D., TE Y.F., JAIN P. (2017), “Predicting business performance through patent applications”, *In 2017 IEEE International Conference on Big Data (Big Data)*, vol. December, pp. 4159-4164.
- MESSENI PETRUZZELLI A., MURGIA G., PARMENTOLA A. (2022), “How can open innovation support SMEs in the adoption of I4. 0 technologies? An empirical analysis”, *R&D Management*, vol. 52 n. 4, pp. 615-632.
- PEDOTA M. (2023), “Big data and dynamic capabilities in the digital revolution: The hidden role of source variety”, *Research Policy*, vol. 52 n. 7, 104812.
- PLANTEC Q., DEVAL M.A., HOOGE S., WEIL B. (2023), “Big data as an exploration trigger or problem-solving patch: Design and integration of AI-embedded systems in the automotive industry”, *Technovation*, vol. 124, 102763.
- SAHEB T., SAHEB T. (2020), “Understanding the development trends of big data technologies: an analysis of patents and the cited scholarly works”, *Journal of Big Data*, vol. 7 n. 1, 12.
- TAMBE P. (2014), “Big data investment, skills, and firm value”, *Management Science*, vol. 60 n. 6, pp. 1452-1469.
- TRABUCCHI D., BUGANZA T., PELLIZZONI E. (2017), “Give away your digital services”, *Research-Technology Management*, vol. 60 n. 2, pp. 43-52.
- VITALE G., CUPERTINO S., RICCABONI A. (2020), “Big data and management control systems change: the case of an agricultural SME”, *Journal of Management Control*, vol. 31, pp. 123-152.